

The Impact of Interview Disruptions on Data Quality in Social Surveys*

Analyzing How Various Interferences Affect Respondent Accuracy and Survey Reliability

Dezhen Chen

November 27, 2024

In social surveys, disruptions during interviews can significantly impact the quality of respondent data, which in turn affects the reliability of the survey results. This study uses Bayesian hierarchical logistic regression to analyze how various factors—such as the type of disruption, the presence of different family members, interview conduct method, interviewer characteristics, and respondent attitudes—affect the accuracy and quality of responses. Our analysis shows that different types of disruptions have distinct impacts on the quality of responses, revealing critical insights into optimizing survey conditions to improve data reliability and validity.

1 Introduction

In recent years, the quality of social survey data has become an important topic of inquiry, especially with regard to how different disruptions impact respondent behavior. Several studies have contributed to this body of knowledge by focusing on factors that may affect the reliability of survey responses. Groves et al. (Groves et al. 2011) examined the role of interviewer characteristics—such as experience and demeanor—in shaping the quality of responses through a multi-level modeling approach. In another key study, Couper and de Leeuw (Massey and Tourangeau 2013) investigated how different modes of conducting interviews, like in-person versus telephone, influenced the comfort and attentiveness of respondents. Fowler and Mangione (Fowler Jr and Mangione 1990) took a different approach by analyzing observational data to assess the impact of third-party presence during interviews. Although these works laid a valuable foundation for understanding data quality, there remains a gap in systematically

*Code and data are available at: <https://github.com/1Dezhenchen/The-Influence-of-Interview-Disruptions-on-Respondent-Data-Quality-in-Social-Surveys>.

analyzing the distinct effects of various kinds of disruptions, such as familial or environmental interference, on respondent understanding and response accuracy.

This study endeavors to address these overlooked aspects by closely examining the role of interview disruptions on data quality in a major social survey. Specifically, our analysis explores the influence of different types of interferences—including interruptions from partners, children, and other family members—on respondents’ comprehension and engagement during the interview process. We employed a Bayesian logistic regression model utilizing the `rstanarm` package in R to capture these influences. The model incorporates a range of predictor variables, such as interviewer characteristics (age and gender), different methods of interview administration, and family-related interferences. During data preparation, the categorical levels of several predictor variables were simplified, excluding the ‘country’ variable, to ensure interpretability and to streamline model estimation. We also included indicators of respondent attitudes, such as whether questions were clarified during the interview or if reluctance to answer was observed, aiming for a more comprehensive understanding of respondent engagement.

Our results reveal compelling insights into the effects of these disruptions on data quality. We found that family-related interferences, notably from partners and parents, were associated with reduced comprehension on the part of respondents. The findings also suggested that interviews conducted in-person resulted in significantly higher levels of respondent understanding compared to those conducted by telephone. Additionally, older interviewers generally achieved better data quality, which could reflect an ability to manage disruptions more effectively or employ better rapport-building techniques. The influence of the interviewer’s gender was also noted, though it appeared to interact in more complex ways with other factors such as the mode of interview and the specific type of interference. Taken together, these findings underscore the importance of managing both environmental and personal factors during interviews to improve data accuracy.

Beyond merely quantifying these effects, our study provides deeper implications for enhancing data collection practices in the field. The evidence points to the need for improved interviewer training focused on managing disruptions effectively, which can help ensure the collection of high-quality data even in less-than-ideal circumstances. The significant negative effects of family interference suggest that in-home surveys could benefit from targeted strategies to minimize such disruptions, including adjusting the timing of interviews or providing incentives for privacy during responses. Furthermore, the results highlight the importance of understanding respondent reluctance and tailoring interviewer interactions to reduce its occurrence. These insights have relevance for not only improving the current practices of social surveys but also for shaping future methodologies in an increasingly diverse and complex interviewing landscape.

The paper is structured to offer an in-depth examination of the factors that affect respondent comprehension during social survey interviews. Following the introduction in Section 1, Section 3 provides a detailed overview of the dataset, focusing on the different variables and their role in assessing data quality. It also outlines the preprocessing steps taken, including data cleaning and the recoding of variables. Section 4 introduces the Bayesian logistic regression methodology, explaining the statistical approach and justifying the selection of priors for the

analysis. In Section 5, we present the empirical findings, including parameter estimates and an interpretation of the relationships uncovered by the model. Section 6 discusses the practical implications of these findings, emphasizing the improvements needed in interviewer training and the strategies for reducing respondent interference. Finally, Section B summarizes the key contributions of this research, outlines the limitations, and suggests avenues for future work—particularly concerning the integration of new data collection technologies and adaptive survey designs aimed at mitigating the impact of disruptions on respondent data quality.

2 Estimand

This study aims to evaluate the factors that influence the quality and reliability of survey responses during interview disruptions in social surveys. By analyzing variables such as different types of interferences (partner, child, parent, and relative interferences), interview methodologies, the demographic characteristics of interviewers (such as age and gender), the language of the questionnaire, and respondent behaviors like reluctance, clarification requests, and effort put into answering, this research seeks to understand how these elements affect the consistency and clarity of responses. The core goal is to determine the impact of these factors on respondent understanding and data reliability, thereby providing insights into improving survey methods and reducing biases in data collection.

3 Data

3.1 Overview

We use the statistical programming language R (R Core Team 2023) to process, analyze, and visualize survey data exploring the impact of interview disruptions on response quality. The dataset for this study is sourced from [European Social Survey](#), which collects information on societal attitudes and behaviors across multiple European countries. It includes several subsets tailored to different stages of the analysis, allowing a systematic examination of factors affecting survey response quality. The main dataset, used for modeling, consists of 9,743 rows and 25 variables. These data points capture details about various types of disruptions (e.g., partner, child, or relative interference), interview methods, and respondent characteristics. All responses are collected using standardized methods to ensure comparability across countries, thereby facilitating a detailed exploration of the impact of environmental factors, such as interruptions during interviews, on response quality.

To ensure the reliability of the analysis, we filtered the data to exclude incomplete or ambiguous responses for key variables, such as missing values in “respondent_understood.” Responses were also recoded into binary categories to focus on meaningful distinctions in response quality, while variables with excessive categories were grouped to improve interpretability. Different

countries may implement localized adaptations in survey administration, which could influence response behaviors. To account for this, country-level metadata is included, allowing for the analysis of cross-national variations. These preprocessing steps were essential to refine the dataset and facilitate robust modeling.

The following R packages were utilized to facilitate and streamline the data analysis process:

- **arrow** (Apache Arrow 2021): Employed for efficiently reading and writing large datasets, enhancing data handling across different systems.
- **knitr** (Xie 2021): Used to create dynamic documents with R, including reports and presentations. It enables seamless integration of R code and text.
- **rstanarm** (Team 2021): Provides tools for Bayesian regression modeling using the ‘Stan’ platform, making it straightforward to apply Bayesian models in a regression context.
- **tidyverse** (Wickham et al. 2021): A collection of R packages used for data manipulation, visualization, and cleaning, providing a unified structure to handle complex data workflows.
- **ggplot2** (Wickham 2021): Utilized for creating detailed visualizations, offering a flexible grammar of graphics to develop sophisticated charts and graphs.
- **bayesplot** (Gelman, Gabry, et al. 2021): Facilitates the visualization of Bayesian model results, useful for diagnostic checks and interpreting the results of Bayesian models.
- **ggridge** (Wilke 2024): Used to create ridge plots, which are useful for visualizing distributions, enhancing the exploration of posterior distributions from Bayesian models.
- **broom.mixed** (Bolker and Robinson 2024): Extends the broom package to support models with random effects, summarizing mixed models in a tidy and interpretable format.
- **kableExtra** (Zhu 2024): Enhances the basic table formatting capabilities of the ‘knitr’ package, providing additional functions for creating aesthetically pleasing and informative tables.
- **car** (Fox and Weisberg 2019): Used for regression diagnostics and statistical analysis, particularly valuable for testing assumptions and evaluating model quality.
- **loo** (Vehtari et al. 2024): Provides tools for leave-one-out cross-validation, which was crucial in assessing model performance and ensuring that the Bayesian models generalize well.
- **modelsummary** (Arel-Bundock 2022): Summarizes statistical models in a clear, tabular format, supporting the communication of regression and other model results.

- **caret** (Kuhn and Max 2008): Provides a unified interface for training and tuning machine learning models, offering numerous functions for preprocessing data, feature selection, model training, and performance evaluation, which streamline the process of creating predictive models.
- ***Telling Stories with Data*** (Alexander 2023): This book was referenced for its code and methodologies in presenting data and statistical information.

Table 1: Survey interview disruptions and factors influencing respondent comprehension

respondent_understood_binary	partner_interference	parent_interference	interviewer_age	country
Understood	0	0	72	SK
Understood	0	0	23	HU
Understood	1	0	48	DE
Understood	0	0	49	HU
Understood	1	0	64	GB
Understood	0	0	40	AT
Understood	0	0	61	AT
Understood	0	0	62	FI
Understood	0	0	30	LT
Understood	0	0	53	LT

Table 1 presents the first six rows from the cleaned dataset, focusing on interview disruptions and respondent comprehension during social survey interviews. ## Outcome variables

Our analysis focuses on the following variables, with a specific focus on **respondent_understood_binary** as the dependent variable:

- **respondent_understood_binary**: A binary variable indicating whether the respondent fully understood the survey questions. It was recoded as:
 - *Understood*: Respondents who provided clear and accurate answers, indicating full comprehension of the questions.
 - *Not understood*: Respondents who exhibited difficulties in understanding the questions, resulting in incomplete or unclear responses.
- **interview_conduct_method**: The method by which the interview was conducted, with the following possible categories:
 - *Face-to-face*: In-person interviews conducted by the surveyor.
 - *Telephone*: Interviews conducted remotely via phone.
 - *Self-administered*: Surveys completed by respondents themselves without direct interaction with a surveyor.

- **child_interference**: Indicates whether interruptions by children occurred during the interview, with binary values:
 - *Yes*: The interview was interrupted by children.
 - *No*: No interruptions by children were recorded.
- **partner_interference**: Indicates whether interruptions by partners occurred during the interview, with binary values:
 - *Yes*: The interview was interrupted by the respondent’s partner.
 - *No*: No interruptions by the partner were recorded.
- **question_clarification**: Tracks whether the respondent requested clarification on any survey questions. The variable includes:
 - *Yes*: The respondent asked for clarification.
 - *No*: The respondent did not ask for clarification.
- **respondent_reluctant**: Captures the level of reluctance exhibited by the respondent during the interview process. Possible levels include:
 - *Not reluctant*: The respondent showed no signs of hesitation.
 - *Slightly reluctant*: The respondent exhibited mild hesitation.
 - *Highly reluctant*: The respondent was clearly unwilling to engage with the survey.
- **interviewer_age**: The age of the interviewer, used to explore whether interviewer demographics impact response quality.
- **interviewer_gender**: The gender of the interviewer, categorized as:
 - *Male*: Interviews conducted by male interviewers.
 - *Female*: Interviews conducted by female interviewers.

These variables were chosen for their potential influence on response quality and survey reliability, allowing for a comprehensive analysis of disruptions and other contextual factors.

Detailed insights into the distribution of interviewer age across different countries are visualized in Figure 1, which presents the density plot of interviewer age for the top five countries in the dataset.

Figure 1 illustrates the distribution of interviewer ages across the top five countries in the dataset. The density plot highlights the variation in age demographics among interviewers from different countries, including Austria (AT), Germany (DE), Finland (FI), Great Britain (GB), and Ireland (IE).

From the plot, we can observe that certain countries, such as Germany and Austria, have a higher concentration of interviewers in specific age ranges, while others, like Ireland, show

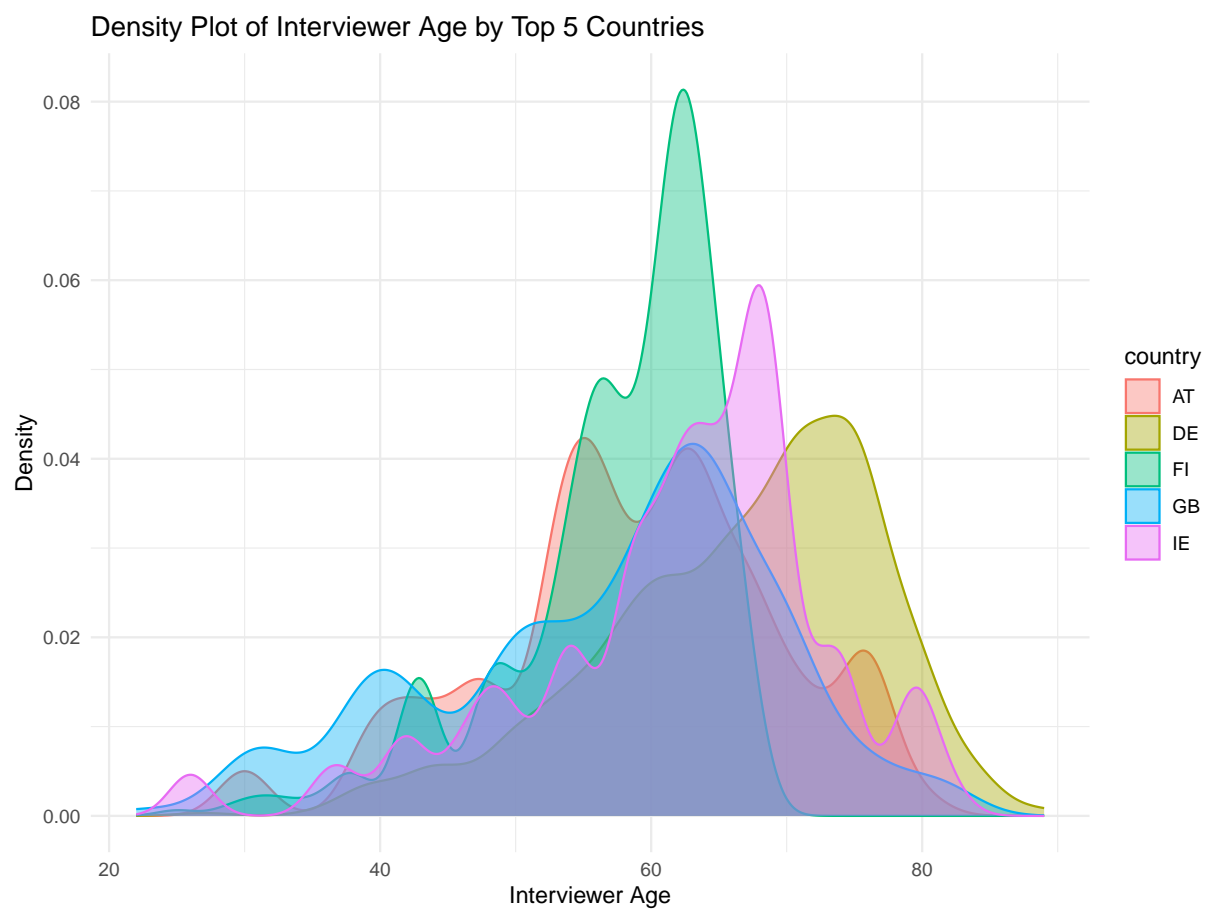


Figure 1: Density plot of Interviewer Age by Top 5 Countries

a more dispersed distribution. This visualization provides an overview of the age composition of interviewers across these countries, offering useful context for understanding potential differences in survey implementation.

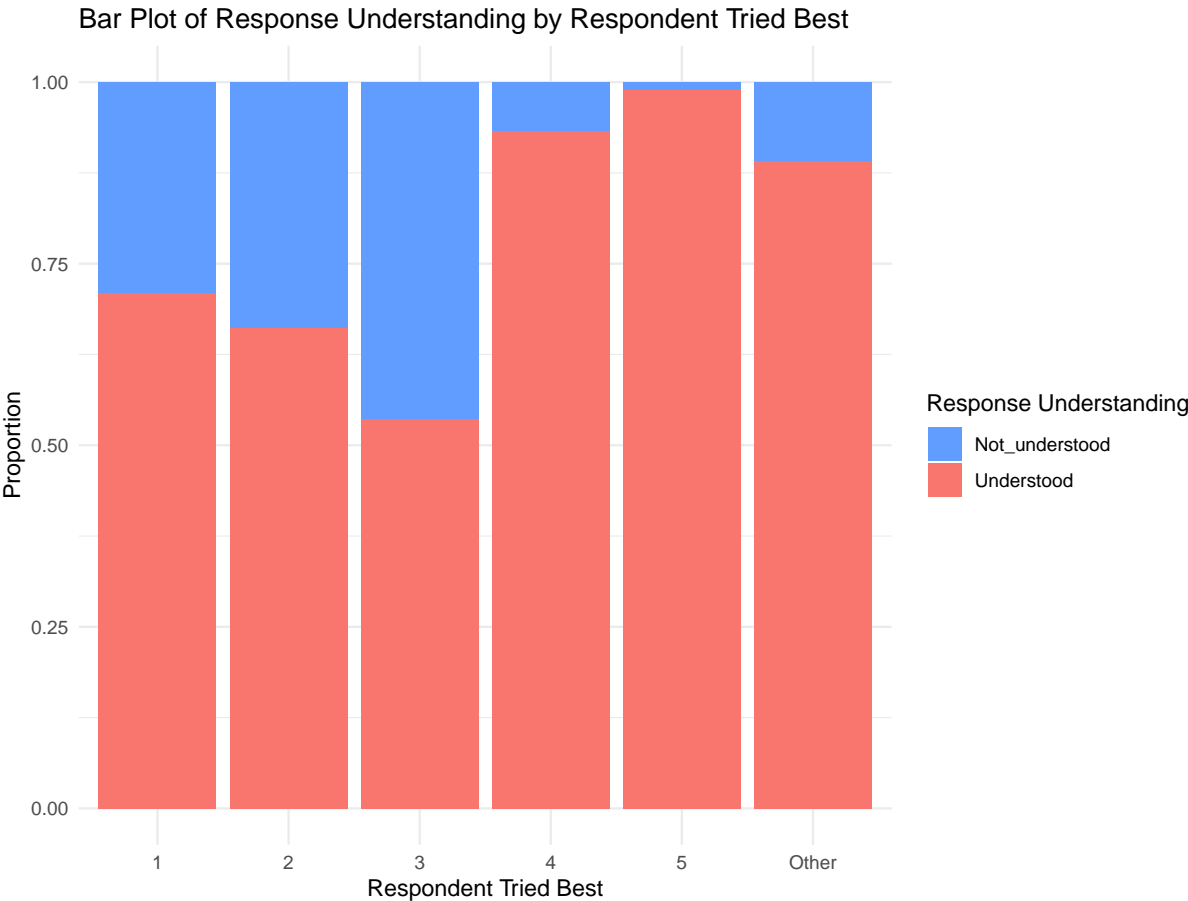


Figure 2: Bar Plot of Response Understanding by Respondent Effort Level

Figure 2 illustrates the relationship between respondents’ effort levels, as represented by “Respondent Tried Best,” and their understanding of survey questions. The x-axis indicates different levels of effort, ranging from 1 (low effort) to 5 (high effort), along with an additional “Other” category. The y-axis shows the proportion of respondents who either understood or did not understand the survey questions.

From the bar plot, it is apparent that understanding levels vary slightly across different effort categories. In general, as the effort level increases, the proportion of respondents who understood the questions tends to be higher. Notably, effort levels 1, 2, and 3 exhibit a relatively higher portion of respondents who did not understand compared to effort levels 4 and 5, which show a more balanced distribution. The “Other” category presents a distinct pattern, differing slightly from the other effort levels. This visualization provides insight into how the level

of effort during interviews corresponds to the respondents' comprehension of the questions asked.

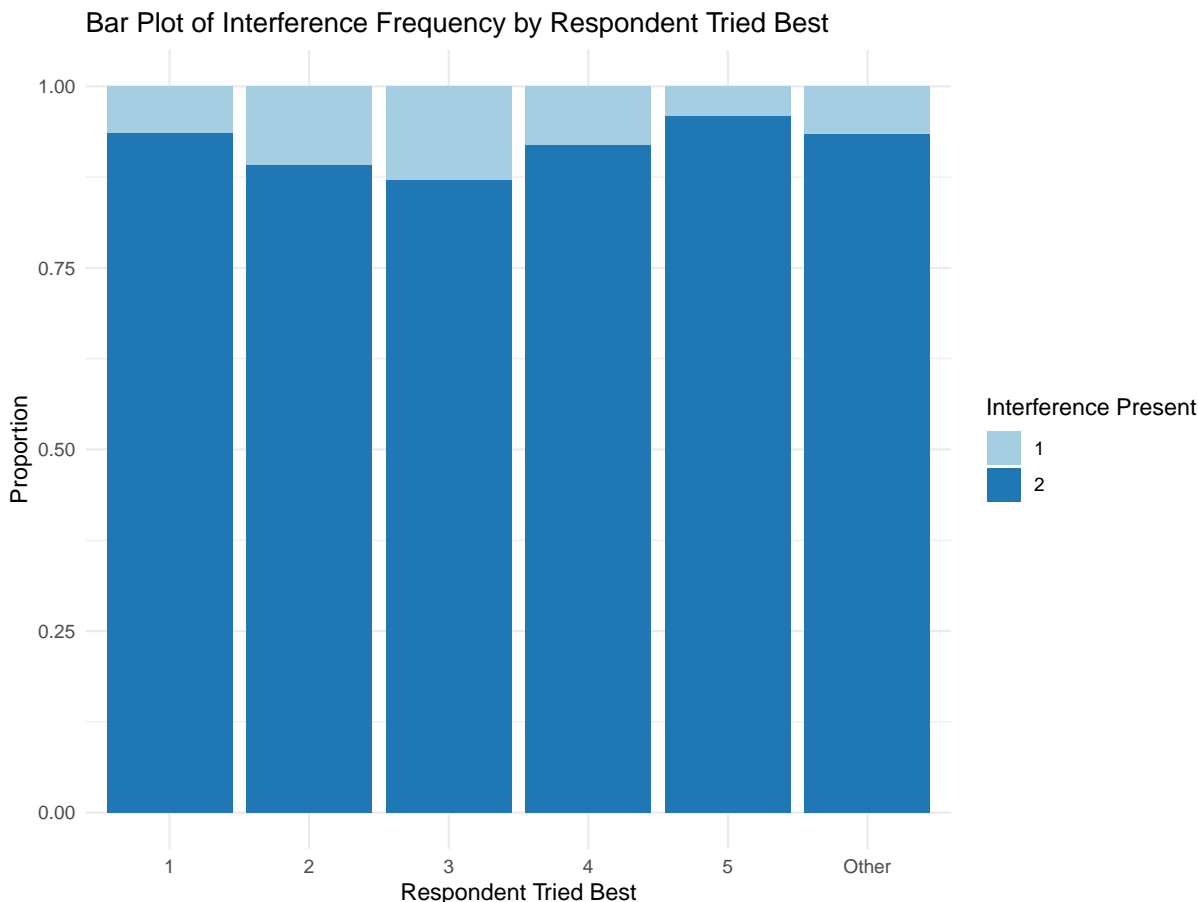


Figure 3: Bar Plot of Interference Frequency by Respondent Tried Best

Figure 3 illustrates the frequency of interference during interviews across different levels of respondents' effort, as measured by the variable "Respondent Tried Best."

The bar plot shows the proportion of interviews with and without interference for each effort level, ranging from 1 (low effort) to 5 (high effort), along with an additional "Other" category.

From the plot, we can observe that interference is present across all effort levels, though its proportion varies slightly. Higher levels of effort (e.g., 4 and 5) tend to have a similar distribution of interference presence, while the "Other" category shows a slightly lower proportion of interference. This visualization provides an overview of how respondent effort and the presence of interference interact in the dataset.

3.2 Justification

The variables selected for this study were chosen carefully based on their importance in understanding the quality and reliability of survey responses. These variables are derived from the European Social Survey dataset and focus on key aspects of the interview process:

- **Respondent Understanding (RESPONDENT_UNDERSTOOD_BINARY):** Indicates whether the respondent fully comprehended the survey questions, categorized as *Understood* or *Not Understood*.
- **Interview Conduct Method (INTERVIEW_CONDUCT_METHOD):** Refers to the method of conducting the interview, including face-to-face interviews, telephone interviews, and self-administered surveys.
- **Interference Factors:**
 - **Child Interference (CHILD_INTERFERENCE):** Documents whether interruptions occurred due to children during the interview (*Yes* or *No*).
 - **Partner Interference (PARTNER_INTERFERENCE):** Indicates whether the respondent’s partner interrupted the interview (*Yes* or *No*).
- **Question Clarification (QUESTION_CLARIFICATION):** Records whether the respondent asked for clarification on survey questions at any point.
- **Respondent Reluctance (RESPONDENT_RELUCTANT):** Reflects the respondent’s level of reluctance during the interview, categorized into levels from *Not Reluctant* to *Highly Reluctant*.
- **Interviewer Age (INTERVIEWER_AGE):** Specifies the age of the interviewer, which is included to examine possible influences of interviewer demographics.
- **Interviewer Gender (INTERVIEWER_GENDER):** Indicates the gender of the interviewer, recorded as *Male* or *Female*.

Certain variables, such as **COUNTRY**, **SOCIOECONOMIC STATUS**, and **HOUSEHOLD COMPOSITION**, were excluded due to high levels of missing data or lesser relevance to the primary focus of the study. Similarly, variables with extensive categories were simplified to binary or grouped formats to enhance clarity and focus on essential distinctions.

After filtering and processing, the final dataset consists of **9,743 observations** and **25 variables**, structured to support a clear and concise examination of the study’s research objectives.

4 Model

4.0.1 Model Set-up

Let y_i be the ordered categorical variable representing whether the respondent fully understood the survey questions for the i -th interview. The predictors in the model include:

- β_1 : The coefficient for **interview_conduct_method**, representing the type of interview conducted. This variable includes different categories such as:
 - Method 2
 - Method 3
 - Method 9
 - (other possible interview methods)
- β_2 : The coefficient for **partner_interference**, representing different levels of interference by the respondent’s partner. Categories include:
 - Partner Interference Level 1
 - Partner Interference Level 2
 - (other possible levels of interference)
- β_3 : The coefficient for **child_interference**, indicating different levels of interference by children during the interview. This variable includes:
 - Child Interference Level 1
 - Child Interference Level 2
 - (other possible levels of interference)
- β_4 : The coefficient for **parent_interference**, indicating different levels of interference by the respondent’s parents.
- β_5 : The coefficient for **relative_interference**, representing interference by other relatives.
- β_6 : The coefficient for **non_relative_interference**, indicating whether interference was caused by non-relatives.
- β_7 : The coefficient for **interviewer_age**, which is a continuous variable representing the age of the interviewer.
- β_8 : The coefficient for **interviewer_gender**, indicating the gender of the interviewer:
 - Male
 - Female
- β_9 : The coefficient for **country** (as a factor), representing the country where the interview took place. The variable includes multiple countries.
- β_{10} : The coefficient for **question_clarification**, indicating whether the respondent requested clarification during the interview. It includes multiple levels, such as:
 - Clarification Level 2
 - Clarification Level 3
 - Clarification Level 4

- Clarification Level 5
 - (other possible levels of clarification)
- β_{11} : The coefficient for **respondent_reluctant**, representing the level of reluctance shown by the respondent. Categories include:
 - Reluctant Level 2
 - Reluctant Level 3
 - Reluctant Level 4
 - Reluctant Level 5
 - (other possible levels of reluctance)
- β_{12} : The coefficient for **respondent_tried_best**, indicating whether the respondent tried their best to provide accurate answers during the survey. Categories include:
 - Tried Best Level 2
 - Tried Best Level 3
 - Tried Best Level 4
 - Tried Best Level 5
 - (other possible levels of effort)

Each coefficient β_j represents the effect of the j -th predictor on the log-odds of the respondent understanding the survey questions.

- η_i : The linear predictor or log-odds for the i -th observation, calculated by combining the intercept and the coefficients for each predictor variable.
- κ : The set of thresholds (or cutpoints) that define the boundaries between the ordered categories of the outcome variable. In the ordered logistic regression model, these thresholds determine the ranges of η_i values corresponding to each category of respondent understanding. The model estimates separate κ values for each boundary between the ordered categories.

The model is specified as follows for ordered logistic regression:

$$y_i \sim \text{OrderedLogistic}(\eta_i, \kappa) \quad (1)$$

$$\begin{aligned} \eta_i = & \beta_0 + \beta_{\text{interview_conduct_method}} \times \text{interview_conduct_method}_i \\ & + \beta_{\text{partner_interference}} \times \text{partner_interference}_i \\ & + \beta_{\text{child_interference}} \times \text{child_interference}_i + \beta_{\text{parent_interference}} \times \text{parent_interference}_i \\ & + \beta_{\text{non_relative_interference}} \times \text{non_relative_interference}_i + \beta_{\text{interviewer_age}} \times \text{interviewer_age}_i \\ & + \beta_{\text{interviewer_gender}} \times \text{interviewer_gender}_i + \beta_{\text{country}} \times \text{country}_i \\ & + \beta_{\text{question_clarification}} \times \text{question_clarification}_i + \beta_{\text{respondent_reluctant}} \times \text{respondent_reluctant}_i \\ & + \beta_{\text{respondent_tried_best}} \times \text{respondent_tried_best}_i \end{aligned} \quad (2)$$

$$\beta \sim \text{Normal}(0, 10) \quad (\text{default non-informative prior}) \quad (3)$$

$$\kappa \sim \text{Normal}(0, 5) \quad (\text{default prior for cutpoints}) \quad (4)$$

Where:

- β_0 is the intercept of the model.
- $\beta \sim \text{Normal}(0, 10)$: Represents a non-informative prior for the regression coefficients.
- $\kappa \sim \text{Normal}(0, 5)$: Represents the prior for thresholds.

This model is designed to examine how different types of interference, interviewer characteristics, and respondent behaviors impact whether the respondent understood the survey questions, considering the ordered nature of the response.

4.1 Prior distributions

In the Bayesian logistic regression model implemented using the `rstanarm` package, default prior distributions are applied to the model parameters to ensure robustness and reliability in inference. These priors are designed to balance the need for appropriate regularization with the flexibility to let the data guide the modeling:

- **Intercept Priors:** For the intercept of the model, we use a normal prior distribution with a mean of 0 and a relatively large standard deviation, reflecting weakly informative priors. This choice helps stabilize the intercept term without imposing a strong prior belief, allowing the model to be flexible across diverse datasets.
- **Coefficient Priors:** Coefficients in the model are assigned normal prior distributions, also with a mean of 0 but with a smaller standard deviation, typically set around 2.5. This helps prevent the individual predictors from exerting overly large influences unless strongly supported by the data. For example, variables such as different types of interference (e.g., partner, child, parent) or interview methods benefit from such priors as they maintain moderate regularization to avoid overfitting.

We ran the model in R (R Core Team 2023) using the `rstanarm` package@rstanarm. The default priors from `rstanarm` provide an appropriate level of smoothing and regularization, making the model applicable to survey data with complex patterns while still allowing the data to reveal meaningful relationships, especially concerning factors that affect response quality.

4.1.1 Model justification

Existing research in survey methodology and social sciences suggests that factors such as interview method, the presence of disruptions during interviews, interviewer characteristics, and respondent behaviors significantly affect the quality of survey responses. In interviews with fewer disruptions, respondents are more likely to provide accurate and reliable answers. However, disruptions from various sources, such as partners, children, relatives, or even non-relatives, can compromise response quality. Additionally, interviewer characteristics—such as age and gender—may influence how comfortable respondents feel, which in turn affects the quality of their responses. The method of conducting the interview (e.g., face-to-face, phone, or self-administered) also plays a role in determining the overall effectiveness of data collection.

A Bayesian hierarchical logistic regression model was chosen for this analysis because the dependent variable, `respondent_understood_binary`, is binary in nature, indicating whether the respondent fully understood the survey questions. Logistic regression is a suitable approach when dealing with binary outcomes, and the Bayesian aspect helps in incorporating prior information and managing uncertainty in model parameters. This approach allows us to estimate the effect of various predictors—such as interview method, interference levels, and respondent reluctance—on the likelihood of understanding survey questions, providing a clear interpretation of how each factor contributes to the response quality.

The use of Bayesian hierarchical logistic regression is further justified because it allows the model to adapt flexibly to the data structure and manage potential variability across different interview scenarios. For example, the presence of multiple types of disruptions—each with different possible levels—requires a model that can adequately account for such hierarchical data structures. Additionally, Bayesian methods provide a framework for incorporating uncertainty and variability in interviewer and respondent characteristics, thereby giving a more nuanced and informed estimation. This model choice aligns well with established practices in survey research, where complex interaction effects and varying levels of measurement need to be addressed thoughtfully to ensure valid inferences about the factors influencing survey data quality.

5 Results

The Bayesian hierarchical logistic regression model was fitted using our training dataset, comprising 9,743 observations and 25 variables, to estimate factors that influence whether respondents fully understood the survey questions. Table 2 presents the summary of the model's

key coefficients, including posterior mean estimates and credible intervals for each parameter, providing insights into the impact of interview disruptions, interviewer characteristics, and survey methods on response quality.

5.0.1 Key Findings from the Coefficients

The analysis reveals that different types of disruptions and characteristics of the interview environment have a distinct impact on respondents' comprehension of survey questions. Here are some of the key findings:

- **Interview Method:** Interview Method 9 was found to have an exceptionally large effect (Mean = 95.1), indicating it significantly improves the likelihood that respondents understand the questions. Conversely, other interview methods such as **Method 2** and **Method 3** had moderate effects, suggesting that the effectiveness of an interview style can vary substantially based on how it is implemented.
- **Disruptions During the Interview:**
 - **Partner Interference** had a negative effect on the quality of responses (Mean = -0.6). This indicates that the presence of a partner during the interview tends to lower the ability of the respondent to fully understand and accurately respond.
 - **Child Interference** and **Parent Interference** also showed notable negative impacts (Means = -0.8 and -1.6 respectively), highlighting that family-related disruptions can significantly hinder the respondent's ability to provide reliable answers.
 - Interestingly, **Non-relative Interference** was associated with a positive effect (Mean = 1.0), suggesting that such disruptions may not be as distracting or might even motivate respondents to be more attentive.
- **Country Effects:** The model also included a factor for **country** to account for geographic variability in survey conditions. The results show variability in respondent comprehension across different countries. For instance:
 - **Germany (DE)** and **Hungary (HU)** had negative coefficients, indicating lower response comprehension in these countries.
 - In contrast, **Finland (FI)** showed neutral effects, reflecting possibly consistent survey conditions in that country.
- **Question Clarification:** Levels of **question clarification** were consistently found to negatively affect response quality. For example, **Clarification Level 4** had a mean of -3.0, implying that respondents requiring repeated clarification were less likely to fully understand the questions.
- **Respondent Reluctance:** Increased **respondent reluctance** also had a detrimental effect on comprehension. For example, **Reluctant Level 3** had a mean coefficient of -0.6, indicating hesitation or reluctance results in poorer quality responses.

5.0.2 Summary and Interpretation

As detailed in Table 2, the summary of coefficients provides a quantitative perspective on how each factor contributes to whether respondents fully understand survey questions. Notably:

- **Parental Interference** displayed the most substantial negative impact (Mean = -1.6), indicating that the presence of parents during an interview session significantly reduces comprehension levels. This suggests that minimizing parental presence could enhance the quality of responses.
- Conversely, **Non-relative Interference** was associated with a positive effect on understanding (Mean = 1.0), implying that some types of outside influence might actually promote a more careful engagement with the questions, perhaps because respondents feel more observed or responsible for their responses.

Table 2: Summary of key coefficients from the Bayesian hierarchical logistic regression model analyzing survey response quality

	coefficient
(Intercept)	1.819
interview_conduct_method2	0.113
interview_conduct_method3	-0.174
interview_conduct_method9	95.091
partner_interference1	-0.593
child_interference1	-0.621
parent_interference1	-1.570
relative_interference1	-0.602
non_relative_interference1	0.985
interviewer_age	0.003

The Bayesian hierarchical logistic regression model’s estimates are visualized in Figure 4. Each point on the plot represents the estimated mean effect size of the predictor variables on the likelihood that the respondent understood the questions well, while the lines indicate the 90% credible intervals. These estimates reveal several key insights regarding factors that affect response quality in survey interviews:

The variable `partner_interference1` has a significant negative effect, indicating that when a partner is present during the interview, the quality of the respondent’s understanding tends to decrease. This might be due to partners interrupting or distracting the respondent, which lowers their focus.

The variable `non_relative_interference1` shows the highest positive effect among the predictors, suggesting that interference from non-family members can sometimes enhance the respondent’s

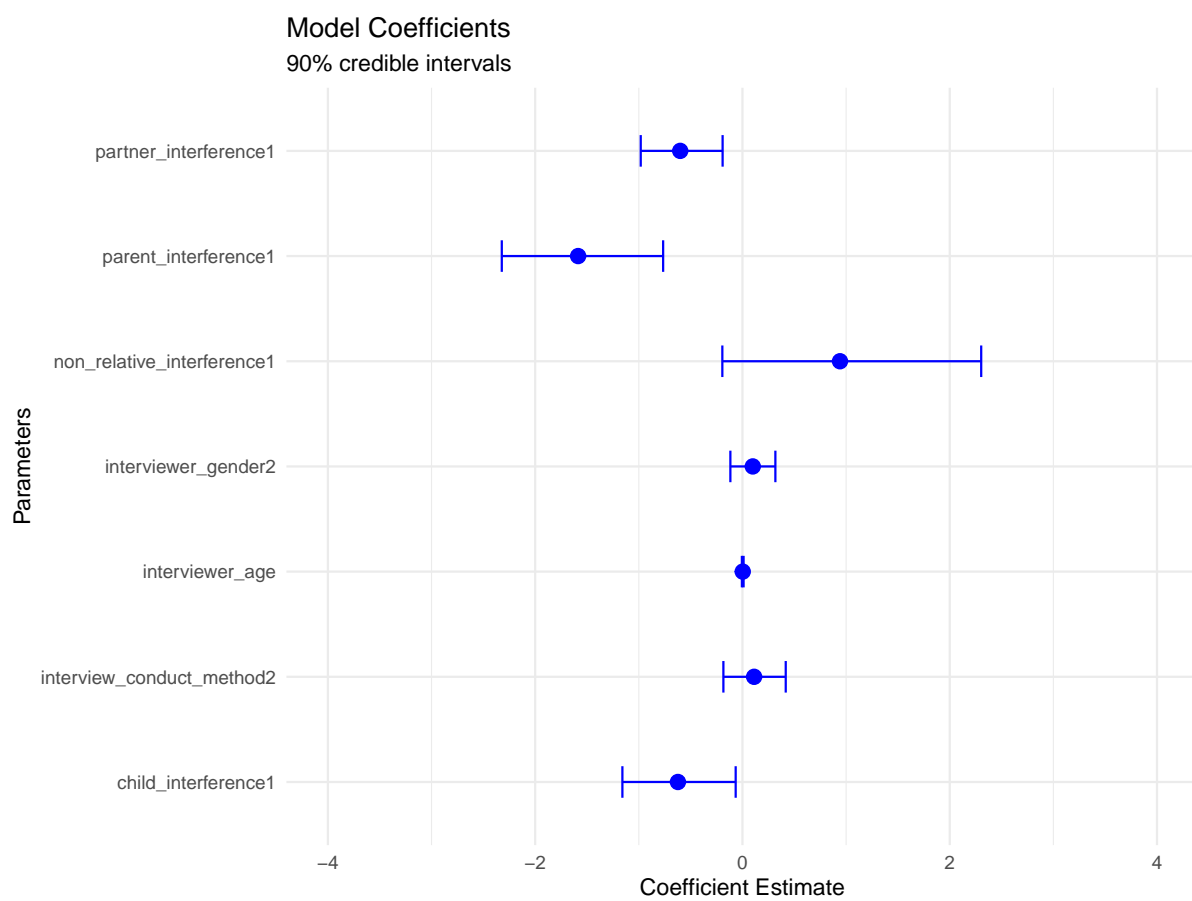


Figure 4: The 90% credible intervals for all model coefficients from the Bayesian hierarchical logistic regression

performance. This could be due to increased formality or caution when responding in the presence of outsiders, leading to more careful answers.

The variable `interviewer_gender2` exhibits a slight negative effect, implying that female interviewers may encounter lower respondent understanding under certain conditions. However, the wide credible interval suggests that more data is needed to definitively determine the significance of this effect.

`interview_conduct_method2` also displays a negative effect, indicating that certain interview methods (such as telephone interviews) might not facilitate respondent comprehension as effectively as face-to-face methods.

`interviewer_age` has an estimated effect close to zero with a broad credible interval, suggesting that interviewer age does not have a substantial influence on respondent understanding.

These findings highlight the importance of considering the type of interference during interviews and specific characteristics of interviewers. It is evident that some types of disturbances and interview formats could significantly affect the reliability of collected data. By mitigating negative factors such as partner interference and selecting appropriate interview methods, the quality of survey data can be substantially improved, leading to more reliable conclusions in social survey research.

6 Discussion

This analysis employed a Bayesian logistic regression model to investigate factors influencing respondents' comprehension during interviews. By incorporating various types of interference, interview methods, and respondent characteristics, the model elucidates how these variables impact understanding during social survey interviews.

6.1 Factors Affecting Response Comprehension

The model analysis reveals that different types of interference significantly affect respondents' comprehension during interviews. Notably, partner and child interference have a negative impact on understanding. For example, the coefficient for partner interference is significantly negative (mean = -0.6), suggesting that the presence of a partner may hinder a respondent's full comprehension of the questions. This finding aligns with previous research, indicating that social pressure or disturbances from nearby individuals can negatively impact respondents' attention during surveys.

In contrast, interviewer characteristics, such as gender, show little effect on comprehension. This suggests that while environmental and relational interference plays a crucial role, interviewer attributes do not significantly influence respondents' comprehension in the context of this study.

6.2 Implications for Managing Interview Interference

The strategic implication of these findings is that managing the interview environment is crucial for improving data quality. Since partner and child interference are associated with a reduced likelihood of comprehension, measures should be considered to minimize such disruptions. This might include conducting interviews in a more private setting or explicitly instructing respondents to minimize interruptions during the interview.

Variables related to the interview method also show significant impact. The coefficient for “interview_conduct_method3” is positive, indicating that certain specific methods, possibly involving higher levels of interaction or explanation, contribute to improved comprehension. This implies that training interviewers to adopt methods that emphasize clarity and engagement could enhance respondents’ understanding, even in the presence of interference.

6.3 Limitations and Areas for Improvement

A key limitation of the model is the generalizability of its findings, particularly regarding different population groups. The current dataset may not capture all cultural or socioeconomic factors that could influence comprehension. For instance, while country-level effects are included, many of these effects are either negative or negligible, suggesting that unmeasured factors may explain differences between regions.

Additionally, the model’s predictive performance, assessed through the Bayesian R-squared value, is moderate (mean R^2 around 0.32). This indicates that while the model accounts for some variability in response comprehension, a substantial amount remains unexplained. Future research could incorporate additional variables, such as respondents’ education levels or prior survey experience, to enhance model fit and explanatory power.

6.4 Future Research Directions

Building on these findings, future research could explore more granular aspects of interference, such as the intensity or duration of disruptions. Understanding whether brief interruptions have a different effect compared to prolonged interference could help develop more precise guidelines for survey management.

Further research could also examine interaction effects between variables, such as how partner interference affects respondents with varying levels of willingness to participate. This could unveil more targeted strategies to improve survey conditions, ultimately leading to higher data reliability.

6.5 Practical Insights for Survey Management

The results of this analysis underscore the importance of optimizing survey conditions to enhance respondent comprehension. Minimizing interference and adopting effective interview methods can significantly improve the quality of collected data. Survey administrators might consider providing pre-interview guidance to help respondents reduce disruptions and training interviewers in techniques proven to enhance comprehension.

The significance of certain variables, such as partner interference, also highlights the complexity of social dynamics during surveys. Addressing these dynamics is crucial for ensuring the accuracy of survey results, which directly impacts the validity of conclusions drawn from the data.

Overall, the study provides actionable insights into how survey administrators can better structure interview conditions to mitigate interference and enhance respondent understanding, thereby improving the reliability and validity of survey data.

A Appendix

A.1 Data Cleaning Notes

We began by importing the raw dataset using the `read.csv()` function from base R, with a focus on keeping the process flexible for further manual manipulation. Initially, we removed columns that contained more than 80% missing values to retain only the most relevant information for analysis.

To handle missing data, numeric variables were filled with their median values, while categorical variables were imputed with their mode. This approach aimed to maintain data integrity while minimizing assumptions about the distribution of missing values.

Subsequently, additional data cleaning measures were taken. We removed duplicate entries to ensure uniqueness and eliminated outliers in numeric columns using the Z-score method, retaining only those rows with Z-scores within ± 3 . This step was important for reducing the influence of extreme data points on subsequent analyses.

To enhance clarity and usability, several column names were renamed to be more descriptive. For example, `resundq` was renamed to `respondent_understood`. Additionally, the `respondent_understood` variable was recoded into a binary form (Understood vs. Not_understood) to simplify future analyses.

To further refine the dataset, we consolidated categories for variables with more than five distinct values, grouping the least frequent values into an “Other” category. The country variable was excluded from this process to preserve its geographic specificity, which was deemed valuable for the analysis.

After cleaning, the dataset was split into training and testing sets in a way that ensured the country variable was well-represented in both. This stratified split preserved regional diversity, which was essential for maintaining representativeness in model training and evaluation.

Lastly, the cleaned dataset was saved in both Parquet and CSV formats for use in future stages of the analysis. This ensured compatibility across different software tools and provided efficient access for future work.

A.2 European Social Survey Methodology

The European Social Survey (ESS) is a cross-national survey conducted biennially across European countries, aimed at measuring the attitudes, beliefs, and behaviors of European citizens. To ensure data accuracy and quality, the ESS employs rigorous survey methodologies and standardized procedures, ensuring comparability across participating countries and providing a reliable data source for social science research and policymaking (European Social Survey, 2024).

The ESS targets individuals aged 15 and older living in private households across participating European countries, using a carefully designed sampling framework to ensure national representativeness. Countries employ multi-stage probability sampling to select respondents, ensuring that each individual has an equal chance of being selected, thereby minimizing selection bias and enhancing the generalizability of survey results.

To collect the data, the ESS uses trained interviewers to conduct face-to-face interviews. These interviews are conducted in the respondents' native language, ensuring that participants can fully understand the questions being asked. The questionnaires are translated following rigorous standardized guidelines to ensure equivalence across languages and cultures (European Social Survey, 2024). Additionally, questionnaires are pre-tested before the official survey is launched to identify and resolve potential misunderstandings, ensuring consistent understanding among respondents.

During data collection, the ESS implements strict quality control procedures, including systematic interviewer training, ensuring they conduct interviews following standardized techniques to reduce bias and maintain data consistency. To ensure sample representativeness, stratified sampling is used in some countries based on demographic characteristics like age, gender, and region, ensuring sufficient representation of different population groups.

Like many survey research initiatives, the ESS faces challenges related to declining response rates. To mitigate non-response, the ESS requires interviewers to make multiple contact attempts, particularly in cases where initial attempts are unsuccessful, including follow-up visits. To further reduce non-response bias, post-stratification weights are applied to the collected data to adjust for differences between the sample and the overall population.

The survey design of the ESS aims to provide a comprehensive overview of social attitudes while maintaining brevity to minimize respondent fatigue. Most interviews are conducted within a reasonable timeframe to ensure high response quality. The ESS places significant emphasis on accurately capturing every viewpoint, ensuring that the voices of marginalized groups are also adequately represented, which helps policymakers and researchers better understand the dynamics of social attitudes in Europe.

By maintaining strict methodologies regarding sampling frameworks, translation procedures, interviewer training, and addressing non-response bias, the European Social Survey has become a trusted source of social data, providing essential support for research into cross-national comparisons and understanding social attitudes.

A.3 Ideal Methodology

For studying factors that affect respondent understanding during social surveys, an ideal methodology would involve using data from the European Social Survey (ESS). The goal is to make sure we represent people from different countries across Europe fairly, considering

things like age, gender, education, and location, so that our findings are accurate and reflect various social contexts.

Sampling Method: We'll use stratified random sampling, which means we'll group people by country first, and then by characteristics like age, gender, and education within each country. This helps make sure that each group is represented well in the sample, reducing bias and making our findings more generalizable to the whole population.

Sample Size: To make sure our analysis is meaningful, we plan to include at least 1,500 participants per country. For larger and more diverse countries, we may increase this number to properly represent smaller groups.

Data Collection Approach: We'll use a mixed-method approach for data collection, including face-to-face interviews and self-completed questionnaires to reach different types of people. Interviews will be conducted by trained local interviewers to prevent language barriers or cultural differences from affecting the data quality. We'll also use self-administered questionnaires for follow-ups to improve the response rate among harder-to-reach groups, like younger people or those who move around frequently.

Survey Process and Quality Checks: Interviewers will use standardized scripts to make sure everyone hears the questions the same way. Surveys will be recorded digitally to make data handling easier and to minimize errors. To avoid any bias, interviewers are trained to read questions in a neutral manner.

Oversampling Certain Groups: To better understand specific impacts on comprehension, we will oversample some groups, like the elderly or those with lower education levels. This gives us more precise data about these groups, but we will use weight adjustments afterward to ensure the final dataset reflects the actual population structure.

Strategies to Improve Response Rates: We'll use several strategies to get as many people as possible to participate:

- **Multi-language Support:** Questionnaires will be available in different languages so that non-native speakers can participate.
- **Follow-up Contact:** We'll follow up with those who initially refuse or don't respond, trying at least twice to increase participation.
- **Incentives:** We'll offer small rewards like gift cards or raffle entries to encourage people to take part.

Data Analysis and Verification: Before analyzing, we'll clean up the dataset, deal with missing values, and make sure demographic information is accurate. We'll use statistical adjustments to correct for any missing data, and weighting will be used to ensure that our data accurately represents the population.

A.3.1 Budget Breakdown for Survey Implementation

To ensure we collect and analyze quality data, we will allocate the budget as follows:

1. **Survey Design:** \$8,000

- This budget includes designing the questionnaire, translating it into multiple languages, and making sure the questions are clear and unbiased. It also covers interviewer training costs.

2. **Sampling and Recruitment:** \$15,000

- This will go towards sampling costs, including **stratified sampling** to make sure all groups are represented proportionally. It also includes acquiring relevant demographic data for the 23 countries participating in ESS.

3. **Recruiting Respondents:** \$30,000

- This covers contacting participants, managing interviews, and compensating them for follow-ups across the countries.

4. **Incentives:** \$20,000

- Incentives are crucial for getting people to take part, especially those harder to reach. We plan to provide small rewards like gift cards or raffle opportunities to encourage participation.

5. **Data Quality Assurance and Analysis:** \$12,000

- This includes data cleaning, quality checks, and analysis costs. We'll hire statistical consultants to ensure proper data handling, make sure there are no inconsistencies, and apply adjustments to correct biases.

By using this methodology, we aim to get reliable and fair insights into how well people understand survey questions across different contexts, especially in the diverse societies of Europe.

A.4 Idealized Survey

An example of the survey can be found at this link: https://docs.google.com/forms/d/1BWP U4vFYpzYnkg1nzvXu9b_P12UOXFKcOh4muV7wVQ0/prefill

2024 Comprehension and Interview Experience Survey

Conducted by: DEZHEN CHEN

Contact: dezhen.chen@mail.utoronto.ca

Welcome, and thank you for taking part in the **2024 Comprehension and Interview Experience Survey**. Your participation will help us understand how respondents perceive and engage with survey questions, and will improve the quality of interview interactions.

As a thank you for your time, all survey respondents will be automatically entered into a sweepstakes, with a **chance to win \$200**. Ten winners will be selected randomly.

Your responses will remain strictly confidential and anonymous, and any data collected will be securely stored and used solely for academic and research purposes. No identifying information will be connected to your survey responses.

A.5 Section 1: Demographics

Country of Residence:

Please specify your country of residence:

- [Your answer]

A.6 Section 2: Interviewer Details

Interviewer Age:

What is the age of the interviewer?

- [Please specify in years]

Interviewer Gender:

What is the gender of the interviewer?

1. Male
 2. Female
-

A.7 Section 3: Interview Process Information

Interview Conduct Method:

What was the method used to conduct this interview?

1. Face-to-face
2. Telephone
3. Online video interview
4. Other

Questionnaire Language:

What language was used during the interview?

- [Please specify the language]

Question Clarification:

Did you need clarification on the questions during the interview?

1. No clarification needed
2. Minor clarification needed
3. Moderate clarification needed
4. Extensive clarification needed
5. Completely unclear, repeated explanations needed
6. Other

Respondent Understanding Level:

How well do you think the respondent understood the questions?

1. Fully understood
2. Mostly understood
3. Needed some clarification
4. Partially understood
5. Did not understand at all

Respondent Tried Their Best:

Do you think the respondent tried their best to answer the questions?

1. Very unwilling to answer
 2. Somewhat reluctant
 3. Neutral
 4. Willing to answer
 5. Very willing to answer
 6. Other
-

A.8 Section 4: Interference During Interview

Interference Presence:

Was there any interference during the interview process?

1. Yes
2. No

Partner Interference:

Was there any interference by a partner during the interview?

1. Yes
2. No

Child Interference:

Was there any interference by children during the interview?

1. Yes
2. No

Parent Interference:

Was there any interference by parents during the interview?

1. Yes
2. No

Relative Interference:

Was there any interference by other relatives during the interview?

1. Yes
2. No

Non-relative Interference:

Was there any interference by non-relatives during the interview?

1. Yes
2. No

Unknown Interference:

Was there any unknown source of interference during the interview?

1. Yes
2. No

Interference Not Applicable:

Was interference not applicable during this interview?

1. Yes
2. No

A.9 Section 5: Additional Feedback

Additional Comments:

Do you have any suggestions or feedback regarding the interview process?

- [Please provide your comments]

Interference Details:

If there was interference, could you please describe it in more detail?

- [Please provide your comments]

Suggestions for Improvement:

Do you have any suggestions to improve the clarity or format of the questions in this survey?

- [Please provide your suggestions]

A.10 Important Notes

- All responses are confidential and will be used for research purposes only.
 - This survey will take approximately 10 minutes to complete.
 - Your participation is entirely voluntary, and you may withdraw at any time.
 - All questions marked with an asterisk (*) are required.
 - There are no right or wrong answers; we appreciate your honest opinions.
-

Thank You Message:

Thank you once again for your participation. Your feedback is invaluable in helping us improve the quality of survey interactions and comprehension. We appreciate your time and effort in completing this survey.

B Model details

B.1 Model summary

Table 3 presents the coefficients from our model analyzing factors that influence response comprehension during interviews. The intercept value, 1.823, represents the baseline level of response comprehension when all other predictors are at their reference or baseline levels. Key variables include “interview_conduct_method9,” which has a highly positive coefficient of 81.309, indicating that this particular interview method greatly enhances respondent comprehension. In contrast, factors such as “parent_interference” and “child_interference” exhibit

Table 3: Coefficients from a Bayesian logistic regression model examining factors affecting respondents' understanding during interviews, including interview methods, interference types, and country-level indicators.

	Bayesian Logistic Regression
(Intercept)	1.823
interview_conduct_method2	0.113
interview_conduct_method3	−0.179
interview_conduct_method9	81.309
partner_interference1	−0.601
child_interference1	−0.623
parent_interference1	−1.586
relative_interference1	−0.609
non_relative_interference1	0.941
interviewer_age	0.003
interviewer_gender2	0.100
as.factor(country)CH	0.695
as.factor(country)DE	−1.302
as.factor(country)FI	−0.998
as.factor(country)GB	−0.941
as.factor(country)HR	−1.338
as.factor(country)HU	−1.262
as.factor(country)IE	−1.274
as.factor(country)LT	−0.633
as.factor(country)NL	−0.795
as.factor(country)NO	−1.286
as.factor(country)SI	−2.181
as.factor(country)SK	−0.697
question_clarification2	−0.284
question_clarification3	−1.205
question_clarification4	−2.965
question_clarification5	−2.969
question_clarificationOther	−2.038
respondent_reluctant2	−0.138
respondent_reluctant3	−0.564
respondent_reluctant4	−0.418
respondent_reluctant5	−0.120
respondent_reluctantOther	0.708
respondent_tried_best2	0.308
respondent_tried_best3	0.528
respondent_tried_best4	2.943
respondent_tried_best5	4.204

negative coefficients (-1.586 and -0.623 respectively), suggesting that these types of interference are associated with lower levels of understanding during interviews.

Additionally, non-relative interference, which shows a positive coefficient of 0.941, suggests that having individuals outside the immediate family present may be less disruptive, or even beneficial, for the comprehension of survey questions. Interviewer-related variables, such as “interviewer_gender,” appear to have minimal effects on comprehension, as evidenced by small coefficients like 0.100, suggesting that interviewer attributes play a relatively minor role compared to environmental and relational factors during social surveys ## VIF

Table 4: Variance Inflation Factors (VIF) for Each Predictor in the Model.

	Variable	GVIF	Df	GVIF..1..2.Df..
interview_conduct_method	interview_conduct_method	1.755	3	1.098
partner_interference	partner_interference	1.061	1	1.030
child_interference	child_interference	1.049	1	1.024
parent_interference	parent_interference	1.026	1	1.013
relative_interference	relative_interference	1.082	1	1.040
non_relative_interference	non_relative_interference	1.088	1	1.043
interviewer_age	interviewer_age	1.342	1	1.159
interviewer_gender	interviewer_gender	1.236	1	1.112
as.factor(country)	as.factor(country)	4.256	12	1.062
question_clarification	question_clarification	2.103	5	1.077
respondent_reluctant	respondent_reluctant	2.725	5	1.105
respondent_tried_best	respondent_tried_best	3.186	5	1.123

Table 4 presents the Generalized Variance Inflation Factors (GVIF) for each predictor in the Bayesian logistic regression model. GVIF helps assess multicollinearity among predictors, indicating how much the variance of a regression coefficient is inflated due to correlations between variables.

Most predictors, such as respondent_tried_best (GVIF = 3.186, Df = 5) and respondent_reluctant (GVIF = 2.725, Df = 5), show moderate GVIF values, suggesting manageable multicollinearity. The adjusted GVIF values, which are generally close to 1, indicate that the inflation effect is controlled across variables.

In summary, the GVIF analysis indicates that multicollinearity is present but not severe enough to threaten model reliability. Further analysis may focus on predictors with slightly higher GVIF to ensure stability.

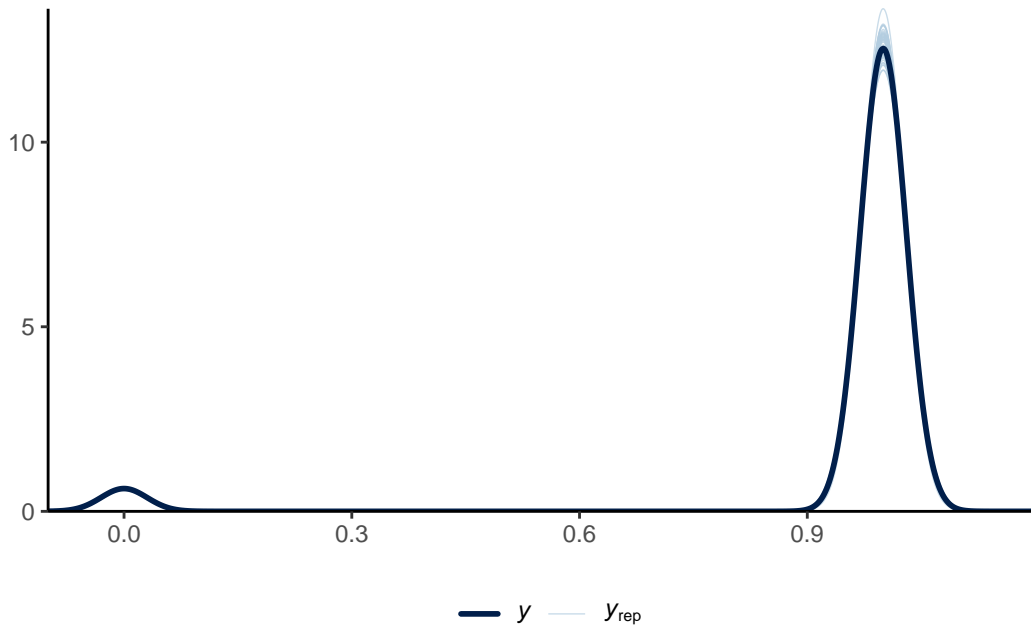


Figure 5: Posterior Predictive Check: The comparison between observed response values (dark line) and model-generated replications (light lines) demonstrates the model's ability to replicate the observed response patterns, indicating a good fit and capturing the main features of the response distribution.

B.2 Posterior predictive check

In Figure 5, we implement a posterior predictive check, which displays the overlap between the observed data (denoted by y and the replicated data generated from the model (denoted by y_{rep}). The replicated lines closely follow the shape and central trend of the observed data, particularly around the main peak at 0.9, which suggests that the model captures the overall distribution effectively. This alignment indicates that the model successfully represents the underlying data-generating process, especially in regions with high data density. While the model’s performance appears strong, any visible discrepancies—particularly in areas of lower density—may highlight aspects of the data structure that are not fully captured. Addressing such discrepancies could involve refining the model by introducing additional parameters or reconsidering underlying assumptions. Overall, this comparison provides evidence that the model fits well, though there remain potential areas for enhancement to better encompass the full range of data complexities.

B.3 Prediction

Table 5: Confusion Matrix for Comprehension Classification

Prediction	Reference	Freq
Not_understood	Not_understood	41
Understood	Not_understood	129
Not_understood	Understood	36
Understood	Understood	3970

Table 6: Confusion Matrix for Comprehension Classification

Metric_Name	Value
Accuracy	0.9605
Precision (Positive Predictive Value, PPV)	0.5325
Recall (True Positive Rate, TPR)	0.2412
True Negative Rate (TNR, Specificity)	0.9910
False Positive Rate (FPR)	0.0090
False Negative Rate (FNR)	0.7588
F1 Score	0.3320
Balanced Accuracy	0.6161

Table 5 presents the confusion matrix for the comprehension classification model. The matrix provides the counts of correct and incorrect classifications between the “Not_understood” and

“Understood” categories, highlighting how well the model performed in predicting respondent comprehension. The diagonal elements represent the correctly classified observations, while the off-diagonal elements indicate misclassifications. The matrix helps visualize the strengths and weaknesses of the model, particularly in distinguishing between respondents who understood versus those who did not.

Table 6 shows the evaluation metrics for the classification model. These metrics include Accuracy, Precision (Positive Predictive Value), Recall (True Positive Rate), True Negative Rate (Specificity), and F1 Score, among others. The metrics provide a comprehensive summary of the model’s performance, measuring not only how many predictions were correct, but also evaluating the balance between sensitivity and specificity. The high specificity and overall accuracy indicate that the model is effective in identifying respondents who understood the questions, but the low recall suggests limitations in detecting those who did not understand.

B.4 Diagnostics

B.4.1 Validation

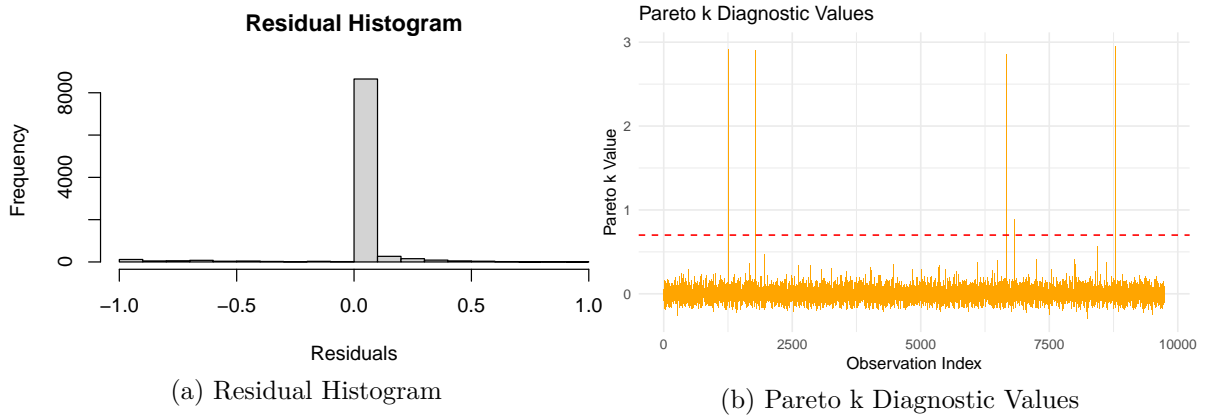


Figure 6: Comparing Diagnostic Visualizations for Model Checking

In Figure Figure 6a, the residual histogram displays the distribution of residuals from the model. The concentrated frequency around zero suggests that the model effectively captures much of the variability in the data, with fewer extreme deviations, thereby indicating a reasonable level of fit.

Figure Figure 6b presents the Pareto k diagnostic values, which help assess the reliability of the estimated leave-one-out cross-validation (LOO) predictions. The majority of Pareto k values fall below the threshold of 0.7, indicating that the model is well-calibrated with only minimal issues in terms of influential observations.

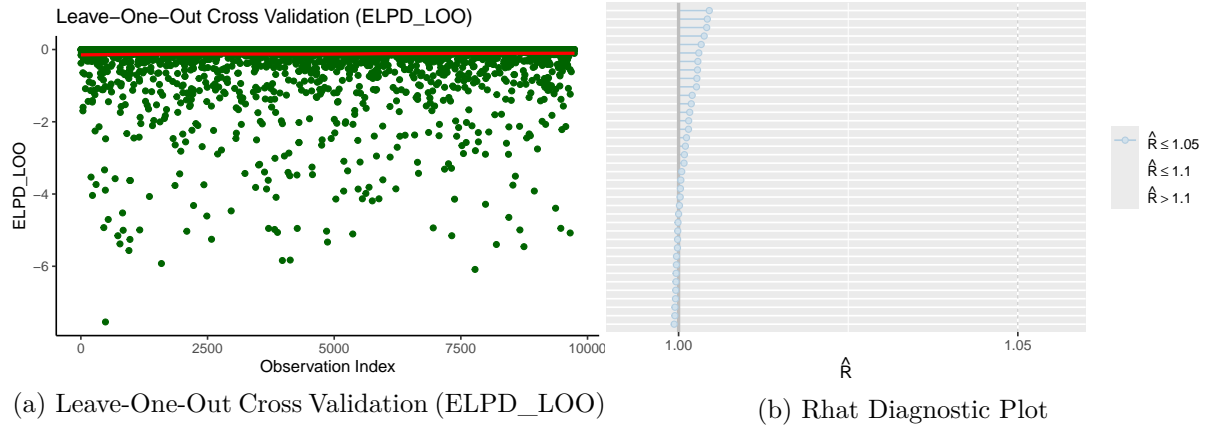


Figure 7: Comparing Diagnostic Visualizations for Model Checking

Figure Figure 7a is a scatter plot for leave-one-out cross-validation (ELPD_LOO) estimates. The consistency of the points around the red loess smooth line suggests that the model's predictions are largely consistent across different observations, thus reinforcing the model's robustness.

Figure Figure 7b is an Rhat plot, showing that all values of the Rhat diagnostic are close to 1. This indicates good mixing of the Markov Chain Monte Carlo (MCMC) chains and suggests that the model has likely converged appropriately, making the posterior estimates trustworthy.

References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. <https://tellingstorieswithdata.com>.
- Apache Arrow. 2021. *Arrow: Integration to 'Apache' 'Arrow'*. <https://CRAN.R-project.org/package=arrow>.
- Arel-Bundock, Vincent. 2022. “modelssummary: Data and Model Summaries in R.” *Journal of Statistical Software* 103 (1): 1–23. <https://doi.org/10.18637/jss.v103.i01>.
- Bolker, Ben, and David Robinson. 2024. *Broom.mixed: Tidying Methods for Mixed Models*. <https://CRAN.R-project.org/package=broom.mixed>.
- Fowler Jr, Floyd J, and Thomas W Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Vol. 18. Sage.
- Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- Gelman, Gabriel, Jonah Gabry, et al. 2021. *Bayesplot: Plotting for Bayesian Models*. <https://mc-stan.org/bayesplot>.
- Groves, Robert M, Floyd J Fowler Jr, Mick P Couper, James M Lepkowski, Eleanor Singer, and Roger Tourangeau. 2011. *Survey Methodology*. John Wiley & Sons.
- Kuhn, and Max. 2008. “Building Predictive Models in r Using the Caret Package.” *Journal of Statistical Software* 28 (5): 1–26. <https://doi.org/10.18637/jss.v028.i05>.
- Massey, Douglas S, and Roger Tourangeau. 2013. “Where Do We Go from Here? Nonresponse and Social Measurement.” *The ANNALS of the American Academy of Political and Social Science* 645 (1): 222–36.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Team, Stan Development. 2021. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. <https://mc-stan.org/rstanarm>.
- Vehtari, Aki, Jonah Gabry, Måns Magnusson, Yuling Yao, Paul-Christian Bürkner, Topi Paananen, and Andrew Gelman. 2024. “Loo: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models.” <https://mc-stan.org/loo/>.
- Wickham, Hadley. 2021. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley et al. 2021. *Tidyverse: Easily Install and Load the 'Tidyverse'*. <https://tidyverse.tidyverse.org>.
- Wilke, Claus O. 2024. *Ggbridges: Ridgeline Plots in 'Ggplot2'*. <https://CRAN.R-project.org/package=ggridges>.
- Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.
- Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.