WHAT THE PEOPLE THINK: ADVANCES IN PUBLIC OPINION MEASUREMENT USING ORDINAL VARIABLES

By

Simon Heuberger

Submitted to the

Faculty of the School of Public Affairs

of American University

in Partial Fulfillment of

the Requirements for the Degree

of Doctor of Philosophy

In

Government

Chair:

Professor Jeff Gill

Professor Ryan T. Moore

Professor Elizabeth Suhay

Professor R. Michael Alvarez

Dean of the College

Date

2020

American University

Washington, D.C. 20016

WHAT THE PEOPLE THINK: ADVANCES IN PUBLIC OPINION MEASUREMENT USING ORDINAL VARIABLES

By

Simon Heuberger

Submitted to the

Faculty of the School of Public Affairs

of American University

in Partial Fulfillment of

the Requirements for the Degree

of Doctor of Philosophy

In

Government

Chair: ______________________________

Professor Jeff Gill

____

Professor Ryan T. Moore

____

Professor Elizabeth Suhay

____

Professor R. Michael Alvarez

____

Dean of the College

____

Date

2020

American University

Washington, D.C. 20016

WHAT THE PEOPLE THINK: ADVANCES IN PUBLIC OPINION MEASUREMENT USING ORDINAL VARIABLES

by
Simon Heuberger

ABSTRACT

Surveys are a central part of political science. Without surveys, we would not know what people think about political issues. Survey experiments further enable us to test how people react to given treatments. Surveys and survey experiments are only as good as the measurements and analytical techniques we as researchers employ, though. For one particular survey variable called ordinal variables, some of our current measurements and techniques are insufficient. Ordinal variables consist of ordered categories where the spacing between each category is uneven and not known. Ordinal variables are highly important because the most important predictor of political behavior is an ordinal variable: Education. Ordinal example categories for education could be "Some High School", "High School Graduate", and "Bachelor's Degree". The literature currently often does not take the special nature of ordinal variables, i.e. their uneven spacing, into account. This could misrepresent the data and potentially distort survey results. It is important that we measure and use education and other ordinal variables correctly. My dissertation develops two methods to do so and applies them in original survey research. Chapter II develops a new method to improve the use of ordinal variables in the assignment of treatment in survey experiments. Chapter III develops a new method to treat missing survey data with ordinal variables. Chapter IV applies both methods in an online survey experiment on political framing.

ACKNOWLEDGEMENTS

I want to thank Fendi and Muesli, the two greatest creatures on the planet. And of course Waldemar Burgo.

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Some general text about surveys and survey experiments. Basically the abstract, but longer, more elaborate.

1.1 Ordinal Variables

Ordinal variables matter in surveys. One of the most important ordinal variables in political science surveys is education. It is widely established that education represents one of the major driving forces behind public opinion and political behavior, such as turnout or donations, in the U.S. (Abramowitz, 2010; Dawood, 2015; Druckman, Peterson, & Slothuus, 2013; Fiorina & Abrams, 2009; Fiorina, Abrams, & Pope, 2011; King, 1997; Leighley & Nagler, 2014). Ordinal variables are part of the larger framework of categorical variables. Categorical variables represent types of data which are commonly divided into three groups: Nominal, interval, and ordinal variables. Nominal variables are categorical variables with two or more categories that are not intrinsically ordered. Examples include gender (Female, Male, Transgender etc.), race (African-American, White, Hispanic etc.), and party ID (Democrat, Republican, Independent) where the categories cannot be ordered sensibly into highest or lowest. Interval variables are ordered categorical variables with evenly spaced values. Examples include income ($20,000, $40,000, $60,000, $80,000 etc.), where the distance between $20,000 and $40,000 is the same as the distance between \$60,000 and \$80,000. Ordinal variables are ordered categorical variables where the spacing between values is not the same. Examples include education (Elementary school, Some high school, High school graduate etc.) where the distance between "Elementary school" and "Some high school" is likely different than the distance between "High school graduate" and "Some college". Each subsequent category has quantitatively more education than the previous, but the exact measure of the distance between the categories is unclear.

For statistical analysis, the categories of nominal variables are often turned into binary variables. This manipulation does not impose any unnatural ordering onto the variable and thus does not require any theoretical assumptions. Interval variables are often made numeric, which is statistically sound. It makes sense to assign numeric values such as 1, 2, 3, and 4 to income categories of \$20,000, \$40,000, \$60,000, and \$80,000. The distance between each of these categories is identical between any adjacent pair and thus translates perfectly into the numeric values with equally identical distances. The distance between \$20,000 and \$40,000 is the same as the distance between 1 and 2. Ordinal variables are also often made numeric for analytic purposes. This is problematic because of their unevenly spaced categories. If the education categories "Elementary school", "Some high school", and "High school graduate" were turned into the numeric values 1, 2, and 3, we would wrongly assume that the distances between the education categories correspond to these evenly spaced values. Do the numbers 1 to 3 really represent the distances between the categories? Perhaps the true spacing between some of the categories is so narrow they should not even be separate categories at all. We cannot answer this by making an arbitrary assumption that is not justified by the data. Alternatively, if "Elementary school", "Some high school", and "High school graduate" were turned into three separate dummy variables, we would wrongly assume that there is no ordering to these values. In both cases, important information would be lost, which could lead to a large degree of distortion (O'Brien, 1981). To truly use the ordinal nature of a variable, we need to use both its quantitative and its inherent unevenly spaced ordered aspects to make a more underlying description of the data possible (Agresti, 2010). To fill this gap, I borrow from machine learning, which has close connections to problems of causal inference (Grimmer, 2015), and propose an ordered probit model that estimates an ordinal variable's underlying latent continuous structure and is trained on external data.

1.2 Ordered Probit Approach

Many approaches in the literature on the analysis of ordinal variables incorporate the distribution of the variable categories (Agresti, 1996). The most promising suggestions focus on natural extensions of probit and logit models (Winship & Mare, 1984) by assigning scores to be estimated from the data (Agresti, 1990) and quantifying each non-quantitative variable according to the empirical distributions of the variable, assuming the presence of a continuous underlying variable for each ordinal indicator (Lucadamoa & Amenta, 2014). In fact, Agresti (2010) states "that the type of ordinal method used is not that crucial" but that the "results may be quite different, however, from those obtained using methods that treat all the variables as nominal" (p. 3). The same applies to methods which treat ordinal variables as interval (Gertheiss & Tutz, 2008). This suggests that a probit or logit model is suitable to uncover the latent continuous variable underlying an ordinal variable, thus using the ordinal information provided and respecting uneven distances. In the literature, this approach is focused exclusively on the analysis of ordinal variables as a response variable. I propose an ordered probit model that applies to ordinal variables as predictive variables.

Let there be $\boldsymbol{X}$, an $n \times k$ matrix of explanatory variables. Let further $\boldsymbol{Y}$ be observed on the ordered categories $\boldsymbol{Y}_i \in [1, \ldots, k]$, for $i = 1, \ldots n$, and let $\boldsymbol{Y}$ be assumed to be produced by the unobserved latent continuous variable $\boldsymbol{Y}^*$. $\boldsymbol{Y}^*$ is continuous on

$\Re$ from $-\infty$ to ∞. The 'response mechanism' for the r^{th} category is $Y = r \Longleftrightarrow \theta_{r-1} < Y^* < \theta_r$. This requires there to be thresholds on $\Re$: $Y_i^*: \ \theta_0 \underset{c=1}{\longleftrightarrow} \theta_1 \underset{c=2}{\longleftrightarrow} \theta_2 \underset{c=3}{\longleftrightarrow} \theta_3 \ldots \theta_{C-1} \underset{c=C}{\longleftrightarrow} \theta_C$. The vector of (unseen) utilities across individuals in the sample, Y^*, is determined by a linear model of explanatory variables: $\boldsymbol{Y^*} = \boldsymbol{X\beta} + \mathbf{E}$, where $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_p]$ does not depend on the θ_j and $\mathbf{E} \sim F_{\mathbf{E}}$. For the observed vector $\boldsymbol{Y}$,

$$\begin{aligned} p(\boldsymbol{Y} \leq r|\boldsymbol{X}) &= p(\boldsymbol{Y^*} \leq \theta_r) = p(\boldsymbol{X\beta} + \mathbf{E} \leq \theta_r) \\ &= p(\mathbf{E} \leq \theta_r + \boldsymbol{X\beta}) = F_{\mathbf{E}}(\theta_r + \boldsymbol{X\beta}) \end{aligned}$$

is called the cumulative model because $p(\boldsymbol{Y} \leq \theta_r|\boldsymbol{X}) = p(\boldsymbol{Y} = 1|\boldsymbol{X}) + p(\boldsymbol{Y} = 2|\boldsymbol{X}) + \ldots + p(\boldsymbol{Y} = r|\boldsymbol{X})$. A logistic distributional assumption on the errors produces the ordered logit specification: $F_{\mathbf{E}}(\theta_r - \boldsymbol{X'\beta}) = P(\boldsymbol{Y} \leq r|\boldsymbol{X}) = [1 + \exp(-\theta_r - \boldsymbol{X'\beta})]^{-1}$. The likelihood function is: $L(\boldsymbol{\beta}, \boldsymbol{\theta}|\boldsymbol{X}, \boldsymbol{Y}) = \prod_{i=1}^{n} \prod_{j=1}^{C-1} [\Lambda(\theta_j + \boldsymbol{X}_i'\boldsymbol{\beta}) - \Lambda(\theta_{j-1} + \boldsymbol{X}_i'\boldsymbol{\beta})]^{z_{ij}}$ where $z_{ij} = 1$ if the i^{th} case is in the j^{th} category, and $z_{ij} = 0$ otherwise. The thresholds on $\Re$ partition the variable into regions corresponding to the ordinal categories. The linear model, Y^*, bins the observations between these thresholds according to the linear predictors.

Practically, we need to estimate a linear combination of meaningful covariates as predictors and an ordinal variable as the dependent variable. We then train this model on externally and internally valid data. This estimates cutoff thresholds between the ordinal categories and bins data cases according to the linear predictors. The binned cases determine which variable categories make sense, given the underlying latent continuous variable. We then replace the original categories with these re-estimated categories and conduct the statistical analysis of interest. In `R`, the ordered probit approach can be implemented with the `polr` function from the `MASS` package (Ripley et al., 2020).

1.3 Blocking

How the ordered probit approach for ordinal variables is used to 'prepare' the data before blocking.

1.4 Missing Data

How the ordered probit approach is used to impute missing data with ordinal variables.

1.5 Application

How the the ordered probit blocking and ordered probit missing data methods are applied in a framing survey experiment.

CHAPTER 2

PRECISION IN SURVEY EXPERIMENTS – A NEW METHOD TO IMPROVE BLOCKING ON ORDINAL VARIABLES

2.1 Introduction

Survey experiments collect background information and attempt to uncover treatment effects on public opinion and/or behavior. In order to identify such potential effects, the treatment groups need to be comparable. All treatment groups need to look the same in every measure, i.e. they must be balanced. This can be achieved through random assignment of participants to treatment groups. Randomization, i.e. flipping a coin to decide which treatment group a participant is assigned to, probabilistically results in balance based on the Law of Large Numbers (Urdan, 2010). For small samples, however, it can lead to serious imbalance. It can easily be that the treatment groups will not look the same. This can leave experimental results in statistically murky waters (Fox, 2015; Imai, 2018; King, Keohane, & Verba, 1994). In survey experiments, the overall sample size is often split across several treatment groups, which can exacerbate the problem. Chong & Druckman (2007), for instance, split 869 participants in a framing experiment on urban growth over 17 treatment groups, which leads to an average of just over 50 participants per group. Randomization is unlikely to lead to balanced treatment groups of this size. Researchers need to employ statistical methods to obtain balanced groups here. Blocking, i.e. arranging participants in groups that are equal in terms of participants' covariates and using random allocation within these groups, can alleviate such worries.

Blocking depends on covariates. In political science, many covariates with high predictive power are categorical variables, i.e. variables where the data can be divided into groups. These include interval (ordered and evenly spaced, e.g. `Income`) and ordinal (ordered and unevenly spaced, e.g. `Education`) variables. To block, these variables are often made numeric, e.g. by assigning the numbers 1-4 to the variable categories. This is acceptable for interval variables as the evenly spaced numbers correspond to the evenly spaced categories. For ordinal variables, however, this can be problematic. An arbitrary evenly spaced string of numbers does not correspond to the unevenly spaced ordinal categories and may misrepresent the data. I propose an ordered probit threshold approach to circumvent this problem: This approach estimates an assumed underlying latent continuous structure underneath ordinal variables whose data-driven categories can then be used for blocking. By training a linear model on meaningful data, it creates numerical thresholds which partition the variable into regions corresponding to the ordinal categories and bins the observations between these thresholds according to the explanatory variables. These binned cases determine which of the original categories make sense given the underlying latent continuous structure. The result is a data-based and non-arbitrary re-estimated set of variable categories. Because of their data-driven estimation, these categories can be safely used for blocking. This approach allows researchers to block on ordinal variables in survey experiments without making unwarranted assumptions in terms of arbitrary numeric values whilst fully utilizing the ordinal information provided and respecting uneven spaces.

The following sections provide a background on survey experiments and blocking, describe the key aspects of ordinal variables, and outline my proposed ordered probit approach. I then demonstrate the benefits and implications of this approach with external survey data and original data from an online survey experiment. Since there currently is no available tool to block in online survey experiments, I create my own survey environment in `shiny`, which will be described in more detail below.

2.2 Theory

2.2.1 Preliminary Notations on Survey Experiments

The simplest of survey experiments has two potential outcomes for participants i, y_{1i} and y_{0i}, with 1 denoting the treatment and 0 referring to the control. Consider a simplified version of a famous survey experiment by Tversky & Kahneman (1981), where researchers want to test the effect of the mortality format on participants' choices. They provide participants with the following scenario:

> Imagine that the US is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. A program to combat the disease has been proposed. Assume that the exact scientific estimates of the consequences of the program are as follows...

Participants in the control group receive the program description in survival format:

> If the program is adopted, 200 out of 600 people will live.

Participants in the treatment group receive the program description in mortality format:

> If the program is adopted, 400 out of 600 people will die.

All participants are subsequently asked whether they support or oppose the program. The treatment effect for each individual participant i is given by $y_{1i} - y_{0i}$. If both groups of participants look the same regarding their covariates (`Age`, `Education`, `Income` etc.), a comparison of the groups' average support reveals the Average Treatment Effect (ATE) across all participants, $\mathbf{E}[\delta] = \mathbf{E}[y_{1i} - y_{0i}]$. A central characteristic of such a comparison is the fundamental problem of causal inference (Holland, 1986; Rubin, 1974): We are unable to observe both potential outcomes for the same participant at once. In our case, we cannot observe how much participant A supports the program if given the survival format whilst also observing how much the same participant A would have supported the program if given the mortality format. If we could, it would be simple to calculate the true average treatment effect, $\mathbf{E}[\delta] = \mathbf{E}[y_{1i}|T = 1] - \mathbf{E}[y_{0i}|T = 0]$, with $T = 0$ denoting the control and $T = 1$ the treatment group. Since the true average treatment effect is unobservable, we need to use statistical means to assess the counterfactuals. This can be done by balancing the treatment and control groups. If both groups of participants look the same in every measure, we can use the participants who received the mortality format (treatment) to estimate what would have happened to the participants who did not receive the mortality format (control). The crucial aspect is whether the two groups do indeed look the same in terms of participants' covariates. The potential outcome of the control needs to mirror what would have happened in the case of treatment, and vice versa. There are two main means by which this may be achieved: Randomization and blocking.

2.2.2 Randomization

Randomization is equivalent to flipping a coin for each participant to be assigned to treatment or control. This chance procedure gives each participant an equal chance of being assigned to either group (or groups, in case of multiple treatment groups) (Lachin, 1988). Randomization increases covariate balance as the number of participants, n, increases (Imai, King, & Nall, 2009). The larger a researcher's sample, the better the resulting balance from randomization in expectation. Probabilistically, randomization enables the comparison of the average treatment effect to be unbiased, which allows the researcher to attribute any treatment effects to the treatment (King et al., 2007).

While randomization thus guarantees balance as the sample size reaches infinity, it often does not do so in the naturally finite sample sizes researchers actually work with. With huge samples, the Law of Large Numbers predicts that treatment groups selected through randomization will be balanced. With small samples, however, it is possible to get unlucky and end up with unbalanced groups (Imai, King, & Elizabeth A. Stuart, 2008). Blocking can help achieve balance in such scenarios (Epstein & King, 2002).

2.2.3 Blocking

Identical levels in terms of covariates across treatment groups represent the key aspect in experimental studies. In randomization, this is achieved by random chance. In blocking, this is achieved by combining covariate information about the participants with randomization. Specifically, participants are blocked into treatment groups that are similar to one another in terms of the their covariates before treatment is assigned. Their similarity is estimated with the Mahalanobis or Euclidian distance. Blocking is better suited to achieving balance in finite samples than randomization, as it "directly controls the estimation error due to differing levels of observed covariates in the treatment and control groups" (Moore, 2012, p. 463). This is particularly relevant with small samples and a high number of treatment groups, as the overall number of participants needs to be divided up. Figures 2.1 and 2.2 show this visually. A numeric discrete variable with levels 1 to 5 is randomized and blocked for different sample sizes and numbers of treatment groups. This is repeated 100 times for each sample size. Figure 2.1 shows the maximum distances between treatment groups across these repetitions for sample sizes up to 1,000 for two, three, five, and ten treatment groups. Blocking outperforms randomization in every scenario. The difference between the two methods is smallest for large samples and a small number of treatment groups. For $n = 998$ and two treatment groups, the largest

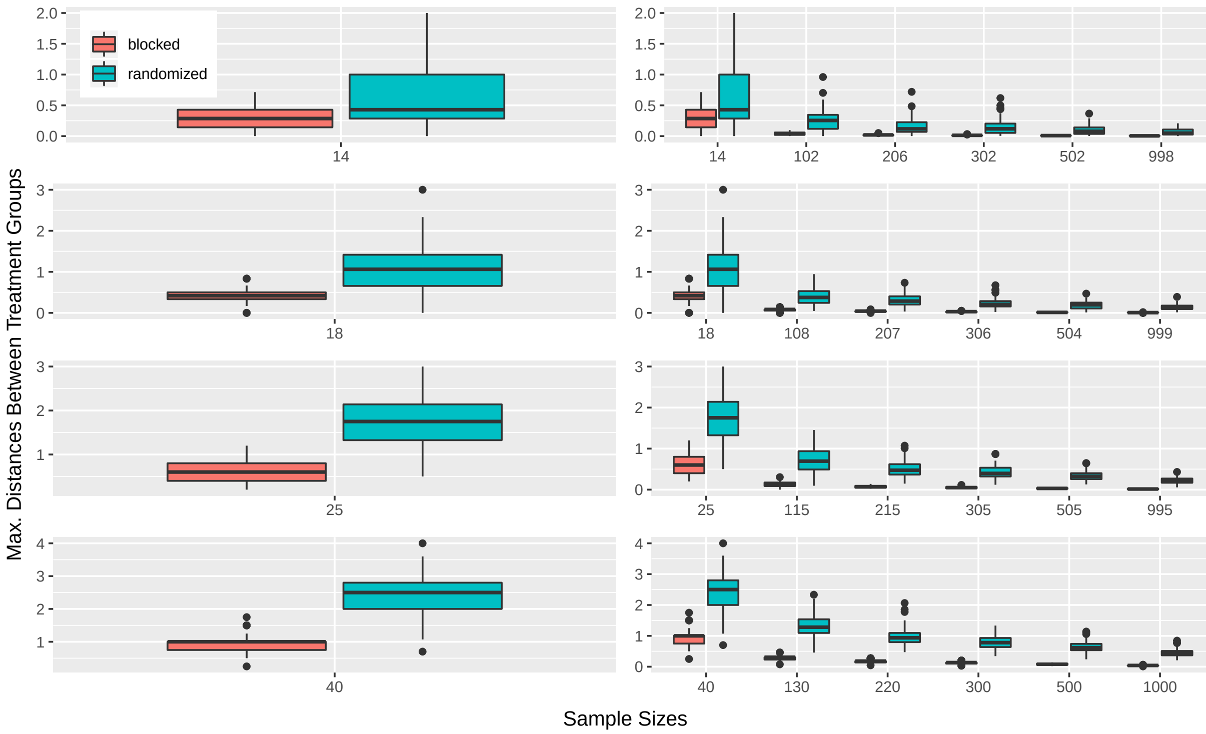

Figure 2.1. Distances between treatment group means in randomized and blocked data. Increasing sample size for 2 (top row), 3 (second row), 5 (third row), and 10 treatment groups (bottom row). Leftmost pair on right panel is exactly the pair on the left panel

distance between randomized treatment groups is 0.208, and the largest distance between blocked treatment groups is 0.01. For small samples and a large number of treatment groups, however, the difference is much starker. For $n = 40$ and ten treatment groups, the largest distance between randomized treatment groups is 4, and the largest distance between blocked treatment groups is 1.75. Figure 2.2 shows the distribution of these imbalances.

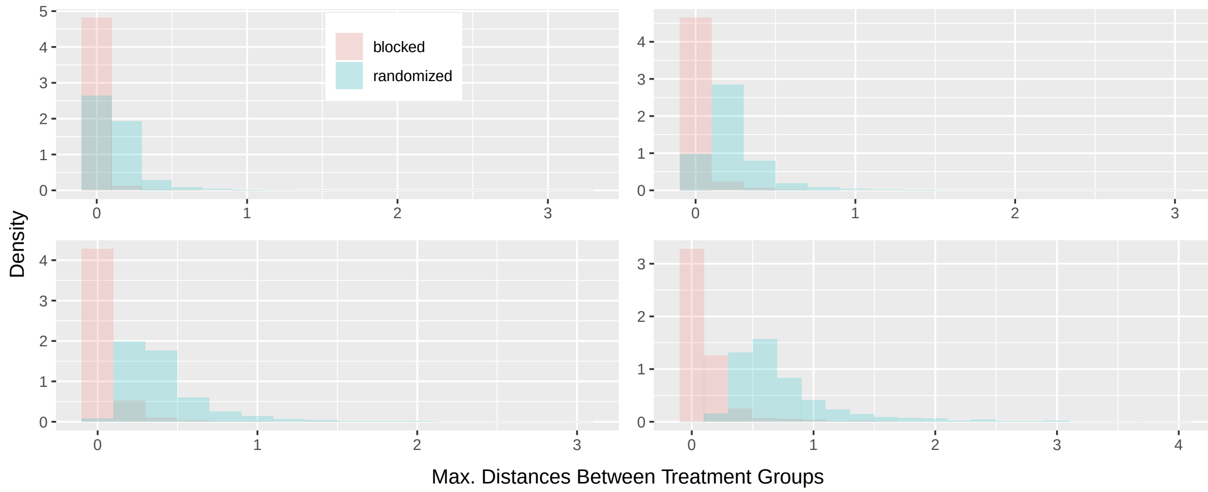

Figure 2.2. Distribution of treatment group differences in randomized and blocked data for 2 (top left), 3 (top right), 5 (bottom left), and 10 (bottom right) treatment groups

Blocking On The Go

In political science, researchers often have an already-collected data set in front of them. One example would be the American National Election Studies (ANES), a pre-existing survey database, which is often used to analyze voter turnout (see for instance Jackman & Spahn, 2018; Leighley & Nagler, 2014), among many others. This setup means all covariate information on all participants is known at the time of assignment, which makes blocking straight-forward. Oftentimes, however, the covariate information of all participants is not known at the time of assignment. This is the case, for instance, for online survey experiments, where each participant completes the survey at differing times. Participants 'trickle in' for treatment assignment as the experiment progresses. 'Traditional' blocking can not be used here, since it relies on covariate information about the entire sample. Instead, we need to block continuously as the experiment progresses, or block 'on the go'. This is called sequential blocking.

Sequential blocking in political science is based on covariate-adaptive randomization, which varies probabilities based on knowledge about previous participants and the current participant (Chow & Chang, 2007). Traditional covariate-adaptive approaches, such as the biased coin design (Efron, 1971) and minimization (Pocock & Simon, 1975), assign the incoming participant to the treatment group with the fewest participants with identical covariate information. This works for discrete covariates as the number of possible covariate levels is finite. For continuous covariates, the number of possible covariate levels rises exponentially. Participants are unlikely to look the same, and identical participants are rare. Blocking on continuous covariates is not possible with these traditional approaches (Eisele, 1995; Markaryan & Rosenberger, 2010; Rosenberger & Lachin, 2002). Moore & Moore (2013) develop a method to do so by exploiting relationships between the current participant's covariate profile and those of all previously assigned participants. They define the similarity between participants with the Mahalanobis distance (MD) between participants q and r with covariate vectors $\boldsymbol{x}_q$ and $\boldsymbol{x}_r$: $MD_{qr} = \sqrt{(\boldsymbol{x}_q - \boldsymbol{x}_r)' \widehat{\sum}^{-1} (\boldsymbol{x}_q - \boldsymbol{x}_r)}$. To aggregate pairwise similarity, they implement the mean, median, and trimmed mean of the pairwise MDs between the current participant and the participants in each treatment condition: Participants are indexed with treatment condition t using $r \in \{1, ..., R\}$. For each condition t, an average MD between the current participant, q, and the participants previously assigned, t. If the distance in terms of MD for the incoming participant is 2 in the control and 5 for the treatment condition, the incoming participant looks more similar to the control condition. To set the probability of assignment, Moore & Moore (2013) calculate the mean Mahalanobis distances for each incoming participant, q, for all treatment conditions, t, and sort the treatment conditions by these averages. Randomization is biased towards conditions with high scores. For each value of k, with $k \in \{2, 3, ..., 6\}$, the condition with the highest average MD is then assigned a probability k times larger than all other assignment probabilities.

Blocking is thus possible when all covariate information is known at the time of assignment and when this information 'trickles in' over time. Covariate information, however, is only one side of the coin. Researchers also need to take into consideration the characteristics of the variable to block on. Not all types of variables can and should be used the same way to be blocked on. Specifically, the current use of ordinal variables as blocking variables is somewhat problematic. I describe these problems and my proposed solution in section 1 above. The following section will apply this solution to differing types of data.

2.3 Data

One of the most respected and recognized externally and internally valid data sets are the American National Election Studies. I thus choose the following ordered probit model with the 2016 ANES data (the predictors are standard linear predictors in political science literature):

$$Education \sim Gender + Race + Age + Income + Occupation + PartyID$$

When trained on the 2016 ANES data, this ordinal probit model estimates the thresholds between each of the education categories shown in Table 2.1. The observations in the data are binned according to the estimated threshold coefficients, which in turn determines what education categories make sense, given the underlying latent continuous variable. Figure 2.3 shows the distribution of both the original and the model-estimated education categories. As we can see, all categories 'below' "High school graduate" and 'above' "Master's" are collapsed because they do not fit the data. The ordered probit model uses the ordinal information with unevenly spaced distances provided and returns categories that do fit the data. We can now use these estimated education categories as the basis for blocking. Assigning numeric values to the new categories is now justifiable because they are based on data-driven estimations. This allows us to block on numerical values with the Mahalanobis distance, which would not be possible without empirical

Table 2.1. Ordered Probit Threshold Estimates

Thresholds	Coefficients	Standard Errors	t-values
Up to 1st|1st-4th	−7.869	1.024	−7.681
1st-4th|5th-6th	−7.146	0.717	−9.965
5th-6th|7th-8th	−5.379	0.326	−16.515
7th-8th|9th	−4.671	0.253	−18.472
9th|10th	−3.920	0.206	−19.070
10th|11th	−3.468	0.188	−18.489
11th|12th	−2.984	0.174	−17.100
12th|HS grad	−2.511	0.166	−15.116
HS grad|Some college	−0.710	0.154	−4.607
Some college|Associate	0.384	0.154	2.500
Associate|Bachelor's	1.045	0.154	6.766
Bachelor's|Master's	2.478	0.160	15.538
Master's|Professional	4.099	0.177	23.144
Professional|Doctorate	4.838	0.197	24.589

justification. The following sections show that the new estimated categories significantly affect analyses and results.

2.3.1 Simulations

I conduct various simulations to compare the Ordered Probit Model and its resulting reestimated education categories with the original ANES categories.

2.3.2 Placebo Regression

We separately block the 2016 ANES on the original and the ordered probit education categories into two treatment groups. We then model the following OLS regression on an interval response variable, a feeling thermometer towards Donald Trump as the Republican presidential candidate:

$$Feel.Trump \sim Group + Dem + Rep + Income + Male + White + Black + Hispanic$$

`Group` indicates a placebo treatment, as no actual treatment is administered. In the absence of actual treatment, the difference between both treatment groups should thus

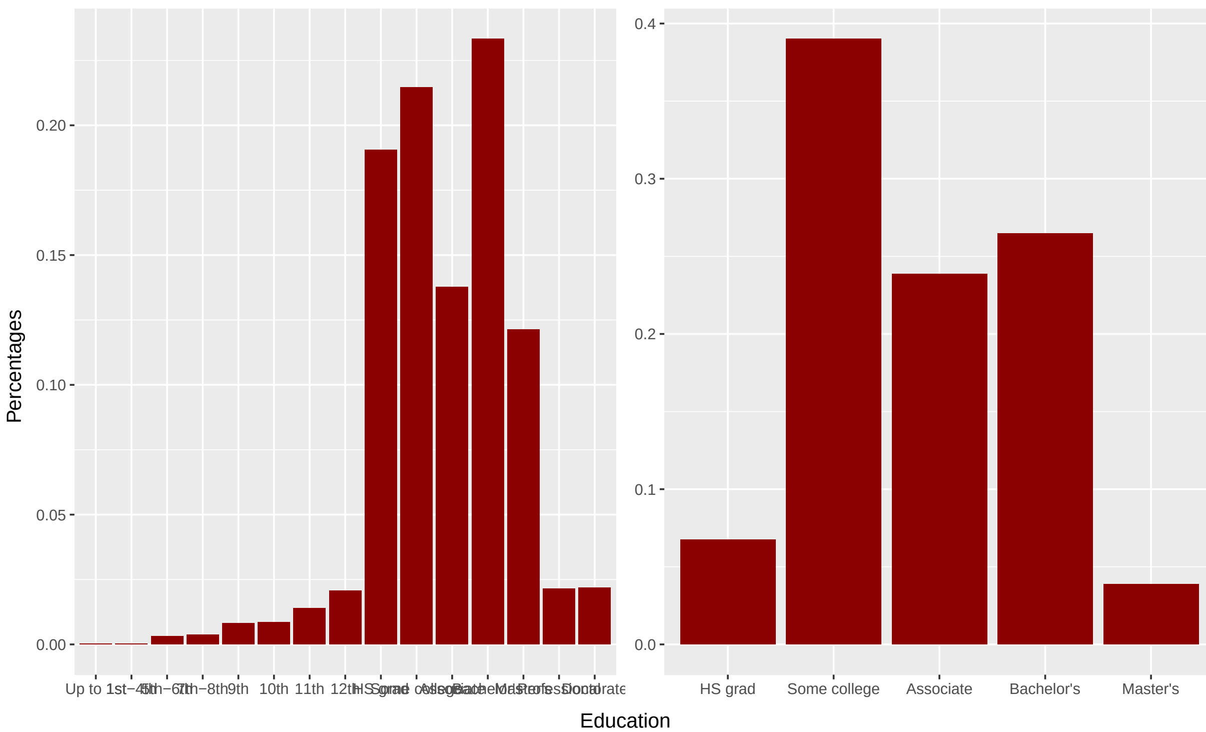

Figure 2.3. Distribution of Education Categories. Original 2016 ANES categories on the left, ordered probit estimated categories on the right

be zero.

2.3.3 Framing Survey

2.4 Results

2.4.1 Simulations

2.4.2 Placebo Regression

To test this, each blocking/regression process for each set of categories is repeated 1,000 times. The distribution of the placebo treatment indicator (`Group`) is visualized in Figure 2.4. Both distributions center around zero, as is the statistical expectation. Upon closer inspection, the ordered probit categories are closer to the true values than the original categories on both mean (0.019 v. -0.05) and median (-0.01 v. -0.07). This

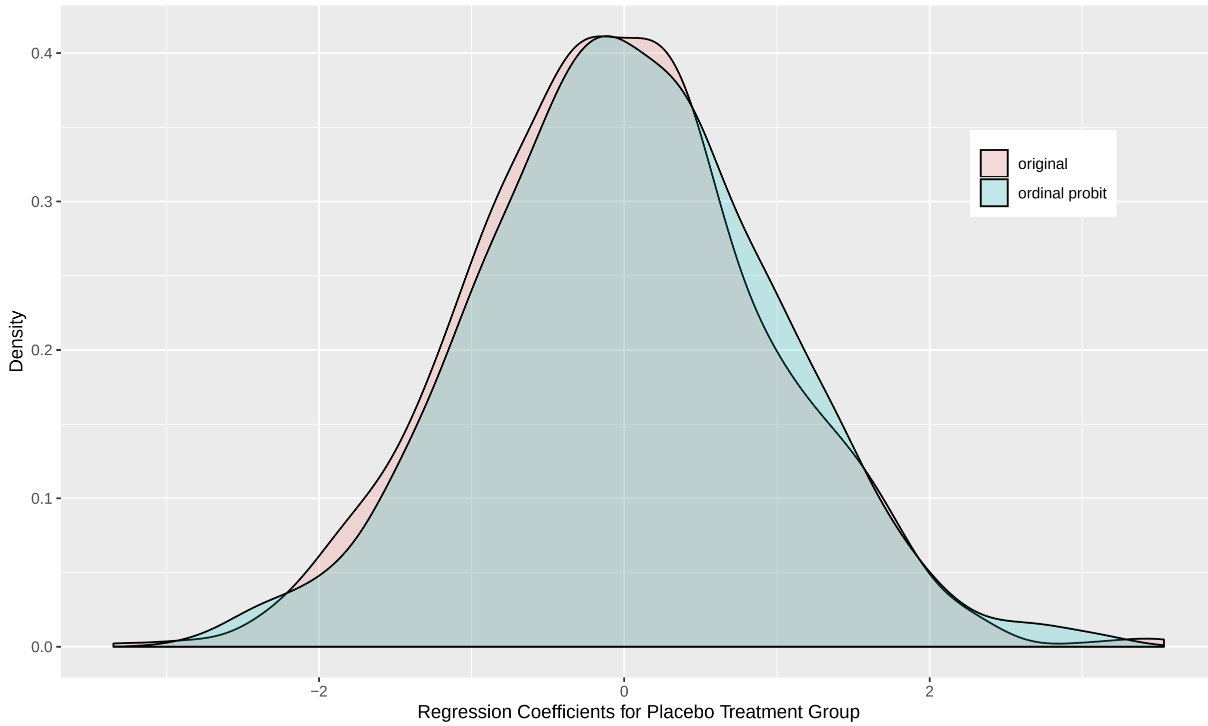

Figure 2.4. Distribution of placebo treatment coefficients by education model

indicates slightly superior performance by the ordered probit categories.

2.4.3 Framing Survey

Put text here

2.5 Conclusion

CHAPTER 3

QUALITY COMPARISON OF MAJOR MISSING DATA SOLUTIONS WITH A PROPOSED NEW METHOD FOR ORDINAL VARIABLES

3.1 Introduction

Missing data are ubiquitous in survey research (Allison, 2002; Raghunathan, 2016). Respondents frequently refuse to answer questions, select "Don't Know" as a response option, or drop out during the response collection process (Honaker & King, 2010). This poses a big problem for researchers because data can typically not be analyzed with statistical software if they contain missing values (Little & Rubin, 2002; Molenberghs & Kenward, 2007). Scholars have developed several ways to treat missing data. These can be roughly categorized into deletion, single imputation, and multiple imputation. Single imputation concerns the replacement of missing values with substitute estimates such as the mean, regression coefficients, or values from randomly drawn 'similar' respondents in the data. Multiple imputation estimates missing values from conditional distributions and subsequently averages the results into a single parameter of inference (Fay, 1996; King, Honaker, Joseph, & Scheve, 2001; Rubin, 1976).

Multiple imputation has become the state of the art in missing data management (Andridge & Little, 2010; Graham & Schafer, 1999; Schafer & Graham, 2002; White, Royston, & Wood, 2011) since it accounts for and incorporates uncertainty around the estimated imputations through repeated draws. This is missing from single imputation techniques which treat the single estimated replacement value as a de-facto data entry on par with observed values. This leads to biased standard errors and confidence intervals. Crucially, uncertainty is not reflected in the imputed values (Gill & Witko, 2013; Kroh, 2006). Similarly, listwise deletion has been shown to induce bias with political data (Bodner, 2008; Collins, Schafer, & Kam, 2001; Pigott, 2001; Rees & Duke-Williams, 1997; Reilly, 1993).

However, parametric multiple imputation as applied by the most popular imputation packages in `R` is not necessarily always the most suitable method for all types of variables. For discrete data, multiple hot deck imputation, a combination of the single imputation method hot decking (see section 3.2.3) and multiple imputation, proves more precise as it avoids the common multiple imputation technique of imputing discrete data on a continuous scale (Gill & Cranmer, 2012). This practically turns discrete variables into continuous variables which changes their nature and can result in non-observable and biased imputation values with artificially smaller standard errors (Fuller & Kim, 2005; Kim, 2004; Kim & Fuller, 2004). Multiple hot deck imputation on the other hand preserves the integrity of discrete data, does not change the size of standard errors, and produces more accurate imputations. It estimates affinity scores for each missing value to measure how similar a respondent with a missing value is to another respondent across all variables except the missing one.

However, multiple hot deck imputation does not account for ordinal variables as its affinity score algorithm assumes even distances between categories in discrete data. This assumption does not apply for ordinal variables. I propose a method designed to impute discrete missing data specifically from ordinal variables. Because of the success of multiple hot deck imputation in its applicability to missing data with discrete variables with a small number of categories (Gill & Cranmer, 2012), this method is based on multiple hot deck imputation and adapted to account for the specific circumstances of ordinal variables. Based on the ordered probit model approach described in section 1.2, it applies a scaled solution with newly estimated numeric thresholds from an assumed underlying latent continuous variable to measure the distances between the categories and calculate affinity scores.

The following is an outline of the remainder of this chapter: Sections 3.2.1 to 3.2.3 will discuss the theory behind missing data mechanisms and provide an overview of listwise deletion and the most common single imputation methods. Section 3.2.4 explains the basics of multiple imputation and outlines major `R` packages implementing it. These include `Amelia`, `mice`, `hot.deck` (implementing multiple hot deck imputation), and my self-penned multiple hot deck imputation function for ordinal variables, `hd.ord`. Since simple imputation is widely condemned as a general imputation method, my focus lies on multiple imputation. Section 3.3 describes the survey data these packages/function are tested on: A selection of the 2016 ANES, a subset of the 2016 CCES, and a randomized experiment on moral framing I ran on MTurk in 2017. Each dataset is complete and randomly amputed to insert missing data. Each of the four `R` packages/functions is applied to each dataset and tested for accuracy and speed for different types of missingness (MAR, MNAR) and variables (binary, ordinal, interval). Section 3.4 shows the results and assesses the usefulness of my proposed new method. Finally, section 3.5 will provide concluding remarks.

3.2 Theory

3.2.1 Missing Data Mechanisms

Let there be Y, an $n \times v$ matrix with data on n respondents for v variables. Let there also be the response indicator R as an $n \times v$ matrix with values of 0 or 1. Let their respective elements be denoted by y_{ij} and r_{ij}, with $i = 1, ..., n$ and $j = 1, ..., v$. If y_{ij} is observed, $r_{ij} = 1$. If y_{ij} is missing, $r_{ij} = 0$. All elements where $r_{ij} = 0$ make up the missing data, Y_{miss}. All elements where $r_{ij} = 1$ make up the observed data, Y_{obs}. Together, Y_{obs} and Y_{miss} form the complete data Y. Missing data can then generally be described by prob$(R = 0|Y_{obs}, Y_{miss}, \theta)$, i.e. the probability of missing data depends on the observed data, the missing data, and a vector of unknown parameters. Depending on the mechanism by which data is missing, this expression can be further simplified.

Data can be missing by three basic mechanisms: It can be missing completely at random (MCAR), missing at random (MAR), nor missing not at random (MNAR). It is not possible to test respective data on the mechanism underlying its missingness. The onus here is on the researcher to make substantive assumptions based on the nature of the data at hand – in other words: Researchers need to make assumptions about how the data came to be missing in the first place.

The simplest and easiest case is missing completely at random (MCAR). Under the MCAR assumption, data is missing at the repeated flip of a coin. There is no process guiding the missingness; it is inserted truly at random. In statistical terms, this means both the unobserved and the missing data independently from each other form a random sample of the population and have the same underlying distribution. Missing data would not pose such a serious problem, if it routinely and ubiquitously occurred MCAR. Unfortunately, however, that is not the case, even when restricted just to survey data: Survey respondents are known to withhold sensitive data. This can stretch from the unwillingness to divulge information deemed to be personally private (income, sexual orientation) to the refusal to answer sensitive questions out of fear of political or social repercussions in the community (union membership, support for polarizing political candidates). These types of missing data occur systematically, i.e. answers have not been refused randomly across the whole range of income levels but are missing only for respondents with very high or very low income. Similarly, only answers criticizing the authorities are missing in surveys in non-democratic states, while the state-loyal responses are present. Notationally, the general missing data expression can be simplified under the MCAR assumption to prob$(R = 0|\theta)$, i.e. the generic probability of missing data, independent of the data themselves and only dependent on θ.

As a result, a MCAR assumption in political science surveys is rare and requires tremendous backing up. It is more commonly assumed among researchers that missing data are missing at random (MAR). MAR means missing data are related to the observed but not the unobserved data. In practical terms, this for instance means missing data on income can be related to observed data on education or occupation. Here, the missing data are not a random sample of the entire data. MAR transforms the general missing data expression to prob$(R = 0|Y_{obs}, \theta)$, i.e. the chances of missing data depend on the observed data and θ.

Finally, missing data can be missing not at random (MNAR)[1] This is the case when missing data are related to unknown and/or unobserved parameters. Continuing the example of missing data on income, under MNAR we do not observe data on education or occupation that can be used to fill the missing income data slot. In the case of data MNAR, the general missing data expression remains unchanged, prob$(R = 0|Y_{obs}, Y_{miss}, \theta)$, i.e. the

[1]Data MNAR is also sometimes called 'non-ignorable' (see for instance Gill & Cranmer (2012) and Allison (2002)).

missingness of data depends on the observed and the missing data.

As mentioned above, it is not possible to test whether missing data is MCAR, MAR, or MNAR. The assessment of the missing data mechanism underlying any respective data comes from researchers and their understanding of the data generating process. The statistical methods to address missing data can be broadly categorized into deletion, imputation, and multiple imputation. Sections 3.2.2 to 3.2.4 outline these below. Given the focus of this chapter, special attention lies on section 3.2.4.

3.2.2 Listwise Deletion

One of the most common methods of handling missing data in quantitative political science is listwise deletion. This involves the removal of any incomplete observations, thereby reducing the sample size. The resulting sample is then ready for analysis. Whilst almost unbeatable in its simplicity and speed, this method often induces bias, depending on how much data are missing, how non-random the pattern of missingness is, and other aspects (Allison, 2002; King et al., 2001; Little & Rubin, 2002).

In the case of data MCAR, listwise deletion is not biased, as it removes a random sample of the population (King et al., 2001; Schafer & Graham, 2002). When missing data are MAR, listwise deletion is likely to be biased, since the observed data is tilted towards respondents with characteristics that make them more likely to respond (Collins et al., 2001; Graham, Hofer, Donaldson, MacKinnon, & Schafer, 1997). The potential for bias increases with data MNAR (Collins et al., 2001; Diggle & Kenward, 1994; Glynn, Laird, & Rubin, 1993; Robins, Rotznitzky, & Scharfstein, 1998).

While the bias inserted by listwise deletion in each individual data analysis may not necessarily be drastic, studies have shown that it can be so severe as to alter substantive conclusions (Brown, 1994; Graham, Hofer, & MacKinnon, 1996; Honaker & King, 2010; Wothke, 2000). Even if that were not or only rarely the case and most data were MCAR, reducing the sample size is generally never a recommended approach as, among other aspects, standard errors from regression models are inflated. As King et al. (2001) put it, the result of listwise deletion "is a loss of valuable information at best and severe selection bias at worst" (p. 49). In `R`, listwise deletion is done very quickly, for instance with the base function `na.omit()`.

3.2.3 Single Imputation

Single imputation means replacing missing data with substituted values, i.e. the structural opposite of deletion. Imputation requires some method of creating a predictive distribution, based on the observed data, from which value substitutions are picked. Single imputation, regardless of its exact nature, is not recommended to impute missing data since, similar to deletion, it biases standard errors and confidence intervals (Honaker & King, 2010). Crucially, uncertainty is not reflected in the imputed values (Little & Rubin, 2002). The following subsections are mere selections of the most common single imputation methods and make no claim of completeness. Since single imputation is widely condemned as a general imputation method and as my focus lies on multiple imputation, they are also brief.

Mean

In mean imputation, the mean is used to substitute missing values within cells. Mean imputation, sometimes also called unconditional mean imputation, means replacing missing values with the mean of the observed values, so $Y_{miss} = \overline{Y_{obs}}$. While it does not change the mean of the sample, this method distorts the empirical distribution of Y, which in turn produces biased estimates of any non-linear quantities such as variances and covariances (Haitovsky, 1968). It is also bound to be inaccurate in most cases, since few values generally fall exactly on the mean, and can be nonsensical for discrete variables (Efron, 1994).

Mean imputation can be done in many ways in `R`, for instance with the `impute()`

function in the `Hmisc` package or by setting `method = "mean"` in the `mice()` function in `mice`. It is also very simple to execute it in base `R`, e.g. for a sample data frame `df` and the numeric column `x`:

```
df <- transform(df, x = ifelse(is.na(x), mean(x, na.rm=TRUE), x))
```

Regression

Imputation by regression, sometimes also called conditional mean imputation, imputes missing values conditional on observed values. Researchers predict observed variable values based on other variables, while the fitted values from the regression model are then used to impute variable values where they are missing. Assume an independent variable in a multiple regression model. Assume further that x contains missing values, x_{miss}, and observed values, x_{obs}. We regress x_{obs} on the other independent variables and use the estimated equation to generate predicted values for x_{miss}, x_{pred}. x_{pred} then replace x_{miss}, completing the dataset. While more accurate than mean imputation, particularly for large samples with data MCAR, regression imputation nonetheless suffers from the same flaw that accompanies all single imputation approaches: Uncertainty is not reflected in the imputed values (Horton & Kleinman, 2007).

Differing variations of imputation by regression exist in `R`, such as the `aregImpute()` function in the `Hmisc` package, which performs additive regression, and setting `method = "norm.predict"` in `mice()` to conduct linear regression. The `predict()` function in base `R` also applies linear regression imputation.

Hot Decking

Hot decking imputation was developed in the 1970s and replaces missing values with values from similar respondents in the sample (Ernst, 1978; Ford, 1983). It is called 'hot decking' as a reference to taking draws from a deck of matching computer punch-cards. The deck was 'hot' since it was currently being processed, as opposed to pre-collected or 'cold' data (Andridge & Little, 2010; Little & Rubin, 2002). In the most general version, researchers select all respondents that are 'similar' to a respondent with missing data and randomly draw one of those respondents (with replacement) to fill in the missing value. The respondent with the initially missing value is termed the *recipient*, while the 'similar' respondent is called the *donor*. Variations of the method include hot decking within adjustment cells, by nearest neighbor, and sequentially ordered by a covariate (Cox, 1980; David, Little, Samuhel, & Triest, 1986; Kaiser, 1983; Kalton & Kish, 1981; Rockwell, 1975).

Contrary to mean or regression imputation, hot decking imputation preserves the integrity of the data, i.e. only actually observed values are used to fill in missing slots (Bailar & Bailar, 1997). In both other single imputation methods, it is possible and sometimes even likely that missing values are replaced by values not found amongst the observed values. Also contrary to regression imputation, hot decking does not require a fitted model and is thus less vulnerable to model misspecification. However, hot decking does necessitate the existence of at least some donors for a respondent at every variable value that is missing. With a lot of missing data and few 'similar' matches, the accuracy of hot decking greatly decreases (Young, Weckman, & Holland, 2011). Hot decking works best for discrete data as continuous data are very unlikely to be matched or 'similar', though the definition of what might constitute a 'similar' respondent is somewhat subjective (Marker, Judkins, & Winglee, 2002). As is the case with all single imputation methods, uncertainty is not reflected in the imputed values. Selecting the initial sample of 'similar' respondents and the subsequent random sampling from that subsampling is treated as factual responses, which leads to smaller standard errors and confidence intervals than statistically valid (Little & Rubin, 2002).

To my knowledge, there is currently no `R` package that applies hot decking. Nonetheless, forms of hot decking are still in use by some government statistics agencies such as the National Center for Education Statistics (for parts of the Current Population Survey) or the U.S. Bureau of the Census (NCES, 2002; USBC, 2002).

3.2.4 Multiple Imputation

Multiple imputation was invented by Rubin in the 1970s to account for the absence of uncertainty in single imputation methods and allow more accurate standard error estimates. It fills missing values with a predictive model that includes observed data and prior knowledge (Honaker & King, 2010). Over the time of its development, it has become the dominant sophisticated strategy for handling missing data (Dempster, Laird, & Rubin, 1977; Glynn et al., 1993; Heitjan & Rubin, 1991; Little & Rubin, 2002; Rubin, 1976, 1987, 1996; Rubin & Schenker, 1986). Multiple imputation involves three general steps:

(1) Impute a dataset with missing values m times. This results in i complete datasets (with $i = 1, ..., m$)

(2) Analyze each of the i complete datasets

(3) Combine the results from each of the i analysis results into one collective result

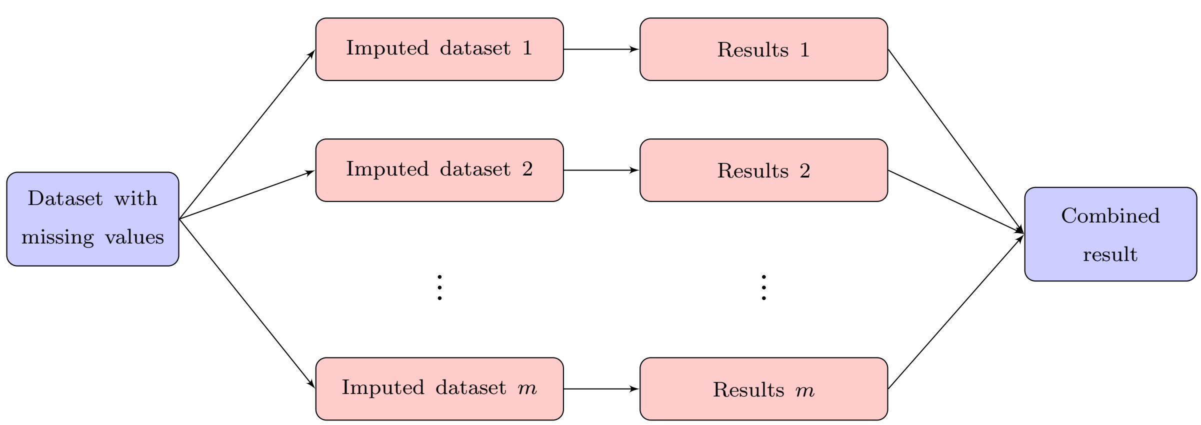

Figure 3.1. Multiple Imputation Workflow

Figure 3.1 provides a graphical overview of this workflow. Each missing value is imputed m times from a conditional distribution using other present values for the respective value to create i imputed complete datasets. The chosen statistical analysis τ, for instance a regression model, is applied to each of these i datasets, resulting in τ_i, with $i = 1, ..., m$. Finally, averaging τ_i gives us the single estimate, $\overline{\tau}$. Together, this is expressed as:

$$\overline{\tau} = \frac{1}{m}\sum_{i=1}^{m}\tau_i$$

Following Rubin (1987), the total variance of $\overline{\tau}$, Var_T consists of the mean variance of τ_i within each dataset i, $\overline{Var_W}$, and the sample variance of τ across all datasets, Var_A:

$$\overline{Var_W} = \frac{1}{m}\sum_{i=1}^{m} SE(\tau_i)^2$$
$$Var_A = \sum_{i=1}^{m}\frac{(\tau_i - \overline{\tau})^2}{m-1}$$
$$Var_T = \overline{Var_W} + Var_A$$

Multiplied by a factor correcting for small numbers of m (as $m < \infty$), Var_T is adjusted to:

$$Var_T = \overline{Var_W} + Var_A(1 + \frac{1}{m})$$

Each imputed complete dataset is identical to all other imputed complete datasets, with the exception of the imputed value. The imputed values for a missing value differ with each imputation of M in order to reflect uncertainty levels. The 'multiple' part of the imputation is a crucial aspect here since each imputation run will produce slightly different parameter estimates. Imputing multiple times and then averaging the results creates variability which adjusts the standard errors upward (Kroh, 2006). This deliberate random variation included in a deterministic multiple imputation run removes the overconfidence from single imputation, where the standard error estimates are too low (Schafer & Graham, 2002). In a case where the utilized multiple imputation model predicts missing values well, variation across the imputed values is small. In other cases, variation may be larger, depending on the level of certainty we have about the missing value. Multiple imputation has been shown to produce consistent, asymptotically efficient and normal estimates for a variety of data MAR (Allison, 2002).

Choosing m, the number of imputations, is somewhat subjective. Originally, $m = 5$ was considered sufficient based on efficiency calculations (Rubin, 1987) and is still the default in most software packages. More recent discussions stress the need for an increase of m in order to estimate more nuanced standard errors. Various approaches continue to coexist, such as focusing on the parameter with the largest fraction of missing information (Kroh, 2006) or starting with $m = 5$ and gradually increasing it in subsequent runs (Raghunathan, 2016). The most common current practice appears to be to set m to the percentage of missing data, i.e. if 20 percent of data are missing, m is set to 20 (Bodner, 2008; White et al., 2011).

There are numerous ways to implement multiple imputation. Up until the late 1990s, this required considerable statistical knowledge and sophisticated methodological skills (see Honaker & King (2010) for an overview). The use of multiple imputation was thus limited to a rather specialized audience of statisticians and methodologists. Since then, numerous `R` packages have emerged to facilitate user-friendliness. The by far most popular packages are `mice` and `Amelia`. Since its inception in 2001, `mice` has been cited 5,276 times on Google Scholar at the time of writing. `Amelia` was created in 2006 and has been cited 1,836 times. They are both considered among the very best implementations of multiple imputation (Horton & Kleinman, 2007). Any improvement in multiple imputation thus needs to be measured against them. `hot.deck`, the method by Gill & Cranmer (2012) upon which my proposed method of multiple hot deck imputation with ordinal variables, hd.ord is based, follows this approach and demonstrates improved results when compared to Amelia. hd.ord extends this and also includes mice as a further benchmark of performance.

The following sections do not cover the full list of functions available in each package, as this would go far beyond the scope of this chapter and could literally fill books of its own, as evidenced by the 182-page package manual for mice (Buuren et al., 2020). Instead, I will focus on the packages' core underlying mechanisms and their major functions to perform imputation, which are named after their package namesakes: mice, amelia, hot.deck, and hd.ord.[2] I extend the focus on simplicity and user-friendliness further by running these major imputation functions with their default settings. Survey analysts usually do not possess the statistical expertise that enable them to dive deeply into distribution or chain properties. The vast majority can be assumed to use imputation functions with their default settings, particularly when applied to surveys and survey experiments that cover a nationally representative population sample. If a package only proved superior over others by setting specific and highly technical function arguments, this would defeat the purpose of making multiple imputation the missing data approach for the masses. I apply only two very minor exceptions to the default settings: The number of imputations is set to the percentage of missingness instead of the default 5, and the console printing options are turned off. Since hot.deck requires the setting of a method, this is set to "best.cell". The same applies to hd.ord.

mice: Multivariate Imputation by Chained Equations

mice is an R package released in 2001 (Buuren & Oudshoorn, 2000). It stands for Multiple Imputation by Chained Equations (MICE), which means imputing incomplete multivariate data by full conditional specification (FCS) (Buuren, 2007; Buuren

[2]For the remainder of this chapter and to avoid confusion, all names will refer to the function unless explicitly stated otherwise.

& Groothuis-Oudshoorn, 2011), a version of the imputation-posterior (IP) (King et al., 2001). The initial release of the `mice` package featured predictor selection, passive imputation, and automatic pooling. Subsequent releases included functionality for imputing multilevel data, post-processing imputed values, specialized pooling, stable imputation of categorical data, and model selection, among many others. Imputation by chained equations is extensively used across domains (see Buuren & Groothuis-Oudshoorn (2011) for a list of over 20 applied fields). Full conditional specification refers to imputation on a variable-by-variable basis, i.e. a set of conditional densities is used to impute data for each individual missing value. This approach does not require the specification of a multivariate distribution for the missing data, which separates it from competing methods like joint modeling (Schafer, 1997).

Chained equations are based on the Gibbs sampler, a randomized Markov chain Monte Carlo algorithm to estimate a sequence of observations from a specified multivariate probability distribution (Gelman et al., 2013; Gill, 2014). In essence, chained equations fill in missing values through an iterative repetition of univariate procedures that are chained together – hence the name for the procedure. As the term univariate signifies, specification happens at the variable level, i.e. each chained equation specifies the imputation model separately for each column of the data. Following deliberations by Rubin (1987) and Buuren & Groothuis-Oudshoorn (2011), imputation by chained equations takes the missing data generating process into account and maintains data relations as well as the uncertainty about these relations. With these conditions satisfied, the imputation model results in statistically valid and factual imputations. This has been shown empirically under various circumstances, for instance for regression models (Giorgi, Belot, Gaudart, & Launoy, 2008; Horton & Kleinman, 2007; Horton & Lipsitz, 2001), continuous data (Yu, Burton, & Rivero-Arias, 2007), missing predictor variables (Moons, Donders, Stijnen, & Harrell, 2006), large surveys (Schunk, 2008), and addressing issues of convergence (Brand, 1999; Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006; Drechsler & Rassler, 2008).

Continuing the notation from section 3.2.1 and incorporating Buuren & Groothuis-Oudshoorn (2011), let there be Y, an $n \times v$ matrix with data on n respondents for v variables, that is formed of missing, Y_{miss}, and observed data, Y_{obs}. As before, let there also be a vector of unknown parameters θ. Now let Y further be a random sample from the z-variate multivariate distribution, $Z(Y|\theta)$, with θ accounting for the multivariate distribution of Y. The proverbial pot of gold here is how to estimate the multivariate distribution of θ. Under the chained equations model, we estimate a posterior distribution of θ by sampling repeatedly from conditional distributions, i.e.:

$$Z(Y_1|Y_{-1}, \theta_1)$$
$$\vdots$$
$$Z(Y_z|Y_{-z}, \theta_z)$$

Any iteration n of chained equations is then a Gibbs sampler that sequentially draws

$$\begin{aligned}
\theta_1^{*(n)} &\sim Z(\theta_1|Y_1^{obs}, Y_2^{(n-1)}, ..., Y_z^{(n-1)}) \\
Y_1^{*(n)} &\sim Z(Y_1|Y_1^{obs}, Y_2^{(n-1)}, ..., Y_z^{(n-1)}, \theta_1^{*(n)}) \\
&\vdots \\
\theta_z^{*(n)} &\sim Z(\theta_z|Y_z^{obs}, Y_1^{(n)}, ..., Y_{z-1}^{(n)}) \\
Y_z^{*(n)} &\sim Z(Y_z|Y_z^{obs}, Y_1^{(n)}, ..., Y_z^{(n)}, \theta_z^{*(n)})
\end{aligned}$$

with the chain starting from a random draw from observed marginal distributions and $Y_i^{(n)} = (Y_i^{obs}, Y_i^{*(n)}$ being the ith imputed variable at iteration n. Note that immediately preceding imputations, $Y_i^{*(n-1)}$, do not affect $Y_i^{*(n)}$ directly but only through connections with other variables.

Figure 3.2 shows the package's main imputation function, `mice`, with all its arguments. As stated above, I will use `mice` with its default settings to ensure simplicity and user-friendliness. The majority of arguments are not of importance to general users. Ar-

```
mice(data, m = 5, method = NULL, predictorMatrix, where = NULL,
     blocks, visitSequence = NULL, formulas, blots = NULL, post = NULL,
     defaultMethod = c("pmm", "logreg", "polyreg", "polr"), maxit = 5,
     printFlag = TRUE, seed = NA, data.init = NULL, ...)
```

Figure 3.2. The `mice` function

guments like `predictorMatrix`, which specifies the set of predictors to be used for each target column, and `blocks`, which provides the option to manually put variables into imputation blocks, require too much statistical knowledge to be of use to non-specialists. Other arguments do not affect the basic workings of the function. This applies for instance to `printFlag`, which sets the console printing preference, `seed`, which is used to offset the random number generator, and `data.init`, which specifies a data frame to be used to initialize imputations before the start of the iterative process.

The only important arguments for general users are `data`, `m`, and `defaultMethod`. Only two of these require input from general users: `data` and `m`. `data` requires a data frame with missing values and `m` should be set to the percentage of missing data, as mentioned above. `defaultMethod` does not require user input but is crucial for insight into the default workings of `mice`. Its options `pmm`, `logreg`, `polyreg`, and `polr` refer to the default imputation methods that are implemented depending on the type of variable in question. `pmm` (predictive mean matching) is used for numeric data, `logreg` (logistic regression imputation) for binary and factor data with two levels, `polyreg` (polytomous regression imputation) for factor data with more than two unordered levels, and `polr` (proportional odds model) for factor data with more than 2 ordered levels. Note that `mice` thus distinguishes between ordered and unordered as well as the number of factor levels, but does not specifically incorporate ordinal variables, which feature ordered but unevenly spaced levels.

`Amelia`: A Program for Missing Data

`Amelia` is an `R` package originally released in 1998 (Honaker, Joseph, King, Scheve, & Singh, n.d.). A second version, `Amelia II`, was released in 2010 (Honaker, King, & Blackwell, 2012). Contrary to `mice`, which is based on IP, both versions of `Amelia` are based on the expectation-maximization (EM) algorithm (Dempster et al., 1977; Gelman et al., 2013; Jackman, 2000; McLachlan & Krishan, 1997; Tanner, 1996). EM functions to a large part like IP, with the crucial exception that deterministic calculations of posterior means replace random draws from the entire posterior. This translates into running regressions to estimate the regression coefficient, imputing a missing value with a predicted value, re-estimating the regression coefficient, and repeating the process until convergence (King et al., 2001). While the iterations and parameters thus represent an entire density in IP, they are single maximum posterior values in EM. This makes EM comparatively much faster in finding the maximum of the likelihood function. On its own, however, EM is unsuitable for multiple imputation as it does not provide the rest of the distribution. `Amelia` circumvents this issue with expectation-maximization importance sampling (EMi) (Gelfand & Smith, 1990; Rubin, 1987; Wei & Tanner, 1990), which combines EM with the iterative simulation approach of importance sampling. This proved unsuitable for large datasets, however, as it led to high running times and system crashes. `Amelia II` addresses this by mixing the inference methods EM and bootstrapping (Efron, 1994; Lahlrl, 2003; Rubin, 1994; Shao & Sitter, 1996), allowing the imputation of more variables for more observations more quickly.

`Amelia II` is based on the assumption that the complete data (Y_{obs} and Y_{miss}) are multivariate normal (MVN): $Y \sim N_v(\mu, \sum)$, with mean vector μ and covariance matrix $\sum$. The MVN model has been proven to work for a variety of variable types (Ezzati-Rice et al., 1995; Graham & Schafer, 1999; Rubin & Schenker, 1986; Schafer, 1997). Continuing the notation from section 3.2.1 and incorporating Honaker & King (2010), let there be a vector of unknown parameters θ, with $\theta = (\mu, \sum)$. Let there further be our missingness matrix$R$ and the likelihood of Y_{obs}, prob$(Y_{obs}, R|\theta)$. Amelia II is explicitly set up for the MAR assumption of missing data, prob$(R = 0|Y_{obs}, \theta)$. Under this assumption, the likelihood can be transformed as

$$\text{prob}(Y_{obs}, R|\theta) = \text{prob}(R|Y_{obs})\text{prob}(Y_{obs}|\theta).$$

Since the missing mechanism is MAR, we are only interested in the inference on complete data parameters, thus the likelihood becomes

$$L(\theta|Y_{obs}) \propto \text{prob}(Y_{obs}|\theta)$$

which further translates into

$$\text{prob}(Y_{obs}|\theta) = \int \text{prob}(Y|\theta)yY_{miss}$$

under the law of iterated expectations. This results in the posterior

$$\text{prob}(\theta|Y_{obs}) \propto \text{prob}(Y_{obs}|\theta) = \int \text{prob}(Y|\theta)yY_{miss}.$$

Taking draws from this posterior is computationally intensive since the contents of μ and $\sum$ increase exponentially as the number of variables increases – this is the perennial crux of multiple imputation, particularly for large datasets with many variables. Amelia II solves this through a combination of EM and bootstrapping. This process bootstraps the data to simulate estimation uncertainty for each posterior draw, runs the EM algorithm to find the mode of the posterior bootstrapped data, and then imputes by drawing from Y_{miss} conditional on Y_{obs} and the respective draws of θ. The latter is a linear regression with parameters that can be estimated from θ. This bootstrapped EM approach is said to be faster than IP as Markov chains do not need be assessed for convergence and an improvement over EMi since the variance matrix of μ and $\sum$ do not need to be calculated, allowing the algorithm to handle larger datasets.

Figure 3.3 shows the package's main imputation function, `amelia`, with all its arguments. As stated above, I will use `amelia` with its default settings to ensure simplicity and user-friendliness. As with `mice`, the majority of arguments are not of importance

```
amelia(x, m = 5, p2s = 1,frontend = FALSE, idvars = NULL,
       ts = NULL, cs = NULL, polytime = NULL, splinetime = NULL, intercs = FALSE,
       lags = NULL, leads = NULL, startvals = 0, tolerance = 0.0001,
       logs = NULL, sqrts = NULL, lgstc = NULL, noms = NULL, ords = NULL,
       incheck = TRUE, collect = FALSE, arglist = NULL, empri = NULL,
       priors = NULL, autopri = 0.05, emburn = c(0,0), bounds = NULL,
       max.resample = 100, overimp = NULL, boot.type = "ordinary",
       parallel = c("no", "multicore", "snow"),
       ncpus = getOption("amelia.ncpus", 1L), cl = NULL, ...)
```

Figure 3.3. The `amelia` function

to general users. Specifications such as `splinetime`, which allows the control of cubic smoothing splines of time, and `lags`, which indicates columns in the data that should have their lags included in the imputation model, will only be used in very particular situations by a small minority of users. Other arguments likewise are not crucial to the basic workings of the function, such as `p2s` to control console printing and `parallel` to identify any type of parallel operation to be used.

The only arguments that require user input are `x` and `m`. `x` requires data with missing values that can be in a variety of formats, while `m`, similar to `mice`, should be adjusted to reflect the percentage of missingness in the data. Three other arguments are important since they arguably comprise the core of `amelia`'s underlying imputation mechanism: `tolerance`, `autopri`, and `boot.type`. `tolerance` sets the convergence threshold for the EM algorithm. `autopri` allows the EM chain to increase the empirical prior if the path strays into an non-positive definite covariance matrix. `boot.type` offers the option to turn off the non-parametric bootstrap that is applied by default.

`amelia` incorporates ordinal variables to some extent. General research treats independent ordinal variables as continuous variables. `amelia` supports this and treats ordinal variables as continuous variables as a default. This means missing ordinal variables are imputed on a continuous scale, rather than preserved as the factual levels present in the observed data. However, the `ords` argument allows users to 'disable' continuous ordinal imputation. In this case, ordinal variables are still imputed on a continuous scale, but these imputations are then scaled and used as the probability of success in a binomial distribution. The draw from this binomial distribution is then transformed into one of the ordinal levels present in the observed data by rounding. Note that none of these `amelia` features address or reflect the spacing between the ordinal variable categories.

`hot.deck`: Multiple Hot Deck Imputation

`hot.deck` is an `R` package released in 2012 (Gill & Cranmer, 2012). It combines a variation of non-parametric hot decking (see 3.2.3) with multiple imputation and aims to fill gaps where parametric multiple imputation, i.e. the approach used in `mice` and `amelia`, falls short (Fuller & Kim, 2005; Kim, 2004; Kim & Fuller, 2004; Reilly, 1993). Like hot decking, `hot.deck` uses draws of actual observable values (*donors*) to fill missing values (*recipients*. In order to account for uncertainty around the drawn values, `hot.deck` iterates these draws over m imputations and pools the results.

The main proposed advantage of `hot.deck` lies in its applicability to missing data with discrete variables with a small number of categories. Approaches like the one used in `amelia`, for instance, by default impute discrete data on a continuous scale. This changes the nature of discrete variables and practically turns them into continuous variables. This can result in non-observable, biased, and sometimes even nonsensical imputation values with artificially smaller standard errors. The proposed `amelia` solution of rounding continuous imputations is problematic as well: Let imputation 1 of a binary variable between 0 and 1 be 0.4. Let further imputation 2 of the binary variable be 0.6. With rounding, these imputations become 0 and 1, when they are in fact 0.4 and 0.6. The rounding problem is further exacerbated for ordinal variables, where the spacing between the discrete variable categories is unknown, since it arbitrarily reduces or lengthens distances between the categories. Rounding thus by definition introduces at least some level of bias. This is not the case in `hot.deck` as it preserves the integrity of discrete data, does not change the size of standard errors, and produces more accurate imputations. `hot.deck` also does not require assumptions of a MVN distribution that is required by `amelia`.

Following Gill & Cranmer (2012), `hot.deck` estimates affinity scores, α, for each missing value to measure how similar a respondent with a missing value, the recipient c, is to another respondent, the potential donor o, across all variables except the missing one. Each score is bounded by 0 and 1. The total set of affinity scores is denoted by α_{co}. For each respondent, let there be vector (p, v), with p being the dependent variable and v a vector of discrete explanatory variables of length k. If recipient c has q_c missing values in v_c, then the potential donor vector, v_o has between 0 and $k - q_c$ exact matches with c. Let w_{co} be the number of variables where c and o have non-identical values. This leaves $k - q_c - w_{co}$ as the number of variables where they have identical values. Scaled by the highest number of possible matches $(k - q_c)$, this value forms the affinity score

$$\alpha_{co} = \frac{k - q_c - w_{co}}{k - q_c}$$

for each missing value recipient c. When the number of identical matches decreases, so does α_{co}. While this might work well for binary variables, it poses a problem for discrete variables with many levels, as the probability to find identical matches decreases. To account for this, `hot.deck` treats potential donors o for the hth variable in $v_{o[h]}$ that are 'close' differently than potential donors o that are further away. 'Close' is defined as $v_{o[h]}$ and $v_{c[h]}$ being in the same concentric standard deviation from $\overline{h}$, the mean of variable h. Values outside of this range are penalized while values within this range are counted as matches. All donors with the highest affinity scores, i.e. all matches, form the best imputation cell B. Since all values of $v_{c[h]}$ in B are part of the same distribution of independent and identically distributed (iid) random variables, which satisfies the MCAR requirement, we can use random draws from B to impute the missing value. As with the other multiple imputation approaches, this process is then repeated m times for each missing value to account for imputation uncertainty.

Figure 3.4 shows the package's main imputation function, `hot.deck`, with all its arguments. As before, I will use `hot.deck` with its default settings. Like `mice` and `amelia`,

```
hot.deck(data, m = 5, method = c("best.cell", "p.draw"), cutoff = 10,
         sdCutoff = 1, optimizeSD = FALSE, optimStep = 0.1, optimStop = 5,
         weightedAffinity = FALSE, impContinuous = c("HD", "mice"),
         IDvars = NULL, ...)
```

Figure 3.4. The ***hot.deck*** *function*

`hot.deck` only features two arguments that require user input, `data` and `m`. Similar to the other functions, `data` requires a data frame or matrix with missing values, while `m` should once more be set to the percentage of missingness. Specialized arguments such as `optimStep` and `optimStop`, which can be tweaked to optimize standard deviation cutoff parameters, as well as `weightedAffinity`, which indicates whether a correlation-weighted affinity score should be used, do not apply to general users.

`method` and `cutoff` form the core of `hot.deck`. The default setting of `best.cell` in the `method` argument implements multiple hot deck imputation. The alternative, `p.draw`, on the other hand, merely conducts random probabilistic draws. `cutoff` allows users to specify which variables the algorithm should treat as discrete. By default, any variable up to and including ten unique values is considered discrete. This thus includes the majority of political science survey measures, with the sensible exceptions of variables like age or widely spread assessments of income levels.

Overall, `hot.deck` is a specialized function to improve the application of multiple imputation for discrete data and has been shown to do so for highly granular discrete data (Gill & Cranmer, 2012). Moreover, political science survey research relies on highly discrete measures. What is missing from `hot.deck`, however, is the incorporation of ordinal variables as a special form of discrete data. I thus identify this gap as the ideal leverage point to improve the use of ordinal variables in the imputation of missing data. To do so, I adapt `hot.deck` to form `hd.ord`, a function specifically designed to utilize the ordered but unevenly spaced information contained in ordinal variables.

`hd.ord`: Multiple Hot Deck Imputation with Ordinal Variables

`hd.ord` is a self-penned `R` function designed specifically to implement multiple hot deck imputation with ordinal variables. It an extension of `hot.deck` and fully utilizes the unevenly spaced yet ordered information contained in ordinal variables. As described in section 1.1, ordinal variables matter in political science surveys because a key variable in such surveys is ordinal: Education. The importance of the spacing between education values is best shown with a simplified example shown in Table 3.1. Respondent B shows

Respondent	Age	Party ID	Education	Income	Gender
A	25	Republican	High School Graduate	$30-40,000	Male
B	40	NA	Some High School	$20-30,000	Female
C	30	Democrat	Bachelor's Degree	$50-60,000	Female

Table 3.1. Illustrative Data

.

missing data for party ID. To impute a fill-in value, we look at how close respondents A and C are to B in terms of age, education, income, and gender. C is closer to B in terms of age and they share the same gender. A is closer to B on education and income. `hot.deck` measures these distances and estimates affinity scores for respondents A and C. The affinity scores measure how close A and C are to B on all variables except the missing one, i.e. party ID. B then receives the party ID fill-in value from whichever respondent has the higher score. The algorithm building the affinity score is based on evenly spaced sequential numeric values, e.g. 1, 2, 3 etc. to represent the distances between the variable categories. This is the case for age, income, and gender, but not for education, since education is an ordinal variable. Applying `hot.deck` to such a numeric representation would misrepresent the data.

Instead, `hd.ord` applies `polr` from the `MASS` package to any specified number of ordinal variables in the data to estimate the underlying latent continuous variable. This estimates cutoff thresholds between the ordinal categories and bins data cases according to the linear predictors. The binned cases determine which variable categories make sense, given the underlying latent continuous variable. This can result in a reduction of education categories if the categories are too finely thinned out. `hd.ord` uses these newly estimated categories. Rather than assigning evenly spaced sequential numeric values to them, the function estimates the mid-cutpoints between each category, based on the `polr` results. We then replace the ordinal variable categories with the newly estimated numeric mid-cutpoints in the data. Finally, these values are scaled and used in the assessment of distance to calculate affinity scores.

Table 3.2 displays illustrative results from running `polr` on survey data, with column "Thresholds" showing the estimated cutoff thresholds between the education categories. Table 3.3 in turn shows the estimated mid-cutpoints for each of the education categories. The mid-cutpoint values for the categories in Table 3.3 fall between the adjacent values in Table 3.2, i.e. the mid-cutpoint of .547 for Some High School lies between the respective thresholds of .032 and 1.063. To estimate the beginning cutpoint for the first category (Less Than High School), we halve the difference between the first and second threshold and subtract this value from the first threshold: $.032 - (1.063 - 0.032) / 2 = -0.484$. The same process is applied the ending cutpoint for the last category (Master's Degree). The mid-cutpoint values are then scaled and used for the calculation of the affinity scores.

Table 3.2. Illustrative Data polr Results

Intercepts	Thresholds
Less Than High School|Some High School	.032
Some High School|High School Graduate	1.063
High School Graduate|Some College	1.783
Some College|Bachelor's Degree	3.302
Bachelor's Degree|Master's Degree	4.907

Table 3.3. Illustrative Data Value Replacements

Original Education Categories	Mid-Cutpoints
Less Than High School	-.484
Some High School	.547
High School Graduate	1.423
Some College	2.542
Bachelor's Degree	4.104
Master's Degree	5.709

Figure 3.5 shows my self-penned imputation function, `hd.ord`, with all its arguments. As before, I will use `hd.ord` with its default settings. Since `hd.ord` is an adaptation of `hot.deck`, the two functions are identical except for the `ord` argument, which allows users to specify the ordinal variables for `polr` treatment.

```
hd.ord(data, ord, m = 5, method=c("best.cell", "p.draw"), cutoff=10,
       sdCutoff=1, optimizeSD = FALSE, optimStep = 0.1, optimStop = 5,
       weightedAffinity = FALSE, impContinuous = c("HD", "mice"),
       IDvars = NULL, ...)
```

Figure 3.5. The ***hd.ord*** *function*

3.3 Data

To test the performance of several imputation methods, we need to work with complete data, as only complete data allow us to obtain the true values needed as a benchmark for comparison. I choose three different sets of survey data: An experiment on moral framing I ran on MTurk in 2017 (original data), data from the 2016 ANES, and data from the 2016 CCES. Data for all selected variables in all three datasets is complete. In order to test the accuracy of several imputation methods, I delete data from these complete data sets with the `ampute()` function from the `mice` package (Buuren et al., 2020). `ampute()` allows the removal of data MCAR, MAR, and MNAR. Particularly the availability of the latter offers unique opportunities: Establishing whether real-life missing data is MNAR a difficult feat. Data that are artificially MNAR, however, circumvent this problem and allow us to test the accuracy of imputation methods for data MNAR as well. `ampute()` has been shown to accurately remove data MCAR, MAR, and MNAR (Schouten, Lugtig, & Vink, 2018).

Each dataset is imputed with five different functions: `amelia`, `mice`, `hot.deck`, `hd.ord`, and `na.omit` (for listwise deletion). As outlined in section 3.2.4, all functions are used with their default settings with only two exceptions: The number of imputations is set to the percentage of missingness instead of the default 5, and the console printing options are turned off.

I test each function for imputation accuracy and speed for binary, ordinal, and interval variables in all three datasets. Each dataset contains two ordinal (`Education`, `Interest`), two interval (`Age`, `Income`) and numerous binary variables. In order to enable factually accurate comparison and unless specified otherwise, each dataset contains 1,000 observations and six levels of the ordinal variable `Education`. 1,000 observations represent a common size for survey experiment data, and the `polr` analysis from section 2.3 estimates five or six levels to best represent `Education` in a US context. Imputation of each dataset was carried out with each of the five imputation methods for 1,000 iterations. With the exception of Table 3.12, 20 percent NA were randomly amputed in each iteration for each dataset.

Following Collins et al. (2001) and Honaker & King (2010), as many relevant predictor variables as possible were used to impute each of the datasets. For the ANES data, up to 14 predictor variables were used: `Ind` (Independent), `Moderate`, `Black`, `Hisp` (Hispanic), `Asian`, `Empl` (Employed), `Stud` (Student), `Religious`, `InternetHome`, `OwnHome`, `Rally` (have you attended a political rally), `Donate` (have you donated to a political candidate), `Married`, and `Separated`. For the CCES data, up to 17 predictor variables were used: `Rep` (Republican), `Moderate`, `Liberal`, `Black`, `Hisp`, `Asian`, `Empl`, `Unempl` (Unemployed), `Stud`, `Gay`, `Bisexual`, `StudLoans` (do you have student loans), `InternetHome`, `NotReligious`, `RentHome`, `Separated`, and `Single`. For the Framing data, up to 13 predictor variables were used: `Ind`, `Conservative`, `Liberal`, `Black`, `Hisp`, `White`, `Asian`, `Unempl`, `Ret` (Retired), `Stud`, `Official` (have you written to a political official), `Media` (how much do you follow public affairs in the media), and `Participation` (how many political activities have you participated in). Highly collinear variables were excluded with a cutoff of 0.6.

The variable mean serves as the baseline of comparison for the performance of each imputation method. Since each dataset is complete, we know the true variable mean of all variables. The closer a method comes to the true mean, the better its performance. Sections 3.4.1 to 3.4.4 show the results in terms of performance accuracy. First, I impute all datasets MAR (section 3.4.1) and MNAR (section 3.4.2) for five and 12 amputed variables. The five amputed variables are consistent across all three datasets: `Democrat` (binary), `Male` (binary), `Interest` (ordinal, scaled from 1 to 4), `Income` (interval), and `Age` (interval). The 12 amputed variables contain additional variables that differ between each dataset due to availability in the data. I do not impute these datasets MCAR for obvious reasons: Imputation is not necessary for data MCAR because simple deletion leads to unbiased and therefore valid results. In section 3.4.3, I increase the number of ordinal variables to be treated by `polr` for `hd.ord` by including `Interest`. Imputations are also conducted MAR and MNAR for four and 11 amputed variables. The variables are the same as in sections 3.4.1 and 3.4.2, but without including `Interest`. Finally, I increase the amount of missing data to 50 and 80 percent (section 3.4.4).

Section 3.4.5 shows the results in terms of performance speed by the number of imputed variables (Table 3.13) and the percentage of missingness (Table 3.14). All running times were achieved on a Code Ocean AWS EC2 instance with 16 cores and 120 GB of memory.

3.4 Results

3.4.1 MAR

This section shows the imputation results for the MAR missing data mechanism. MAR amputation was achieved by setting the `mech` argument in the `ampute` function to `MAR`. Table 3.4 shows the results of imputing all datasets MAR for five amputed variables. For the two binary variables, `Dem` and `Male`, `hd.ord` performs on par or worse than `hot.deck` for all datasets, while `mice` and `amelia` perform best. `hd.ord` is relatively close for CCES `Dem` (+.0001 `amelia` vs. –.0004 `hd.ord`) but further away for ANES (+.0003 `mice` vs. –.0011 `hd.ord` `Dem`; +.0001 `mice` vs. –.0013 `hd.ord` `Male`) and CCES `Male` (–.0001 `amelia` vs. –.0014 `hd.ord`). For the ordinal variable, `Interest`, `hd.ord` performs worst for all datasets, with considerable distance to `hot.deck` (–.0191 vs. –.0130 ANES; –.0196 vs. –.0125 CCES; –.0248 vs. –.0213 Framing). `mice` and `amelia` perform by far best across all datasets. The performance differences between the methods are far larger for the ordinal than for the binary variables: `mice` is not more than +.0003 (ANES) away from the true value across all datasets, while the maximum difference for `hd.ord` amounts

Table 3.4. Accuracy of Multiple Imputation Methods. MAR, 5 Variables with NA

Method	Variable	ANES	CCES	Framing
true	Dem	.3420	.3770	.4660
hot.deck	Dem	−.0010	+.0000	+.0001
hd.ord	Dem	−.0011	−.0004	+.0008
amelia	Dem	+.0004	+.0001	−.0001
mice	Dem	+.0003	+.0002	+.0000
na.omit	Dem	−.0290	−.0229	−.0340
true	Male	.4890	.4830	.5260
hot.deck	Male	−.0013	−.0011	+.0001
hd.ord	Male	−.0013	−.0014	+.0005
amelia	Male	+.0002	−.0001	+.0000
mice	Male	+.0001	−.0001	−.0001
na.omit	Male	−.0392	−.0414	−.0256
true	Interest	2.9340	3.3290	3.2170
hot.deck	Interest	−.0130	−.0125	−.0213
hd.ord	Interest	−.0191	−.0196	−.0248
amelia	Interest	+.0003	+.0003	−.0002
mice	Interest	+.0003	+.0000	−.0003
na.omit	Interest	−.0705	−.0724	−.0714
true	Inc	16.6140	6.4810	3.0890
hot.deck	Inc	−.1068	−.0259	−.0119
hd.ord	Inc	−.1278	−.0407	−.0192
amelia	Inc	+.0008	−.0004	+.0003
mice	Inc	+.0003	−.0002	+.0014
na.omit	Inc	−.5631	−.2468	−.1367
true	Age	50.0410	52.8230	37.9120
hot.deck	Age	−.3888	−.2616	−.3650
hd.ord	Age	−.4597	−.3895	−.3923
amelia	Age	+.0007	−.0033	−.0039
mice	Age	+.0017	−.0073	−.0049
na.omit	Age	−1.1875	−1.2361	−1.0641

Note: Each `true` value shows the true variable mean. All other values show the differences between the imputation means and the true mean, indicated with a + or – sign.

to –.0248 (Framing).

For the interval variables, `Inc` and `Age`, `hd.ord` also performs worst for all datasets. The distance to `hot.deck` is once more considerable. `mice` performs best for `Inc` and `amelia` shows the best results for `Age`. The performance differences between the methods are even larger here: For `Inc`, `mice` is not more than +.0014 (Framing) away from the true value across all datasets, but the maximum difference for `hd.ord` is –.1278 (ANES). Similarly, `amelia`'s largest deviation from the true value for `Age` is –.0039 (Framing) as opposed to `hd.ord`'s –.4597 (ANES).

Table 3.5 shows the results of imputing all datasets MAR for 12 amputed variables. The first five listed variables are the same as for the MAR analysis for five amputed variables in Table 3.4. The remaining variables were chosen based on availability in each dataset. As much as possible, the same variables were selected across all datasets. `Black`, `Empl`, `Religious`, `Married`, `OwnHome`, `Rally`, `Donate`, `Gay`, `StudLoans`, `Hisp`, `Official`, and `Stud` are binary variables. `Black` indicates whether a respondent is of African-American origin, `Empl` whether she is currently employed, `Religious` whether she follows a religious belief, `Married` whether she is currently married, `OwnHome` whether she owns her home, `Rally` whether she has attended a political rally, `Donate` whether she has donated to a political candidate, `Gay` whether she identifies as homosexual, `StudLoans` whether she currently has student loans, `Hisp` whether she is of Hispanic origin, `Official` whether she has contacted her political representative, and `Stud` whether she currently is a student. `Media` is an ordinal variable and indicates how much she follows public affairs in the media (scaled from 1 to 5). `Participation` is a interval variable and shows the accumulative count of political activities she has participated in (scaled from 0 to 4).

Table 3.5: Accuracy of Multiple Imputation Methods. MAR, 12 Variables with NA

Method	Variable	ANES	CCES	Framing
true	Dem	.3420	.3770	.4660
hot.deck	Dem	–.0005	–.0003	+.0001
hd.ord	Dem	–.0005	–.0004	+.0004
amelia	Dem	+.0000	+.0000	+.0001
mice	Dem	+.0000	+.0001	+.0001
na.omit	Dem	–.0191	–.0172	–.0280
true	Male	.4890	.4830	.5260
hot.deck	Male	–.0004	–.0002	–.0002

hd.ord	Male	−.0001	−.0003	+.0000
amelia	Male	+.0001	−.0001	−.0001
mice	Male	+.0000	−.0002	−.0001
na.omit	Male	−.0256	−.0364	−.0154
true	Interest	2.9340	3.3290	3.2170
hot.deck	Interest	−.0053	−.0041	−.0095
hd.ord	Interest	−.0077	−.0067	−.0106
amelia	Interest	+.0001	−.0001	+.0002
mice	Interest	+.0000	−.0001	−.0001
na.omit	Interest	−.0620	−.0515	−.0721
true	Inc	16.6140	6.4810	3.0890
hot.deck	Inc	−.0470	−.0130	−.0060
hd.ord	Inc	−.0591	−.0212	−.0089
amelia	Inc	−.0007	−.0005	−.0002
mice	Inc	−.0013	−.0003	+.0000
na.omit	Inc	−.6303	−.2860	−.0960
true	Age	50.0410	52.8230	37.9120
hot.deck	Age	−.1391	−.0883	−.1494
hd.ord	Age	−.1835	−.1435	−.1592
amelia	Age	+.0056	−.0015	−.0041
mice	Age	+.0048	−.0050	−.0029
na.omit	Age	−.8638	−.5974	−.6434
true	Black	.0790	.0950	.0690
hot.deck	Black	+.0000	+.0000	−.0001
hd.ord	Black	+.0000	+.0000	−.0003
amelia	Black	+.0000	+.0001	+.0000
mice	Black	+.0000	+.0001	+.0001
na.omit	Black	−.0092	−.0090	−.0110
true	Empl	.6610	.4370	.7430
hot.deck	Empl	+.0006	+.0000	+.0007
hd.ord	Empl	+.0006	+.0001	+.0006
amelia	Empl	+.0000	+.0000	+.0000
mice	Empl	+.0000	−.0001	−.0001
na.omit	Empl	−.0087	−.0301	−.0157
true	Religious	.6460	.6420	—
hot.deck	Religious	−.0006	−.0003	—
hd.ord	Religious	−.0005	−.0003	—
amelia	Religious	−.0001	−.0001	—
mice	Religious	−.0001	−.0002	—
na.omit	Religious	−.0166	−.0234	—
true	Married	.5290	.6310	—
hot.deck	Married	+.0002	−.0001	—
hd.ord	Married	+.0002	+.0001	—
amelia	Married	−.0001	−.0002	—
mice	Married	−.0001	−.0002	—
na.omit	Married	−.0384	−.0326	—
true	OwnHome	.6820	.7010	—
hot.deck	OwnHome	−.0001	−.0002	—
hd.ord	OwnHome	+.0000	+.0000	—
amelia	OwnHome	+.0001	+.0000	—
mice	OwnHome	−.0001	−.0001	—
na.omit	OwnHome	−.0334	−.0304	—

true	Rally	.0830	—	—
hot.deck	Rally	−.0001	—	—
hd.ord	Rally	−.0002	—	—
amelia	Rally	+.0001	—	—
mice	Rally	+.0001	—	—
na.omit	Rally	−.0191	—	—
true	Donate	.1390	—	—
hot.deck	Donate	−.0002	—	—
hd.ord	Donate	−.0005	—	—
amelia	Donate	+.0000	—	—
mice	Donate	+.0001	—	—
na.omit	Donate	−.0320	—	—
true	Gay	—	.0420	—
hot.deck	Gay	—	+.0001	—
hd.ord	Gay	—	+.0000	—
amelia	Gay	—	+.0000	—
mice	Gay	—	+.0000	—
na.omit	Gay	—	−.0112	—
true	StudLoans	—	.1910	—
hot.deck	StudLoans	—	+.0003	—
hd.ord	StudLoans	—	+.0002	—
amelia	StudLoans	—	+.0000	—
mice	StudLoans	—	−.0001	—
na.omit	StudLoans	—	−.0117	—
true	Hisp	—	—	.0550
hot.deck	Hisp	—	—	−.0002
hd.ord	Hisp	—	—	−.0002
amelia	Hisp	—	—	+.0000
mice	Hisp	—	—	+.0000
na.omit	Hisp	—	—	−.0072
true	Official	—	—	.3560
hot.deck	Official	—	—	+.0000
hd.ord	Official	—	—	−.0003
amelia	Official	—	—	+.0001
mice	Official	—	—	+.0001
na.omit	Official	—	—	−.0385
true	Stud	—	—	.0440
hot.deck	Stud	—	—	+.0000
hd.ord	Stud	—	—	+.0000
amelia	Stud	—	—	+.0000
mice	Stud	—	—	+.0001
na.omit	Stud	—	—	−.0012
true	Media	—	—	1.7240
hot.deck	Media	—	—	−.0049
hd.ord	Media	—	—	−.0060
amelia	Media	—	—	+.0001
mice	Media	—	—	+.0001
na.omit	Media	—	—	−.0984
true	Participation	—	—	.9540
hot.deck	Participation	—	—	−.0021
hd.ord	Participation	—	—	−.0027
amelia	Participation	—	—	−.0003

mice	Participation	—	—	–.0001
na.omit	Participation	—	—	–.1003

The results are consistent with those presented in Table 3.4. For the binary variables, `amelia` and `mice` again perform best with results actually matching the true variable values, though `hd.ord` arguably shows closer results than in Table 3.4 with a maximum difference to a true value of +.0006 (ANES and Framing `Empl`). For the ordinal variable, `hd.ord` continues to perform worst across all datasets. The same applies to Framing `Media`. `mice` and `amelia` again show the best results. Note, however, that `hd.ord` is less worse in terms of performance differences when compared to the MAR analysis of five imputed variables. The maximum difference to the true `Interest` value is –.0106 (Framing) for 12 variables but –.0248 (Framing) for five variables. This is possibly explained by a thinner spread of missing values across a higher number of variables, resulting in a lower number of NAs in each amputed variable.

The results for the interval variables follow the same pattern. `hd.ord` displays the worst results for all datasets. The difference to `hot.deck` is still present though less pronounced than in the MAR analysis of five imputed variables. `mice` overall performs better than `amelia` for `Inc` and `Age`, with the exceptions of ANES `Inc` (–.0007 vs. –.0013) and CCES `Age` (–.0015 vs. –.0050). The results for Framing `Participation` confirm those for `Inc` and `Age`.

3.4.2 MNAR

This section shows the imputation results for the MNAR missing data mechanism. MNAR amputation was achieved by setting the `mech` argument in the `ampute` function to `MNAR`. All MAR and MNAR analyses are otherwise identical. Table 3.6 shows the results of imputing all datasets MNAR for five amputed variables. It is immediately noticeable that the differences between the methods' imputation results and the true values is much higher for all methods for all variables for all datasets. The results for `Dem`, for instance, hover around .0100 for ANES/CCES and around .004 for Framing, while the `Male` numbers center around .0125 across all datasets. In the corresponding MAR analysis, however, the results for `Dem` and `Male` revealed around .0005 for all datasets. For the binary variables,

Table 3.6. Accuracy of Multiple Imputation Methods. MNAR, 5 Variables with NA

Method	Variable	ANES	CCES	Framing
true	Dem	.3420	.3770	.4660
hot.deck	Dem	−.0114	−.0099	−.0038
hd.ord	Dem	−.0120	−.0105	−.0033
amelia	Dem	−.0106	−.0102	−.0046
mice	Dem	−.0099	−.0101	−.0036
na.omit	Dem	−.0176	−.0140	−.0185
true	Male	.4890	.4830	.5260
hot.deck	Male	−.0136	−.0116	−.0127
hd.ord	Male	−.0133	−.0124	−.0125
amelia	Male	−.0132	−.0121	−.0133
mice	Male	−.0132	−.0120	−.0135
na.omit	Male	−.0214	−.0219	−.0133
true	Interest	2.9340	3.3290	3.2170
hot.deck	Interest	−.0288	−.0246	−.0333
hd.ord	Interest	−.0335	−.0296	−.0369
amelia	Interest	−.0167	−.0146	−.0133
mice	Interest	−.0167	−.0146	−.0135
na.omit	Interest	−.0379	−.0372	−.0379
true	Inc	16.6140	6.4810	3.0890
hot.deck	Inc	−.2299	−.0928	−.0591
hd.ord	Inc	−.2554	−.1038	−.0648
amelia	Inc	−.1225	−.0578	−.0463
mice	Inc	−.1229	−.0566	−.0445
na.omit	Inc	−.2770	−.1334	−.0740
true	Age	50.0410	52.8230	37.9120
hot.deck	Age	−.6319	−.4596	−.7147
hd.ord	Age	−.7415	−.5929	−.7477
amelia	Age	−.2450	−.2266	−.2875
mice	Age	−.2369	−.2160	−.2888
na.omit	Age	−.6427	−.6392	−.5853

Note: Each `true` value shows the true variable mean. All other values show the differences between the imputation means and the true mean, indicated with a + or − sign.

hd.ord performs more closely on par with amelia and mice than in the corresponding MAR analysis above, sometimes more (–.0133 hd.ord vs. –.0132 amelia ANES Male) and sometimes less so (–.0120 hd.ord vs. –.0099 mice ANES Dem). In addition, hd.ord actually represents the best method for both binary variables in the Framing data. Note the interesting results for na.omit here: It performs as well as amelia and mice for Framing Male.

The results for the ordinal variables confirm those of the MAR analysis: hd.ord represents the worst method across all datasets. amelia and mice show by far the best results and are virtually identical with each other, though the differences to the true values are much higher than in the MAR analysis – as is the case for the entire MNAR analysis. na.omit shows results that are close to hd.ord's for Framing.

The results for the interval variables paint the same picture as the ordinal ones. hd.ord shows the worst performance. Note that na.omit now actually outperforms hot.deck for Framing Age and hd.ord for Framing and ANES Age.

Table 3.7 shows the results of imputing all datasets MNAR for 12 amputed variables. The results are consistent with those obtained for five amputed variables MNAR. amelia and mice perform better than hd.ord for the binary variables overall, but often not by much. Occasionally, hd.ord eclipses them (–.0013 vs. –.0014 mice Framing Dem; –.0053 vs. –.0055 mice and amelia ANES Male). na.omit once more performs close to the other methods.

Table 3.7: Accuracy of Multiple Imputation Methods. MNAR, 12 Variables with NA

Method	Variable	ANES	CCES	Framing
true	Dem	.3420	.3770	.4660
hot.deck	Dem	–.0049	–.0046	–.0015
hd.ord	Dem	–.0049	–.0049	–.0013
amelia	Dem	–.0043	–.0045	–.0018
mice	Dem	–.0040	–.0044	–.0014
na.omit	Dem	–.0092	–.0081	–.0106
true	Male	.4890	.4830	.5260
hot.deck	Male	–.0055	–.0049	–.0052

hd.ord	Male	–.0053	–.0051	–.0051
amelia	Male	–.0055	–.0050	–.0054
mice	Male	–.0055	–.0049	–.0055
na.omit	Male	–.0093	–.0119	–.0060
true	Interest	2.9340	3.3290	3.2170
hot.deck	Interest	–.0113	–.0090	–.0139
hd.ord	Interest	–.0134	–.0113	–.0150
amelia	Interest	–.0068	–.0061	–.0053
mice	Interest	–.0068	–.0061	–.0054
na.omit	Interest	–.0236	–.0161	–.0252
true	Inc	16.6140	6.4810	3.0890
hot.deck	Inc	–.0899	–.0350	–.0239
hd.ord	Inc	–.1046	–.0421	–.0262
amelia	Inc	–.0495	–.0223	–.0184
mice	Inc	–.0503	–.0218	–.0177
na.omit	Inc	–.2088	–.0970	–.0333
true	Age	50.0410	52.8230	37.9120
hot.deck	Age	–.2571	–.1732	–.2811
hd.ord	Age	–.3081	–.2251	–.2989
amelia	Age	–.1100	–.1014	–.1159
mice	Age	–.1047	–.0986	–.1159
na.omit	Age	–.3397	–.1367	–.2239
true	Black	.0790	.0950	.0690
hot.deck	Black	–.0035	–.0038	–.0028
hd.ord	Black	–.0037	–.0038	–.0029
amelia	Black	–.0037	–.0040	–.0011
mice	Black	–.0034	–.0038	–.0010
na.omit	Black	–.0045	–.0052	–.0045
true	Empl	.6610	.4370	.7430
hot.deck	Empl	–.0034	–.0053	–.0022
hd.ord	Empl	–.0033	–.0053	–.0022
amelia	Empl	–.0031	–.0040	–.0024
mice	Empl	–.0031	–.0040	–.0025
na.omit	Empl	–.0014	–.0111	–.0037
true	Religious	.6460	.6420	—
hot.deck	Religious	–.0045	–.0039	—
hd.ord	Religious	–.0043	–.0038	—
amelia	Religious	–.0040	–.0040	—
mice	Religious	–.0040	–.0040	—
na.omit	Religious	–.0049	–.0073	—
true	Married	.5290	.6310	—
hot.deck	Married	–.0041	–.0038	—
hd.ord	Married	–.0040	–.0037	—
amelia	Married	–.0042	–.0037	—
mice	Married	–.0042	–.0037	—
na.omit	Married	–.0122	–.0096	—
true	OwnHome	.6820	.7010	—
hot.deck	OwnHome	–.0030	–.0030	—
hd.ord	OwnHome	–.0028	–.0027	—
amelia	OwnHome	–.0027	–.0030	—
mice	OwnHome	–.0027	–.0030	—
na.omit	OwnHome	–.0111	–.0090	—

true	Rally	.0830	—	—
hot.deck	Rally	–.0043	—	—
hd.ord	Rally	–.0042	—	—
amelia	Rally	–.0040	—	—
mice	Rally	–.0039	—	—
na.omit	Rally	–.0077	—	—
true	Donate	.1390	—	—
hot.deck	Donate	–.0051	—	—
hd.ord	Donate	–.0054	—	—
amelia	Donate	–.0049	—	—
mice	Donate	–.0048	—	—
na.omit	Donate	–.0122	—	—
true	Gay	—	.0420	—
hot.deck	Gay	—	–.0024	—
hd.ord	Gay	—	–.0025	—
amelia	Gay	—	–.0026	—
mice	Gay	—	–.0024	—
na.omit	Gay	—	–.0035	—
true	StudLoans	—	.1910	—
hot.deck	StudLoans	—	–.0058	—
hd.ord	StudLoans	—	–.0058	—
amelia	StudLoans	—	–.0051	—
mice	StudLoans	—	–.0050	—
na.omit	StudLoans	—	–.0070	—
true	Hisp	—	—	.0550
hot.deck	Hisp	—	—	–.0027
hd.ord	Hisp	—	—	–.0027
amelia	Hisp	—	—	–.0011
mice	Hisp	—	—	–.0010
na.omit	Hisp	—	—	–.0033
true	Official	—	—	.3560
hot.deck	Official	—	—	–.0054
hd.ord	Official	—	—	–.0054
amelia	Official	—	—	–.0049
mice	Official	—	—	–.0048
na.omit	Official	—	—	–.0142
true	Stud	—	—	.0440
hot.deck	Stud	—	—	–.0026
hd.ord	Stud	—	—	–.0025
amelia	Stud	—	—	–.0023
mice	Stud	—	—	–.0020
na.omit	Stud	—	—	–.0017
true	Media	—	—	1.7240
hot.deck	Media	—	—	–.0142
hd.ord	Media	—	—	–.0152
amelia	Media	—	—	–.0101
mice	Media	—	—	–.0101
na.omit	Media	—	—	–.0363
true	Participation	—	—	.9540
hot.deck	Participation	—	—	–.0115
hd.ord	Participation	—	—	–.0115
amelia	Participation	—	—	–.0099

mice	Participation	—	—	–.0094
na.omit	Participation	—	—	–.0358

For the ordinal variable, `hd.ord` shows the worst performance across all datasets. `mice` and `amelia` perform far better and display virtually identical result. This also applies to Framing `Media`. `na.omit` does not perform as well as in the corresponding MAR analysis. `mice` and `amelia`'s superior performance is also visible in the results for the ordinal variables, including Framing `Participation`. `na.omit` performs better than `hot.deck` and `hd.ord` for Framing `Age` (–.2239 vs. –.2811 and –.2989), and CCES `Age` (–.1367 vs. –.1732 and –.2251).

3.4.3 Increased Number of Ordinal Variables

This section shows the imputation results where I increase the number of ordinal variables to be treated by `polr` for `hd.ord`. Specifically, I include `Interest` in the `polr` treatment. The intuition behind this is a strengthening of the underlying latent continuous variable assumption. The results so far do not show superior performance by `hd.ord`. However, this might be due to a lack of 'influence' so far. Perhaps one ordinal variable treated with `polr` is not enough to manifest itself in improved results. By including another ordinal variable in the treatment, this 'influence' is strengthened and the `polr` assumption is put to another test. As in sections 3.4.1 and 3.4.2, imputations are conducted MAR and MNAR. Because `Interest` is moved to the `polr` treatment, the number of imputed variables is reduced to four and 11, respectively, to ensure accurate comparison. This means the amputed variables do not include an ordinal variable any more, since no suitable replacement could be found in the ANES and CCES data. The remaining variables are the same as before.

Table 3.8 shows the results of imputing all datasets with two `polr`-treated variables MAR for four amputed variables. `hd.ord` displays the worst results for `Dem` across all datasets and beats only `hot.deck` for `Male`. `amelia` and `mice` perform best and show virtually identically results that often match the true variable means. Comparison with the MAR analysis of five imputed variables reveals that `hd.ord` consistently performs slightly worse here: –.0018, –.0005, +.0012 vs. –.0011, –.0004, +.0008 for `Dem` and –.0015, –.0018, +.0011 vs. –.0013, –.0014, +.0005 for `Male`. `hd.ord` also performs worst for all

Table 3.8. Accuracy of Multiple Imputation Methods. Two Ordinal Variables (Education, Interest), MAR, 4 Variables with NA

Method	Variable	ANES	CCES	Framing
true	Dem	.3420	.3770	.4660
hot.deck	Dem	–.0008	+.0002	+.0005
hd.ord	Dem	–.0018	–.0005	+.0012
amelia	Dem	+.0002	+.0001	+.0001
mice	Dem	+.0001	+.0002	+.0000
na.omit	Dem	–.0333	–.0294	–.0357
true	Male	.4890	.4830	.5260
hot.deck	Male	–.0022	–.0019	+.0015
hd.ord	Male	–.0015	–.0018	+.0011
amelia	Male	+.0001	+.0000	+.0002
mice	Male	+.0000	+.0000	+.0001
na.omit	Male	–.0396	–.0407	–.0297
true	Inc	16.6140	6.4810	3.0890
hot.deck	Inc	–.0830	–.0246	–.0093
hd.ord	Inc	–.1523	–.0516	–.0225
amelia	Inc	+.0010	–.0006	+.0009
mice	Inc	–.0008	+.0002	+.0020
na.omit	Inc	–.5771	–.2564	–.1400
true	Age	50.0410	52.8230	37.9120
hot.deck	Age	–.2889	–.2350	–.3546
hd.ord	Age	–.5431	–.4664	–.4583
amelia	Age	+.0018	+.0085	–.0002
mice	Age	+.0024	–.0002	–.0022
na.omit	Age	–1.1521	–1.1228	–.9688

Note: Each `true` value shows the true variable mean. All other values show the differences between the imputation means and the true mean, indicated with a + or – sign.

datasets across both interval variables. `amelia` and `mice` again perform best. `mice` does better for CCES, while `amelia` does better for Framing. Both perform equally well for ANES. As for the binary variables, the results for `hd.ord` consistently get slightly worse in the switch from one to two ordinal variables in `polr`-treatment: –.1523, –.0516, –.0225 vs. –.1278, –.0407, –.0192 for `Inc` and –.5431, –.4664, –.4583 vs. –.4597, –.3895, –.3923 for `Age`.

Table 3.9 shows the results of imputing all datasets with two `polr`-treated variables MAR for 11 amputed variables. A similar picture for Table 3.8 emerges for the binary variables, with `amelia` and `mice` once more displaying the best results, though `hd.ord` appears to perform somewhat closer here. Consistent with the deterioration of `hd.ord` results from Table 3.4 to Table 3.8, `hd.ord` again consistently performs slightly worse when compared to the MAR analysis with 12 amputed variables and only `Education` treated by `polr`: –.0005, –.0005, +.0005 vs. –.0005, –.0004, +.0004 for `Dem` and –.0001, –.0004, +.0002 vs. –.0001, –.0003, +.0000 for `Male`.

Table 3.9: Accuracy of Multiple Imputation Methods. Two Ordinal Variables (Education, Interest), MAR, 11 Variables with NA

Method	Variable	ANES	CCES	Framing
true	Dem	.3420	.3770	.4660
hot.deck	Dem	–.0002	–.0004	+.0002
hd.ord	Dem	–.0005	–.0005	+.0005
amelia	Dem	+.0000	–.0001	+.0000
mice	Dem	+.0000	+.0000	+.0001
na.omit	Dem	–.0200	–.0213	–.0294
true	Male	.4890	.4830	.5260
hot.deck	Male	–.0002	–.0005	–.0001
hd.ord	Male	–.0001	–.0004	+.0002
amelia	Male	+.0000	–.0001	+.0001
mice	Male	+.0000	–.0001	+.0001
na.omit	Male	–.0254	–.0350	–.0174
true	Inc	16.6140	6.4810	3.0890
hot.deck	Inc	–.0346	–.0114	–.0054
hd.ord	Inc	–.0657	–.0241	–.0095
amelia	Inc	–.0018	–.0007	–.0004
mice	Inc	–.0027	–.0005	–.0001
na.omit	Inc	–.6432	–.2888	–.0983
true	Age	50.0410	52.8230	37.9120
hot.deck	Age	–.0887	–.0704	–.1379
hd.ord	Age	–.1951	–.1532	–.1640
amelia	Age	+.0010	–.0021	+.0001

mice	Age	−.0016	−.0055	+.0022
na.omit	Age	−.8025	−.4330	−.5290
true	Black	.0790	.0950	.0690
hot.deck	Black	+.0001	−.0001	+.0000
hd.ord	Black	+.0000	−.0001	−.0002
amelia	Black	+.0000	+.0001	+.0000
mice	Black	+.0001	+.0001	+.0000
na.omit	Black	−.0103	−.0122	−.0134
true	Empl	.6610	.4370	.7430
hot.deck	Empl	+.0007	+.0003	+.0008
hd.ord	Empl	+.0008	+.0001	+.0007
amelia	Empl	+.0001	+.0000	+.0001
mice	Empl	+.0001	−.0001	+.0000
na.omit	Empl	−.0109	−.0328	−.0159
true	Religious	.6460	.6420	—
hot.deck	Religious	−.0001	−.0002	—
hd.ord	Religious	−.0005	−.0003	—
amelia	Religious	+.0000	−.0001	—
mice	Religious	+.0000	−.0002	—
na.omit	Religious	−.0174	−.0241	—
true	Married	.5290	.6310	—
hot.deck	Married	−.0001	−.0002	—
hd.ord	Married	+.0003	+.0001	—
amelia	Married	−.0001	−.0002	—
mice	Married	−.0001	−.0002	—
na.omit	Married	−.0390	−.0324	—
true	OwnHome	.6820	.7010	—
hot.deck	OwnHome	−.0005	−.0001	—
hd.ord	OwnHome	−.0001	+.0002	—
amelia	OwnHome	+.0000	+.0001	—
mice	OwnHome	−.0001	+.0000	—
na.omit	OwnHome	−.0341	−.0295	—
true	Rally	.0830	—	—
hot.deck	Rally	+.0000	—	—
hd.ord	Rally	−.0002	—	—
amelia	Rally	+.0001	—	—
mice	Rally	+.0002	—	—
na.omit	Rally	−.0186	—	—
true	Donate	.1390	—	—
hot.deck	Donate	−.0003	—	—
hd.ord	Donate	−.0007	—	—
amelia	Donate	−.0001	—	—
mice	Donate	+.0000	—	—
na.omit	Donate	−.0310	—	—
true	Gay	—	.0420	—
hot.deck	Gay	—	+.0001	—
hd.ord	Gay	—	+.0000	—
amelia	Gay	—	+.0000	—
mice	Gay	—	+.0001	—
na.omit	Gay	—	−.0113	—
true	StudLoans	—	.1910	—
hot.deck	StudLoans	—	+.0001	—

hd.ord	StudLoans	—	+.0001	—
amelia	StudLoans	—	−.0001	—
mice	StudLoans	—	−.0001	—
na.omit	StudLoans	—	−.0146	—
true	Hisp	—	—	.0550
hot.deck	Hisp	—	—	−.0002
hd.ord	Hisp	—	—	−.0001
amelia	Hisp	—	—	+.0001
mice	Hisp	—	—	+.0000
na.omit	Hisp	—	—	−.0091
true	Official	—	—	.3560
hot.deck	Official	—	—	−.0006
hd.ord	Official	—	—	−.0004
amelia	Official	—	—	+.0000
mice	Official	—	—	+.0001
na.omit	Official	—	—	−.0373
true	Stud	—	—	.0440
hot.deck	Stud	—	—	+.0000
hd.ord	Stud	—	—	−.0001
amelia	Stud	—	—	+.0000
mice	Stud	—	—	+.0000
na.omit	Stud	—	—	−.0026
true	Media	—	—	1.7240
hot.deck	Media	—	—	−.0047
hd.ord	Media	—	—	−.0059
amelia	Media	—	—	+.0002
mice	Media	—	—	+.0002
na.omit	Media	—	—	−.0921
true	Participation	—	—	.9540
hot.deck	Participation	—	—	−.0019
hd.ord	Participation	—	—	−.0026
amelia	Participation	—	—	−.0002
mice	Participation	—	—	+.0000
na.omit	Participation	—	—	−.0963

The same can be observed for the interval variables, with `hd.ord` again claiming last place across all datasets. The difference to `hot.deck` is still there but less pronounced than in Table 3.8. `amelia` does best for `Age` while `mice` performs better for `Inc`, with the exception of the ANES (–.0027 `mice` vs. –.0018 `amelia`). As for the binary variables, `hd.ord` also consistently performs slightly worse when compared to the MAR analysis with 12 amputed variables and only `Education` treated by `polr`: –.0657, –.0241, –.0095 vs. –.0591, –.0212, –.0089 for `Inc` and –.1951, –.1532, –.1640 vs. –.1835, –.1435, –.1592 for `Age`.

Table 3.10 shows the results of imputing all datasets with two `polr`-treated variables MNAR for four amputed variables. For the binary variables, `hd.ord` performs on the same level as `amelia` and `mice` when compared to the MNAR analysis of five imputed variables with only `Education` treated by `polr`; sometimes more (–.0131 `hd.ord` vs. –.0130 `mice` CCES `Dem`), sometimes less so (–.0155 `hd.ord` vs. –.0127 `mice` ANES `Dem`). `hd.ord` actually shows the best results of all methods for Framing for both variables. `na.omit` does not perform as well as it does in Table 3.6. `hd.ord` again consistently performs slightly worse with the two-ordinal-variable-`polr`-treatment: –.0155,–.0131, –.0042 vs. –.0120, –.0105, –.0033 for `Dem` and –.0172, –.0160, –.0157 vs. –.0133, –.0124, –.0125 for `Male`. The results for the interval variables are consistent with the previous analyses. In addition, note that `na.omit` performs better than `hd.ord` for `Age` across all three datasets and for ANES `Inc`. Once more, `hd.ord` consistently performs slightly worse with more than two ordinal variables: –.3174, –.1303, –.0812 vs. –.2554, –.1038, –.0648 for `Inc` and –.8997, –.7259, –.9349 vs. –.7415, –.5929, –.7477 for `Age`.

Table 3.11 shows the results of imputing all datasets with two `polr`-treated variables MNAR for 11 amputed variables. For the binary variables, `amelia` and `mice` perform better, but often not by much. Occasionally, `hd.ord` eclipses them (–.0057 `hd.ord` vs. –.0060 `mice` ANES `Male`; –.0057 `hd.ord` vs. –.0059 `amelia` Framing `Male`). As was the case in the comparison between the MNAR analysis with four amputed variables and the MNAR analysis with five amputed variables, `hd.ord` consistently demonstrates slightly worse results in the switch from 12 (Table 3.7) to 11 and one to two `polr`-treated variables: –.0053, –.0053, –.0024 vs. –.0049, –.0049, –.0013 for `Dem` and –.0057, –.0056, –.0057 vs. –.0053, –.0051, –.0051 for `Male`.

Table 3.11: Accuracy of Multiple Imputation Methods. Two Ordinal Variables (Education, Interest), MNAR, 11 Variables with NA

Method	Variable	ANES	CCES	Framing
true	Dem	.3420	.3770	.4660

hot.deck	Dem	–.0050	–.0053	–.0022
hd.ord	Dem	–.0053	–.0053	–.0024
amelia	Dem	–.0047	–.0049	–.0024
mice	Dem	–.0044	–.0048	–.0019
na.omit	Dem	–.0097	–.0097	–.0109
true	Male	.4890	.4830	.5260
hot.deck	Male	–.0061	–.0058	–.0061
hd.ord	Male	–.0057	–.0056	–.0057
amelia	Male	–.0060	–.0054	–.0059
mice	Male	–.0060	–.0054	–.0060
na.omit	Male	–.0088	–.0116	–.0068
true	Inc	16.6140	6.4810	3.0890
hot.deck	Inc	–.0841	–.0349	–.0258
hd.ord	Inc	–.1142	–.0473	–.0291
amelia	Inc	–.0540	–.0250	–.0206
mice	Inc	–.0549	–.0249	–.0199
na.omit	Inc	–.2112	–.0982	–.0342
true	Age	50.0410	52.8230	37.9120
hot.deck	Age	–.2124	–.1607	–.2819
hd.ord	Age	–.3329	–.2424	–.3271
amelia	Age	–.1194	–.1135	–.1257
mice	Age	–.1141	–.1102	–.1244
na.omit	Age	–.2978	–.0974	–.2074
true	Black	.0790	.0950	.0690
hot.deck	Black	–.0036	–.0045	–.0024
hd.ord	Black	–.0040	–.0042	–.0031
amelia	Black	–.0041	–.0044	–.0012
mice	Black	–.0037	–.0041	–.0011
na.omit	Black	–.0050	–.0066	–.0054
true	Empl	.6610	.4370	.7430
hot.deck	Empl	–.0042	–.0057	–.0027
hd.ord	Empl	–.0036	–.0058	–.0024
amelia	Empl	–.0034	–.0045	–.0025
mice	Empl	–.0033	–.0044	–.0026
na.omit	Empl	–.0022	–.0122	–.0033
true	Religious	.6460	.6420	—
hot.deck	Religious	–.0045	–.0044	—
hd.ord	Religious	–.0047	–.0042	—
amelia	Religious	–.0043	–.0045	—
mice	Religious	–.0043	–.0045	—
na.omit	Religious	–.0055	–.0077	—
true	Married	.5290	.6310	—
hot.deck	Married	–.0047	–.0043	—
hd.ord	Married	–.0042	–.0040	—
amelia	Married	–.0046	–.0039	—
mice	Married	–.0045	–.0039	—
na.omit	Married	–.0120	–.0095	—
true	OwnHome	.6820	.7010	—
hot.deck	OwnHome	–.0037	–.0035	—
hd.ord	OwnHome	–.0030	–.0030	—
amelia	OwnHome	–.0031	–.0033	—
mice	OwnHome	–.0030	–.0033	—

na.omit	OwnHome	−.0112	−.0088	—
true	Rally	.0830	—	—
hot.deck	Rally	−.0047	—	—
hd.ord	Rally	−.0047	—	—
amelia	Rally	−.0045	—	—
mice	Rally	−.0044	—	—
na.omit	Rally	−.0080	—	—
true	Donate	.1390	—	—
hot.deck	Donate	−.0057	—	—
hd.ord	Donate	−.0060	—	—
amelia	Donate	−.0054	—	—
mice	Donate	−.0053	—	—
na.omit	Donate	−.0122	—	—
true	Gay	—	.0420	—
hot.deck	Gay	—	−.0027	—
hd.ord	Gay	—	−.0026	—
amelia	Gay	—	−.0028	—
mice	Gay	—	−.0027	—
na.omit	Gay	—	−.0037	—
true	StudLoans	—	.1910	—
hot.deck	StudLoans	—	−.0067	—
hd.ord	StudLoans	—	−.0064	—
amelia	StudLoans	—	−.0057	—
mice	StudLoans	—	−.0056	—
na.omit	StudLoans	—	−.0082	—
true	Hisp	—	—	.0550
hot.deck	Hisp	—	—	−.0027
hd.ord	Hisp	—	—	−.0028
amelia	Hisp	—	—	−.0012
mice	Hisp	—	—	−.0012
na.omit	Hisp	—	—	−.0037
true	Official	—	—	.3560
hot.deck	Official	—	—	−.0067
hd.ord	Official	—	—	−.0061
amelia	Official	—	—	−.0055
mice	Official	—	—	−.0054
na.omit	Official	—	—	−.0139
true	Stud	—	—	.0440
hot.deck	Stud	—	—	−.0026
hd.ord	Stud	—	—	−.0025
amelia	Stud	—	—	−.0025
mice	Stud	—	—	−.0021
na.omit	Stud	—	—	−.0021
true	Media	—	—	1.7240
hot.deck	Media	—	—	−.0156
hd.ord	Media	—	—	−.0164
amelia	Media	—	—	−.0110
mice	Media	—	—	−.0110
na.omit	Media	—	—	−.0345
true	Participation	—	—	.9540
hot.deck	Participation	—	—	−.0138
hd.ord	Participation	—	—	−.0130

amelia	Participation	—	—	–.0111
mice	Participation	—	—	–.0105
na.omit	Participation	—	—	–.0348

Finally, the results for the interval variables confirm previous results, with `amelia` and `mice` demonstrating the best performance. In addition, note that `na.omit` delivers better results than `hd.ord` for `Age` for all datasets and represents the best method for CCES `Age`. Similar to the MNAR results for four variables, the performance of `hd.ord` in the MNAR analysis of 11 variables deteriorates in the switch from one to two ordinal variables: –.1142, –.0473, –.0291 vs. –.1046, –.0421, –.0262 for `Inc` and –.3329, –.2424, –.3271 vs. –.3081, –.2251, –.2989 for `Age`.

3.4.4 Increased Percentage of Missingness

This brief section shows the imputation results when the percentage of missingness is increased. I conduct this here MAR for five variables in the Framing data. The results are shown in Table 3.12. As was to be expected at this point, `hd.ord` displays the worst performance for all percentages for all binary variables. `amelia` and `mice` show virtually identical results. Note that `hot.deck` appears to outperform `amelia` and `mice` for CCES `Dem` and `Male`. `hd.ord` also represents the worst method for all percentages for all ordinal and interval variables. `amelia` and `mice` show virtually identical results and far outperform the other methods for the ordinal variables. Note that `amelia` also consistently outperforms `mice` for all percentages for all interval variables.

3.4.5 Speed

This section shows the running times for all methods. Specifically, I outline the speed differences by the number of imputed variables for all datasets (Table 3.13) and by the percentage of missingness for the Framing data (Table 3.14). Both analyses are conducted MAR, with the latter only for five imputed variables. All running times are

Table 3.10. Accuracy of Multiple Imputation Methods. Two Ordinal Variables (Education, Interest), MNAR, 4 Variables with NA

Method	Variable	ANES	CCES	Framing
true	Dem	.3420	.3770	.4660
hot.deck	Dem	−.0142	−.0133	−.0043
hd.ord	Dem	−.0155	−.0131	−.0042
amelia	Dem	−.0136	−.0131	−.0059
mice	Dem	−.0127	−.0130	−.0046
na.omit	Dem	−.0211	−.0185	−.0207
true	Male	.4890	.4830	.5260
hot.deck	Male	−.0180	−.0162	−.0167
hd.ord	Male	−.0172	−.0160	−.0157
amelia	Male	−.0170	−.0154	−.0170
mice	Male	−.0170	−.0153	−.0172
na.omit	Male	−.0233	−.0241	−.0169
true	Inc	16.6140	6.4810	3.0890
hot.deck	Inc	−.2481	−.1034	−.0728
hd.ord	Inc	−.3174	−.1303	−.0812
amelia	Inc	−.1555	−.0741	−.0589
mice	Inc	−.1568	−.0730	−.0565
na.omit	Inc	−.3114	−.1513	−.0852
true	Age	50.0410	52.8230	37.9120
hot.deck	Age	−.6020	−.4844	−.8613
hd.ord	Age	−.8997	−.7259	−.9349
amelia	Age	−.3103	−.2831	−.3702
mice	Age	−.2994	−.2702	−.3705
na.omit	Age	−.6726	−.6482	−.6213

Note: Each `true` value shows the true variable mean. All other values show the differences between the imputation means and the true mean, indicated with a + or − sign.

given in minutes and apply to all 1,000 amputation/imputation iterations combined.

We can see in Table 3.13 that `hd.ord` and `hot.deck` show virtually identical running times for all datasets for both numbers of imputed variables. This is to be expected as both methods are very similar in terms of their code build-up. More importantly, however, we observe that both methods are much faster than `amelia` and `mice`: `amelia` is at least 3.2 times slower than `hd.ord` (for the Framing data for both numbers of imputed variables).

Table 3.12. Accuracy of Multiple Imputation Methods for Increasing Percentages of Missing Data (Framing Experiment Data)

Method	Variable	20% NA	50% NA	80% NA
true	Dem	.4660	.4660	.4660
hot.deck	Dem	+.0001	+.0001	+.0016
hd.ord	Dem	+.0008	+.0018	+.0037
amelia	Dem	−.0001	−.0003	+.0004
mice	Dem	+.0000	−.0002	+.0004
na.omit	Dem	−.0340	−.0841	−.1393
true	Male	.5260	.5260	.5260
hot.deck	Male	+.0001	+.0000	−.0004
hd.ord	Male	+.0005	+.0009	+.0014
amelia	Male	+.0000	−.0001	−.0003
mice	Male	−.0001	−.0003	−.0006
na.omit	Male	−.0256	−.0632	−.1051
true	Interest	3.2170	3.2170	3.2170
hot.deck	Interest	−.0213	−.0604	−.0908
hd.ord	Interest	−.0248	−.0679	−.1007
amelia	Interest	−.0002	−.0003	+.0003
mice	Interest	−.0003	−.0004	+.0001
na.omit	Interest	−.0714	−.1833	−.3219
true	Inc	3.0890	3.0890	3.0890
hot.deck	Inc	−.0119	−.0371	−.0594
hd.ord	Inc	−.0192	−.0550	−.0805
amelia	Inc	+.0003	+.0001	+.0009
mice	Inc	+.0014	+.0028	+.0039
na.omit	Inc	−.1367	−.3076	−.4850
true	Age	37.9120	37.9120	37.9120
hot.deck	Age	−.3650	−1.0406	−1.5789
hd.ord	Age	−.3923	−1.1319	−1.6962
amelia	Age	−.0039	+.0043	−.0005
mice	Age	−.0049	+.0047	+.0034
na.omit	Age	−1.0641	−2.3051	−3.4885

Note: Each `true` value shows the true variable mean. All other values show the differences between the imputation means and the true mean, indicated with a + or − sign.

The highest difference amounts to 3.9 (for the CCES data with 12 imputed variables). `mice`, however, shows by far the highest running times. At its worst, `mice` is 42.3 times slower than `hd.ord` (for the CCES data with 12 imputed variables). At its best, this deficit still amounts 18.9 times slower (for the Framing data with five imputed variables). Table

Table 3.13. Runtimes of Multiple Imputation Methods (in Minutes) by Number of Imputed Variables

	5 Variables with NA			12 Variables with NA		
	ANES	CCES	Framing	ANES	CCES	Framing
hd.ord	2.632	2.778	2.559	2.614	2.679	2.555
hot.deck	2.628	2.786	2.567	2.640	2.690	2.582
amelia	9.052	10.611	8.079	9.038	10.345	8.109
mice	50.445	57.205	48.436	104.143	113.390	103.953

3.14 shows that `hd.ord` and `hot.deck` remain the fastest methods across all percentages of missingness, but the gap to `amelia` and `mice` narrows as the missingness increases. `amelia` improves from 3.2 times slower than `hd.ord` for 20 percent NA to 1.3 times slower for 80 percent NA. Similarly, `mice` speeds up from 18.9 to 8.7 times slower.

Table 3.14. Runtimes of Multiple Imputation Methods (in Minutes) by Percentage of Missingness (Framing Experiment Data)

Method	20% NA	50% NA	80% NA
hd.ord	2.560	11.750	28.170
hot.deck	2.570	12.050	29.390
amelia	8.080	21.150	36.780
mice	48.440	137.930	244.680

3.5 Conclusion

Make sure to include in the MNAR discussion that MNAR is next to impossible to prove/definitively assert 'in the wild'

–> mice.pdf: 6.2. Sensitivity analysis under MNAR

Amelia is very finicky when it comes to collinearity `optimizeSD` in `hot.deck`: Logical indicating whether the sdCutoff parameter should be optimized such that the smallest possible value is chosen that produces no thin cells from which to draw donors. Thin cells are those where the number of donors is less than m –> potential explanation why `hot.deck` isn't better than the others could be something to do with the warnings saying there aren't enough donors

CHAPTER 4

MORAL ARGUMENTS AS A SOURCE OF FRAME STRENGTH

4.1 Introduction

Barack Obama presented the first outline of the Affordable Care Act in the summer of 2009. The content of the reform was put online for everyone to see, but since the administration was still working on details, it refrained from actively communicating it. Published press releases simply stated that the ACA would expand coverage and lower health care costs for everyone. This hesitancy turned out to be a big mistake. At the end of July, support for the ACA hovered around 43 percent. Then Sarah Palin, John McCain's choice for running mate in 2008, posted the following statement on Facebook on August 7$^{\text{th}}$: "The America I know and love is not one in which my parents or my baby with Down Syndrome will have to stand in front of Obama's 'death panel' so his bureaucrats can decide, based on a subjective judgment of their 'level of productivity in society' " (Palin, 2009). Palin implied that federal government workers would be able to refuse treatment to any patients and thus 'decide their fate'. Over the next two weeks, support for the ACA dropped to 35 percent while opposition rose to 52 percent. Republican lawmakers jumped at the opportunity and repeated the claim of 'death panels' whenever possible. The reform never recovered from this drop. In December 2009, four months after the statement, support and opposition were virtually identical to August. While the public was still uncertain about the exact contents of the law, Palin had asserted that it would include a Big Brother type panel that decided whether people would live or die. This drowned out any efforts by the Obama administration to show the law as a cost-reducing reform. Palin's frame of the ACA, in other words, drastically influenced public opinion of the reform.

Framing is the practice of presenting an issue to affect the way people see it (Aaroe, 2011; Druckman, 2001a; Gross, 2008). We learn about healthcare reform through articles, reports, speeches, commercials and social media. This mediated communication possesses tremendous potential influence on our perception of political issues (Iyengar, 1996; Kam & Simas, 2010; Tversky & Kahneman, 1981). Framing research has established that a variety of frames substantively influence how people view and think about issues, such as the ACA (Andsager, 2000; Callaghan & Schnell, 2005; Entman, 1993, 2004; Gamson & Modigliani, 1989; Lahav & Courtemanche, 2012; Pan & Kosicki, 1993; Price, Tewksbury, & Powers, 1997; Slothuus & Vreese, 2010; Sniderman & Theriault, 2004; Vreese, 2004). But we do not know why these frames elicit these effects. A major challenge for framing research thus "concerns the identification of factors that make a frame strong" (Chong & Druckman, 2007, p. 116).

I conduct a series of tests to examine whether moral arguments form a part of what makes a frame strong. These tests include two online polls and an online survey experiment. Both methods developed in chapters II and III are applied in the survey experiment and their performance is analyzed.

4.2 Theory

4.2.1 Framing

Despite the mass of experimental framing research, we still have little insight into what makes a frame strong. The larger persuasion literature "is not as illuminating as one might suppose... It is not yet known what it is about the 'strong arguments'... that makes them persuasive" (O'Keefe, 2002). One research direction is that frames are stronger overall when they cohere with an individual's personal value system. Feinberg & Willer (2012) frame environmental issues as a matter of 'purity', a theme that supposedly correlates with conservative ideology, and find this approach leads to increased conservative support of environmental policies. (Arceneaux, 2012, p. 280) on the other hand finds that "individuals are more likely to be persuaded by political arguments that evoke cognitive biases". Particularly, he asserts that messages which highlight out-group threats resonate to a greater extent than other, more coherent, arguments. In a study investigating the use of scientific data, Druckman & Bolsen (2011) report that adding factual information to messages about carbon nanotubes does nothing to enhance their strength. Providing more scientific evidence seems to have the opposite effect, making the messages weaker. Overall, "it seems as if frame strength increases with frames that highlight specific emotions, invoke threat against one's own group interests, contain some incivility, include multiple, frequently appearing, arguments, and/or have been used in the past" (Druckman, Klar, Robison, & Gubitz, 2018, p. 22). I attempt to provide an avenue of clarification by testing whether moral arguments are part of what makes frames strong.

4.2.2 Moralization

Moralization literature conceptually defines moral arguments as (1) near-universal standards of truth, (2) almost objective facts about the world, and (3) independent of institutional authority (Skitka, 2010). Moral arguments are said to engage a distinctive mode of processing that invokes a whole range of emotions and to be distinct from other forms of arguments (Ryan, 2014b). Scholars find that moral arguments are ubiquitous in political issues because they are essential to how people perceive and make sense of the world around them (Frank, 2005; Mooney, 2001; Tatalovich, Smith, & Bobic, 1994). Ryan (2014b) finds evidence that some people perceive distinctly economic issues such as labor relations laws or social security reform in moral ways. Other studies similarly assert that the strength of attitudes meaningfully differs when they are held with moral conviction (Baron & Spranca, 1997; Bennis, Medin, & Bartels, 2010; Ditto, Pizarro, & Tannenbaum, 2009; Tetlock, 2003). It is also asserted that moral conviction represents an important force that guides citizen behavior and development of public opinion (Converse, 1964; Skitka, Bauman, & Sargis, 2005; Skitka & Wisneski, 2011; Smith, 2002; Tatalovich & Daynes, 2011; Zaller, 1992). It is widely argued that people rely to a disproportionate extent on moral arguments to form their opinions and apply this moralization to political issues (Ryan, 2014a, 2014b; Smith, 2002). Moral arguments can achieve a much higher emotional connection with people because they invoke people's values and feelings (Haidt, 2003; Skitka et al., 2005; Skitka & Wisneski, 2011; Tatalovich & Daynes, 2011). These conceptual definitions are all encompassed in Moral Foundations Theory (MFT), developed by Haidt (2012) and presented in Table 4.1 below. These aspects of moralization

Table 4.1. Foundations of Moral Arguments

Positive		**Negative**	
Care	Cherishing, protecting others	*Harm*	Hurting others
Fairness	Rendering justice by shared rules	*Cheating*	Flouting justice, shared rules
Loyalty	Standing with your group	*Betrayal*	Opposing your group
Respect	Submitting to tradition, authority	*Subversion*	Resisting tradition, authority
Sanctity	Repulsion at disgust	*Degradation*	Enjoyment of disgust

Based on Haidt (2012). Positive and negative foundations are conceptual opposites.

theory have not been applied to frame strength in experimental research. An abundance of framing research has shown that frames elicit significant changes in issue positioning (Chong & Druckman, 2010, 2013; Druckman, 2001b; Druckman, Fein, & Leeper, 2012; Druckman et al., 2013; Nelson, Clawson, & Oxley, 1997; Slothuus, 2008). Brewer & Gross (2005), for instance, find significant effects for the frames 'School vouchers create an unfair advantage' and 'School vouchers provide help for those who need it'. Druckman et al. (2012) provide similar evidence for 'The Affordable Care Act gives more people equal access to health insurance' and 'The ACA increases government costs', while Druckman et al. (2013) do so for 'Oil drilling provides economic benefits' and 'Oil drilling endangers marine life'. While we know these frames elicit changes, we do not know why they do so. We do not know why these frames 'work'. I propose a way to understand why some of these frames work, which is based on moralization theory: The presence of moral arguments.

Applying Moral Foundation Theory, both frames in Brewer & Gross (2005) could be categorized as containing moral arguments, even though the authors do not explicitly do so. 'School vouchers create an unfair advantage' can be argued to contain the negative moral foundation of *Cheating*, while 'School vouchers provide help for those who need it' contains the positive moral foundations of *Care* and *Fairness*. 'The ACA gives more people equal access to health insurance' (Druckman et al., 2012) also could be said to contain the positive moral foundations of *Care* and *Fairness*, while 'Oil drilling endangers marine life' (Druckman et al., 2013) contains the negative moral foundation of *Harm*.

It is important to note that 'The ACA increases government costs' (Druckman et al., 2012) and 'Oil drilling provides economic benefits' (Druckman et al., 2013) do not directly appeal to morality, yet research has shown them to be strong. Of course, some people might see increasing costs immediately as bad and thus morally detrimental for the future of the country, but these frames do not make such an appeal directly, on the surface. This distinguishes them from moral frames.

This might lead one to assert that moral arguments do not form a part of frame strength – after all, if a frame does not contain a direct moral argument but is proven to be strong nonetheless, surely then frame strength does not depend on the presence of moral arguments. This hypothetical argument is flawed, however. For one, the search for the source of frame strength is not the search for a universal 'holy grail' argument whose presence is a precondition for frame strength. There probably are many aspects that can make a frame strong, with moral arguments potentially being one of them, not the only one. Second, the studies with these two amoral and two moral arguments do not distinguish between the directions in which the moral and amoral frames act.

In the case of Druckman et al. (2012), 'The Affordable Care Act gives more people equal access to health insurance' contains a positive, i.e. support-inducing, moral argument, while 'The ACA increases government costs' contains a negative, i.e. opposition-inducing, amoral argument. They act in opposite directions. A comparison of these frames alone does not yield sufficient results as we would not be able to identify the exact cause of the framing effect. Is 'The Affordable Care Act gives more people equal access to health insurance' strong because it supports the issue or because it contains a moral argument? Similarly, is 'The ACA increases government costs' strong because it opposes the issue or because it contains a amoral argument? This set-up cannot answer these questions.

To establish whether moral arguments are part of what makes a frame strong, we need a design that assesses the strength of frames with moral and amoral arguments whilst accounting for both signs, opposing and supporting, in both sets of frames. I provide such a design.

Overall, experimental framing research has shown that many frames have significant framing effects. It is still unclear, however, what makes these, or indeed any, frame strong (Druckman et al., 2018). Moralization theory claims that moral arguments possess enormous power shape human behavior and influence public opinion (Haidt, 2003). I combine these two sets of literature and analyze whether moral arguments form a part of what makes frames strong. I use a variety of data sources to investigate the following hypothesis:

H. Moral arguments form a part of what makes political frames strong.

4.3 Data

First, I field an online poll that asks participant to assess how moral frames in published research are. Second, I design and test moral and amoral complementary frames to the ones used in previous research in a second online poll. Finally, I conduct the main online survey experiment.

Testing and pre-testing frames is crucial in frame analysis. Frames need to be tested with participants who are not part of the main survey. These participants are exposed to the designed frames and asked whether they reflect the core ideas behind moral and pragmatic arguments. This pre-test structure builds on work by Slothuus & Vreese (2010) and Chong & Druckman (2007), following the mass communication and persuasion literature (O'Keefe, 2002). The pre-test is carried out on Amazon's online platform MTurk. Together with the post-test included in the survey, these tests represent a thorough, robust, spread-out safety net to ensure that the designed frames connect with participants, which in turn ensures meaningful survey results.

4.3.1 Online Poll #1: The Moral Content of Frames Used in Previous Studies

I test frames from previous in a simple online poll, which asks participants to rate how moral each argument in each frame is on a scale of 1 (Not moral at all) to 5 (Very moral). This provides me with a list of how moral or amoral the strong frames from previous research. The poll is fielded on MTurk. Survey questions are best designed if researchers have a firm grasp of the underlying realities that participants have to report (Carpini & Keeter, 1993; Conover, Crewe, & Searing, 1991; Stanley, 2016).

Because the design in previous experimental framing studies does not account for direction (see section 4.2.2), this list is a mixture of positive moral frames, negative moral frames, positive amoral frames, and negative amoral frames. Crucially, there will not be 'perfect pairings' that account for direction of support and moral content. To fill these missing 'slots', I design moral and amoral frames that complement the existing moral and amoral frames. Depending on what the missing 'slot' is, these frames are either positive moral, negative moral, positive amoral, or negative amoral. In order to do so, however, I first conduct another online poll, insights of which inform the design of said frames.

4.3.2 Online Poll #2: The Moral Content of Designed Complementary Frames

I use the insights from the second online poll to design several frames for each of the missing 'slots' in previous experiments. I develop several frames for each missing 'slot'. Once I have designed these frames, I will conduct another simple online poll which asks participants to rate how moral each argument in each designed frame is on a scale of 1 (Not moral at all) to 5 (Very moral). Similar to the online poll in section 4.3.1, this assesses the moral content of the frames I designed. This is to make sure that the designed frames connect with people in terms of moral content in the intended way. Like the previous poll, this poll is also fielded on MTurk.

4.3.3 Survey Experiment

Finally, I design a questionnaire for a survey experiment that combines all the insights from the previous tests and thoroughly examines whether moral arguments form a part of what makes a frame strong. The experiment is fielded online for a random sample of U.S. adults on Lucid. It features frames with moral and amoral arguments that have been shown to elicit effects (tested for moral content in the first online poll) and frames with moral and amoral arguments that fill missing ‘slots’ (designed based on insights from the second online poll). The survey also collects demographic information and includes post-treatment evaluations in terms of rating the arguments by their moral content. The political issues the frames apply are determined by the issues used in previous framing experiments. Both methodological contributions from chapters II and III are applied.

Participants are separately sequentially blocked into five treatment groups. The blocking algorithm is performed once on the original ANES edcuation categories for one half of participants and once on the ordered probit education categories for the other half. We then conduct an ordered probit regression on an ordinal response variable on a 5-point Likert scale, ranging from “Strongly oppose” to “Strongly support”. We then compare the differences between the ordered probit and the original results whilst knowing which one is closer to the truth based on the simulation results from section 2.3. This thus tests the performance of the ordered probit method (chapter II).

Participants are not given the option to select “Don’t Know” or “Refuse”, resulting in completely observed data. We then introduce random missing data, and show how the ordinal affinity score method (chapter III) performs on ordinal variables compared with other methods. We can assess the performance of the ordinal affinity score method because we know the true values of the completely observed data.

4.4 Results

4.5 Conclusion

CHAPTER 5

CONCLUSION

Education is the most important predictor of political behavior in political science. It is crucial that we measure and use this variable correctly to obtain results that reflect the true data structure. As an ordinal variable, education contains special characteristics: Its categories are ordered, but unevenly spaced. We need modern statistical methods to fully utilize all this information contained in this variable. So far, this aspect has been largely ignored in the literature. If we want to know what people think and how they act, we need to make sure our measurements are as good as they can possibly be. My dissertation outlines two new methods that contribute to this undertaking. They significantly improve how we handle ordinal variables in surveys and survey experiments in political science and thus increase precision when we analyze public opinion.

Whether my methods are suitable in a specific survey or survey experiment depends on the situation, as no method works for all circumstances. For a survey experiment with a very large sample and few treatment groups, there is no need for my ordered probit method and blocking. Simple randomization does the job here. Similarly, if survey results do not contain important ordinal predictor variables, my method to impute missing data from ordinal variables is not applicable. Overall, my dissertation adds two important new tools to the empirical political scientist's toolbox to choose from.

APPENDIX A

ORDINAL BLOCKING

In order to conduct this experiment, I created an online survey environment based on `R` with `shiny` (Boas & Hidalgo, 2013), as there currently is no available tool to block sequentially online. Popular online survey platforms, such as Qualtrics, do not have offer this functionality, and none of the attempts to combine `R` code work with Qualtrics concern the 'injection' of `R` code into the Qualtrics randomization engine, which blocking would require (Barari et al., 2017; Ginn, 2018; Hainmueller, Hopkins, & Yamamoto, 2014; Testa, 2017). The following is a basic outline of the mechanisms behind this survey environment.

The survey questions, i.e. questions that collect demographic information and questions that apply treatment, need to be designed as `.txt` files and incorporated into a local `shiny` environment. This local environment is then hosted in the cloud and publicly accessible. The hosted website sequentially blocks each incoming participant based on her covariate information and covariate information from all previous participants through constant interaction with the `R` code. The workflow for any incoming participant is illustrated in Figure A.1 below.

A participant clicks on the survey link and answers the demographic question. After she selects her level of education, `R` code in the background pulls previous participants' covariate information from a Dropbox server. Based on this information and her chosen

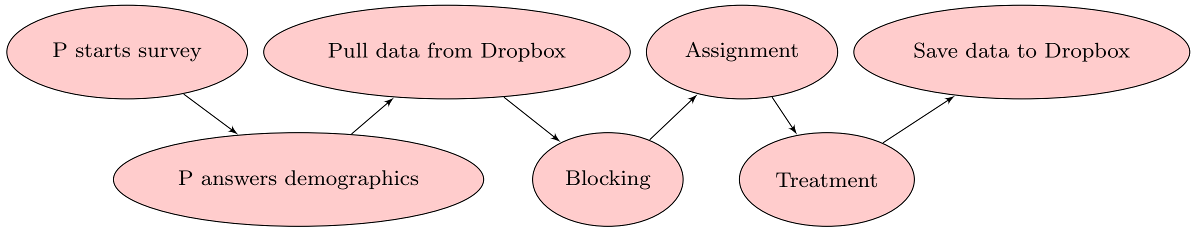

Figure A.1. Online survey experiment workflow

education level, the R code sequentially blocks and assigns her to a treatment group. The participant then sees and answers the respective treatment question(s). Her responses are then saved on the same Dropbox server. This process is repeated for all incoming participants. If the participant is the first person to take the survey, i.e. if there is no covariate information from previous participants yet, the code randomly assigns her to one of the treatment groups. All subsequent participants are then blocked and assigned as just described.

To recruit participants, the cloud-based website can easily be linked to online market platforms, such as MTurk. MTurk is a service where researchers can host tasks to be completed by anonymous participants. Participants receive financial compensation for their work and Amazon collects a commission. MTurk samples have been shown to be internally valid in survey experiments (Berinsky, Huber, & Lenz, 2012). The use of MTurk in political science experiments has increased dramatically over the past decade and is now common practice (Hauser & Schwarz, 2016). I use Lucid for my experiment, which has been shown to be equally reliable and performs well on a national scale in survey experiments (Coppock & McClellan, 2019).

APPENDIX B

ORDINAL MISSING

B.1 All Observations

ANES all obs. 1000 iterations, CCES all obs. 10 iterations (maxed out RAM) 5 variables: Dem, Male, Interest, Inc, Age MAR (Table B.1) and MNAR (B.2)

Table B.1. Accuracy of Multiple Imputation Methods. MAR, 5 Variables with NA, all observations

Method	Variable	ANES	CCES
true	Dem	.3370	.4032
hot.deck	Dem	−.0001	+.0006
hd.ord	Dem	−.0002	+.0006
amelia	Dem	+.0000	+.0003
mice	Dem	+.0000	+.0003
na.omit	Dem	−.0302	−.0250
true	Male	.4868	.4521
hot.deck	Male	−.0004	−.0001
hd.ord	Male	−.0008	−.0002
amelia	Male	+.0001	+.0001
mice	Male	+.0001	+.0001
na.omit	Male	−.0365	−.0436
true	Interest	2.8806	3.3301
hot.deck	Interest	−.0087	−.0033
hd.ord	Interest	−.0135	−.0046
amelia	Interest	+.0001	+.0001
mice	Interest	+.0000	+.0000
na.omit	Interest	−.0741	−.0763
true	Inc	16.6894	6.5830
hot.deck	Inc	−.0606	−.0009
hd.ord	Inc	−.1030	−.0073
amelia	Inc	+.0009	−.0021
mice	Inc	−.0007	−.0022
na.omit	Inc	−.5574	−.2592
true	Age	50.3745	52.8639
hot.deck	Age	−.2355	−.0221
hd.ord	Age	−.3698	−.0790
amelia	Age	+.0056	−.0006
mice	Age	+.0053	−.0132
na.omit	Age	−1.2785	−1.2190
Observations		2395	42205
Iterations		1000	10

Note: Due to the very high number of observations, the CCES data can only be run for a low number of iterations. Anything above that maxes out the 120 GB of RAM available to me. The framing data are not included since their total number of observations (1,003) is virtually identical to the sample of 1,000.

Table B.2. Accuracy of Multiple Imputation Methods. MNAR, 5 Variables with NA, all observations

Method	Variable	ANES	CCES
true	Dem	.3370	.4032
hot.deck	Dem	−.0109	−.0105
hd.ord	Dem	−.0113	−.0105
amelia	Dem	−.0110	−.0109
mice	Dem	−.0103	−.0106
na.omit	Dem	−.0185	−.0145
true	Male	.4868	.4521
hot.deck	Male	−.0138	−.0129
hd.ord	Male	−.0136	−.0131
amelia	Male	−.0134	−.0132
mice	Male	−.0134	−.0131
na.omit	Male	−.0196	−.0244
true	Interest	2.8806	3.3301
hot.deck	Interest	−.0257	−.0171
hd.ord	Interest	−.0299	−.0179
amelia	Interest	−.0178	−.0147
mice	Interest	−.0179	−.0148
na.omit	Interest	−.0407	−.0418
true	Inc	16.6894	6.5830
hot.deck	Inc	−.1874	−.0691
hd.ord	Inc	−.2287	−.0696
amelia	Inc	−.1292	−.0637
mice	Inc	−.1297	−.0634
na.omit	Inc	−.2729	−.1375
true	Age	50.3745	52.8639
hot.deck	Age	−.5240	−.2331
hd.ord	Age	−.6609	−.2764
amelia	Age	−.2533	−.2342
mice	Age	−.2474	−.2371
na.omit	Age	−.7188	−.6476
Observations		2395	42205
Iterations		1000	10

Note: Due to the very high number of observations, the CCES data can only be run for a low number of iterations. Anything above that maxes out the 120 GB of RAM available to me. The framing data are not included since their total number of observations (1,003) is virtually identical to the sample of 1,000.

B.2 Increased Missingness

CCES 10,000 iterations +++ 5 variables: Dem, Male, Interest, Inc, Age MAR 20, 50, 80 percent

B.3 Speed

CCES 10,000 iterations +++ 5 variables: Dem, Male, Interest, Inc, Age MAR 20, 50, 80 percent As can be seen in Table B.4, `hd.ord` and `hot.deck` show virtually identical imputation times for the old framing data (n = 1,003), with `hd.ord` being 12 seconds faster. `amelia`, however, is 2.3 times slower than `hd.ord`. `mice` is 17.9 times slower than `hd.ord`. This is a dramatic speed gain. The ANES (n = 3,223) data show only a reduced speed gain. `mice` is still drastically slower than `hd.ord` (by a magnitude of 9), but the speed gain is cut in half. The previous speed gain over `amelia` is no longer observable. `hd.ord` and `amelia` now show virtually identical runtimes (with a difference of 25 seconds in favor of `hd.ord`).

Table B.3. Runtimes of Multiple Imputation Methods (in Minutes) with All Available Observations

Method	ANES	CCES
hd.ord	9.406	21.829
hot.deck	9.374	21.429
amelia	12.393	2.602
mice	99.455	100.253
Observations	2395	42205
Iterations	1000	10

Note: Due to the very high number of observations, the CCES data can only be run for a low number of iterations. Anything above that maxes out the 120 GB of RAM available to me. The framing data are not included since their total number of observations (1,003) is virtually identical to the sample of 1,000.

Table B.4. Runtimes of Multiple Imputation Methods (in Minutes)

	OldFraming	ANES2016	ANES2016_5lev	OldFramingQuadrObs
hd.ord	34.457	32.642	32.668	31.519
hd.norm	34.660	32.669	32.881	31.810
amelia	77.956	33.051	33.691	30.035
mice	616.023	293.722	298.919	277.021
Observations	1,003	3,223	3,145	4,012
Iterations	12,500	2,396	2,500	1,500
Levels	7	16	5	7

Three factors might explain this sudden and surprising increase in the performance of `amelia`: The levels of the ordinal variable in question (`education`), the number of iterations, and the number of observations in a data set. The ANES (n = 3,223) data contain 17 levels of `education`, whereas the old framing data (n = 1,003) contain only 7. However, when run with 5 levels of `education` and otherwise virtually identical data, the method performances remain unchanged (see column "ANES2016_5lev"). This rules out the levels of the ordinal variable. Another explanation could be the number of observations and the corresponding possible number of iterations. A high number of observations reduces the computationally feasible number of iterations that can be performed before even powerful machines with 120 GB RAM are maxed out. The old framing data (n = 1,003) contain 1,003 observations and can be computationally run for 12,500 iterations. The 2016 ANES data (n = 3,223) contain more than triple the observations than the old framing data (n = 1,003) and can only be run for 2,396 iterations. Indeed, when we quadruple the number of observations in the old framing data (n = 4,012) and are thus forced to reduce the number of possible iterations to 1,500, we observe that `amelia` is now the fastest method (see column "OldFramingQuadrObs"). It thus appears that `amelia` is fast with a low number of iterations and slow with a high number of iterations.

This would lead us to predict that `amelia` should be the fastest method when the old framing data (n = 1,003) is run for a low number of iterations (2,000). That is not the case: `amelia` is 2.2 times slower than `hd.ord` (see column "OldFraming_2000it"); on the same level as the 12,500-iteration-run of the old framing data (n = 1,003). This means it's not the number of iterations but the number of observations that affects `amelia`'s relative lack of speed compared to `hd.ord`. `amelia` appears to be much slower than `hd.ord` for around 1,000 observations but on equal footing for 3,000+ observations.

This is indeed confirmed by Table B.5, where the old framing, the ANES, and the CCES data have all been reduced to 1,000 observations. The left half of the table shows the results for 10,000 iterations and low numbers of variables with NAs (Note: `education` in the ANES data was reduced to 7 levels to make this number of iterations computationally feasible). The right half of the table shows the results for 1,000 iterations and high numbers of variables with NAs. `amelia` is consistently between 2.1 and 2.9 times slower than `hd.ord` for all three data sets, regardless of the number of iterations and the number of variables with NAs.

Table B.5. Runtimes of Multiple Imputation Methods (in Minutes), 1000 Observations

	10,000 Iterations			1,000 Iterations		
	CCES	OldFraming	ANES	CCES	OldFraming	ANES
hd.ord	24.394	26.461	24.294	2.714	2.588	4.103
hd.norm	24.540	26.620	24.316	2.723	2.617	4.300
amelia	57.403	68.862	50.630	7.787	6.926	8.758
mice	449.027	527.532	390.690	158.583	116.609	336.038
Ordinal Levels	6	7	7	6	7	7
Variables with NAs	5	5	5	16	13	22

B.4 With Modified `ampute()`

Table B.6 shows the results of using `ampute()` with the options `bycases=FALSE` and `cont=FALSE` on the old framing data (n = 1,003) and inserts NAs MAR for 5 variables at

20 percent for 9,644 iterations. `hd.ord` overall performs worse in this scenario.

Table B.6. Accuracy of Multiple Imputation Method. ampute() with bycases, old framing data (n = 1,003)

Method	Variable	Value	Diff
true	Dem	.4666	.0000
hd.ord	Dem	.4674	.0008
hd.norm	Dem	.4686	.0020
amelia	Dem	.4669	.0003
mice	Dem	.4667	.0001
na.omit	Dem	.2363	.2303
true	inc	3.0927	.0000
hd.ord	inc	3.0466	.0461
hd.norm	inc	3.0245	.0682
amelia	inc	3.0941	.0014
mice	inc	3.0973	.0046
na.omit	inc	2.4208	.6719
true	age	37.9252	.0000
hd.ord	age	36.5063	1.4189
hd.norm	age	36.3438	1.5814
amelia	age	37.9287	.0035
mice	age	37.9398	.0146
na.omit	age	31.6958	6.2294
true	Female	.4666	.0000
hd.ord	Female	.4619	.0047
hd.norm	Female	.4600	.0066
amelia	Female	.4663	.0003
mice	Female	.4666	.0000
na.omit	Female	.2168	.2498
true	interest	3.2164	.0000
hd.ord	interest	3.1362	.0802
hd.norm	interest	3.1223	.0941
amelia	interest	3.2150	.0014
mice	interest	3.2147	.0017
na.omit	interest	2.7280	.4884

B.5 With My Own Amputation Methods

Table B.7 shows the results of using my own function, `own.NA()`, to insert 20 percent NAs MAR into three variables in the old framing data (n = 1,003). In general, something is MAR if you ampute values in column A based on values in column B, e.g. if you ampute the values for `age` where `income = 1` and where `income = 5`. `own.NA()` applies this procedure for any combination of columns. Here, the chosen columns to be amputed are `Dem`, `age`, and `interest`. The chosen columns the amputations depend on are `inc`, `Female`, and `Black`. The function samples 20 percent of observations for each unique value of `inc`. For those observations, the values of `Dem` are amputed. Accordingly, the function samples 20 percent of observations for each unique value of `Female`. For those observations, the values of `age` are amputed. The same occurs for `Black` and `interest`. The resulting data frame is then imputed. Note that the number of amputed and imputed variables is generally lower, as a pair of variables is needed to ampute one variable. As Table B.7 shows, `hd.ord` does not perform particularly well. It beats `hd.norm` for all three variables but falls considerably short of `amelia` and `mice`. It is notable, however, that my method seems closer to being MCAR than MAR, as `na.omit` performs well and sometimes even outperforms other methods, for instance for `interest`. It is thus questionable how much use `own.NA()` is in its current form. `ampute()` seems to spread NAs evenly across columns. This means that observations are mostly complete, with not more than one or two missing values. I wrote another function, `own.NA.rows()`, that changes this. `own.NA.rows()` inserts missingness for a percentage of observations MAR across all columns except `education`. This means that the majority of observations are complete but a percentage of observations misses data on almost all variables. Table B.8 shows the results, with the missingess percentage set to 20 and NAs inserted into 17 variables.

Table B.7. Accuracy of Multiple Imputation Methods. own.NA(), old framing data (n = 1,003)

Method	Variable	Value	Diff
true	Dem	.4666	.0000
hd.ord	Dem	.4695	.0029
hd.norm	Dem	.4721	.0055
amelia	Dem	.4667	.0001
mice	Dem	.4669	.0003
na.omit	Dem	.4668	.0002
true	age	37.9252	.0000
hd.ord	age	36.1537	1.7715
hd.norm	age	35.8843	2.0409
amelia	age	37.9245	.0007
mice	age	37.9371	.0119
na.omit	age	37.9221	.0031
true	interest	3.2164	.0000
hd.ord	interest	3.1160	.1004
hd.norm	interest	3.1012	.1152
amelia	interest	3.2166	.0002
mice	interest	3.2156	.0008
na.omit	interest	3.2165	.0001

Table B.8: Accuracy of Multiple Imputation Methods. own.NA.rows(), old framing data (n = 1,003)

Method	Variable	Value	Diff
true	Dem	.4666	0
hd.ord	Dem	.4666	0
hd.norm	Dem	.4667	.0001
amelia	Dem	.4667	.0001
mice	Dem	.4670	.0004
na.omit	Dem	.4667	.0001
true	Ind	.2802	0
hd.ord	Ind	.2795	.0007
hd.norm	Ind	.2801	.0001
amelia	Ind	.2801	.0001
mice	Ind	.2817	.0015
na.omit	Ind	.2801	.0001
true	Cons	.2832	0
hd.ord	Cons	.2830	.0002

hd.norm	Cons	.2831	.0001
amelia	Cons	.2831	.0001
mice	Cons	.2823	.0009
na.omit	Cons	.2831	.0001
true	Lib	.5174	0
hd.ord	Lib	.5182	.0008
hd.norm	Lib	.5176	.0002
amelia	Lib	.5176	.0002
mice	Lib	.5173	.0001
na.omit	Lib	.5176	.0002
true	Black	.0698	0
hd.ord	Black	.0702	.0004
hd.norm	Black	.0698	0
amelia	Black	.0698	0
mice	Black	.0731	.0033
na.omit	Black	.0698	0
true	Hisp	.0548	0
hd.ord	Hisp	.0546	.0002
hd.norm	Hisp	.0548	0
amelia	Hisp	.0548	0
mice	Hisp	.0589	.0041
na.omit	Hisp	.0548	0
true	White	.7717	0
hd.ord	White	.7712	.0005
hd.norm	White	.7717	0
amelia	White	.7717	0
mice	White	.7613	.0104
na.omit	White	.7717	0
true	Asian	.0808	0
hd.ord	Asian	.0812	.0004
hd.norm	Asian	.0808	0
amelia	Asian	.0809	.0001
mice	Asian	.0835	.0027
na.omit	Asian	.0808	0
true	Female	.4666	0
hd.ord	Female	.4665	.0001
hd.norm	Female	.4665	.0001
amelia	Female	.4665	.0001
mice	Female	.4673	.0007
na.omit	Female	.4665	.0001
true	Unempl	.1615	0
hd.ord	Unempl	.1610	.0005
hd.norm	Unempl	.1616	.0001

amelia	Unempl	.1616	.0001
mice	Unempl	.1624	.0009
na.omit	Unempl	.1616	.0001
true	Ret	.0508	0
hd.ord	Ret	.0506	.0002
hd.norm	Ret	.0508	0
amelia	Ret	.0509	.0001
mice	Ret	.0505	.0003
na.omit	Ret	.0509	.0001
true	Stud	.0439	0
hd.ord	Stud	.0446	.0007
hd.norm	Stud	.0439	0
amelia	Stud	.0439	0
mice	Stud	.0466	.0027
na.omit	Stud	.0439	0
true	interest	3.2164	0
hd.ord	interest	3.2161	.0003
hd.norm	interest	3.2163	.0001
amelia	interest	3.2164	0
mice	interest	3.2122	.0042
na.omit	interest	3.2163	.0001
true	media	1.7268	0
hd.ord	media	1.7294	.0026
hd.norm	media	1.7270	.0002
amelia	media	1.7270	.0002
mice	media	1.7259	.0009
na.omit	media	1.7269	.0001
true	part	.9561	0
hd.ord	part	.9575	.0014
hd.norm	part	.9560	.0001
amelia	part	.9561	0
mice	part	.9546	.0015
na.omit	part	.9561	0
true	inc	3.0927	0
hd.ord	inc	3.0906	.0021
hd.norm	inc	3.0926	.0001
amelia	inc	3.0926	.0001
mice	inc	3.0949	.0022
na.omit	inc	3.0926	.0001
true	age	37.9252	0
hd.ord	age	37.9000	.0252
hd.norm	age	37.9250	.0002
amelia	age	37.9243	.0009

mice	age	37.8789	.0463
na.omit	age	37.9250	.0002

`hd.norm`, `amelia`, and `na.omit` perform very well. No difference to the true value for any variable is greater than .0009 and most are .0002.

`hd.ord` performs on similar levels except for the nominal variables (`media` (.0026), `part` (.0014), `inc` (.0021), `age` (.0252)).

Somewhat surprisingly, `mice` performs worst overall for many variables (`Ind` (.0015), `Black` (.0033), `Hisp` (.0041), `White` (.0104), `Asian` (.0027), `Stud` (.0027), `interest` (.0042), `part` (.0015), `inc` (.0022), `age` (.0463)) by some margin.

It is also noteable how often the difference amounts to zero and how well `na.omit` performs overall. Overall, `own.NA` and `own.NA.rows` result in a strong performance of `na.omit`, which should not happen in a missingness mechanism that is supposed to be MAR. The functions thus likely represents a version of MCAR, which puts the usefulness of their imputation results in doubt.

REFERENCES

Aaroe, L. (2011). Investigating Frame Strength: The Case of Episodic and Thematic Frames. *Political Communication*, *28*(2), 207–226.

Abramowitz, A. I. (2010). *The Disappearing Center: Engaged Citizens, Polarization, and American Democracy*. New Haven, CT: Yale University Press.

Agresti, A. (1990). *Categorical data analysis*. Hoboken, NJ: Wiley-Interscience.

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. Hoboken, NJ: Wiley-Interscience.

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

Allison, P. D. (2002). *Missing Data*. Thousand Oaks, CA: SAGE Publications.

Andridge, R. R., & Little, R. J. (2010). A Review of Hot Deck Imputation for Survey Non-Response. *International Statistical Review*, *78*(1), 40–64.

Andsager, J. (2000). How Interest Groups Attempt to Shape Public Opinion with Competing News Frames. *Journalism and Mass Communication Quarterly*, *77*(3), 577–592.

Arceneaux, K. (2012). Cognitive Biases and the Strength of Political Arguments. *American Journal of Political Science*, *56*(2), 271–285.

Bailar, J. C., & Bailar, B. A. (1997). Comparison of the Biases of the "Hot Deck" Imputation Procedure with an "Equal Weights" Imputation Procedure. *Symposium on Incomplete Data: Panel on Incomplete Data of the Committee on National Statistics, National Research Council*, 422–447.

Barari, S., Berwick, E., Hainmueller, J., Hopkins, D. J., Liu, S., Strezhnev, A., & Yamamoto, T. (2017). *Package "cjoint"*. https://cran.r-project.org/web/packages/cjoint/cjoint.pdf, package version 2.0.6.

Baron, J., & Spranca, M. (1997). Protected Values. *Organizational Behavior and Human Decision Processes*, *70*, 1–16.

Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The Costs and Benefits of Calculation and Moral Rules. *Perspectives on Psychological Science*, *5*, 187–202.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.Com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368.

Boas, T. C., & Hidalgo, F. D. (2013). Fielding Complex Online Surveys Using rApache and Qualtrics. *The Political Methodologist*, *20*(2), 21–26.

Bodner, T. E. (2008). What Improves with Increased Missing Data Imputations? *Structural Equation Modeling*, *15*, 651–675.

Brand, J. P. (1999). *Development, Implementation and Evaluation of Multiple Imputation Strategies for the Statistical Analysis of Incomplete Data Sets* (PhD thesis). Erasmus University, Rotterdam.

Brewer, P., & Gross, K. (2005). Values, Framing, and Citizens' Thoughts About Policy Issues: Effects on Content and Quantity. *Political Psychology*, *26*(6), 929–948.

Brown, R. L. (1994). Efficacy of the Indirect Approach for Estimating Structural Equation Models with Missing Data: A Comparison of Five Methods. *Structural Equation Modeling*, *1*, 287–316.

Buuren, S. van. (2007). Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification. *Statistical Methods in Medical Research*, *16*(3), 219–242.

Buuren, S. van, Brand, J. P., Groothuis-Oudshoorn, C., & Rubin, D. (2006). Fully Conditional Specification in Multivariate Imputation. *Journal of Statistical Computation and Simulation*, *76*(12), 1049–1064.

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67.

Buuren, S. van, Groothuis-Oudshoorn, K., Vink, G., Schouten, R., Robitzsch, A., Doove, L., … Gray, B. (2020). *Package "mice"*. https://cran.r-project.org/web/packages/mice/mice.pdf R package version 3.9.0.

Buuren, S. van, & Oudshoorn, K. (2000). *Multiple Imputation by Chained Equations: MICE V1.0 User's Manual.* TNO Prevention; Health, Leiden.

Callaghan, K., & Schnell, F. (2005). Introduction: Framing Political Issues in American Politics. In K. Callaghan & F. Schnell (Eds.), *Framing American politics* (pp. 1–17). Pittsburgh, PA: University of Pittsburgh Press.

Carpini, M. X. D., & Keeter, S. (1993). Measuring political knowledge: Putting first things first. *American Journal of Political Science*, *37*(4), 1179–1206.

Chong, D., & Druckman, J. (2007). Framing Public Opinion in Competitive Democracies. *American Political Science Review*, *101*(4), 637–655.

Chong, D., & Druckman, J. (2010). Dynamic Public Opinion: Communication Effects Over Time. *American Political Science Review*, *104*(4), 663–680.

Chong, D., & Druckman, J. (2013). Counterframing Effects. *Journal of Politics*, *75*(1), 1–16.

Chow, S.-C., & Chang, M. (2007). *Adaptive design methods in clinical trials.* Boca Raton, FL: Chapman; Hall.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A Comparison of Inclusive and Restrictive Strategies in Modern Missing-Data Procedures. *Psychological Methods*, *6*, 330–351.

Conover, P., Crewe, I., & Searing, D. (1991). The nature of citizenship in the united states and great britain: Empirical comments on theoretical themes. *Journal of Politics*, *53*(3), 800–832.

Converse, P. E. (1964). The Nature of Mass Belief Systems. In D. Apter (Ed.), *Ideology and Discontent* (pp. 206–261). New York, NY: Free Press.

Coppock, A., & McClellan, O. A. (2019). Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey

Respondents. *Research and Politics*, 1–14.

Cox, B. G. (1980). The Weighted Sequential Hot Deck Imputation Procedure. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 721–726.

David, M., Little, R. J., Samuhel, M. E., & Triest, R. K. (1986). Alternative Methods for CPS Income Imputation. *Journal of the American Statistical Association*, *81*(393), 29–41.

Dawood, Y. (2015). Campaign Finance and American Democracy. *Annual Review of Political Science*, *18*, 329–348.

Dempster, A. P., Laird, N., & Rubin, D. (1977). Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm (with Discussion). *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Diggle, P. J., & Kenward, M. G. (1994). Informative Dropout in Longitudinal Data Analysis (with Discussion). *Applied Statistics*, *43*, 49–73.

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated Moral Reasoning. In B. H. Ross, D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Moral Judgment and Decision-Making* (pp. 307–338). San Diego, CA: Academic Press.

Drechsler, J., & Rassler, S. (2008). Does Convergence Really Matter? In C. Shalabh (Ed.), *Recent Advances in Linear Models and Related Areas* (pp. 341–355). Berlin, GER: Springer.

Druckman, J. (2001a). Evaluating Framing Effects. *Journal of Economic Psychology*, *22*, 91–101.

Druckman, J. (2001b). On the Limits of Framing Effects: Who Can Frame? *Journal of Politics*, *63*(4), 1041–1066.

Druckman, J., & Bolsen, T. (2011). Framing, Motivated Reasoning, and Opinions About Emergent Technologies. *Journal of Communication*, *61*, 659–688.

Druckman, J., Fein, J., & Leeper, T. J. (2012). A Source of Bias in Public Opinion Stability. *American Political Science Review*, *106*(2), 430–454. http://doi.org/10.1017/S0003055412000123

Druckman, J., Klar, S., Robison, J., & Gubitz, S. R. (2018). Political Dynamics of Framing. In T. N. Ridout (Ed.), *New Directions in Media and Politics* (2nd ed.). New York, NY: Routledge.

Druckman, J., Peterson, E., & Slothuus, R. (2013). How Elite Partisan Polarization Affects Public Opinion Formation. *American Political Science Review*, *107*(1), 57–79.

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika*, *58*, 403–417.

Efron, B. (1994). Missing Data, Imputation, and the Bootstrap. *Journal of the American Statistical Association*, *89*(426), 463–475.

Eisele, J. R. (1995). Biased coin designs: Some properties and applications. *Institute of Mathematical Statistics Lecture Notes – Monograph Series*, *25*, 48–64.

Entman, R. M. (1993). Framing: Towards Clarification of a Fractured Paradigm. *Journal of Communication*, *43*(4), 51–58.

Entman, R. M. (2004). *Projections of Power: Framing News, Public Opinion, and US Foreign Policy*. Chicago, IL: Chicago University Press.

Epstein, L., & King, G. (2002). The rules of inference. *University of Chicago Law Review*, *69*, 1–133.

Ernst, L. R. (1978). Weighting to Adjust for Partial Nonresponse. *Proceedings of the American Statistical Association*.

Ezzati-Rice, T. M., Johnson, W., Khare, M., Little, R. J., Rubin, D., & Schafer, J. L. (1995). A Simulation Study to Evaluate the Performance of Model-Based Multiple Imputations in NCHS Health Examination Surveys. In *Proceedings of the Annual Research Conference* (pp. 257–266). Washington, DC: Bureau of the Census.

Fay, R. E. (1996). Alternative Paradigms for the Analysis of Imputed Survey Data. *Journal of the American Statistical Association*, *91*, 490–8.

Feinberg, M., & Willer, R. (2012). The moral roots of environmental attitudes. *Psychological Science*, *24*, 56–62.

Fiorina, M. P., & Abrams, S. J. (2009). *Disconnect: The Breakdown of Representation in American Politics*. Norman, OK: University of Oklahoma Press.

Fiorina, M. P., Abrams, S. J., & Pope, J. C. (2011). *Culture War? The Myth of a Polarized Electorate* (3rd ed.). New York, NY: Longman.

Ford, B. L. (1983). *An Overview of Hot-Deck Procedures: Incomplete Data in Sample Surveys* (Vol. 2). New York, NY: Academic Press.

Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models*. Washington, DC: Sage.

Frank, T. (2005). *What's the Matter with Kansas? How Conservatives Won the Heart of America*. New York, NY: Henry Holt.

Fuller, W., & Kim, J. K. (2005). Hot Deck Imputation for the Response Model. *Survey Methodology*, *31*(2), 139–149.

Gamson, W., & Modigliani, A. (1989). Media Discourse and Public Opinion on Nuclear Power. *American Journal of Sociology*, *95*(1), 1–37.

Gelfand, A. E., & Smith, A. F. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, *85*, 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. (2013). *Bayesian data analysis*. (3rd, Ed.). Boca Raton, FL: CRC Press.

Gertheiss, J., & Tutz, G. (2008). Penalized regression with ordinal predictors. *Technical Report Number 015, Department of Statistics, University of Munich*.

Gill, J. (2014). *Bayesian methods: A social and behavioral sciences approach* (3rd ed.). Boca Raton, FL: CRC Press.

Gill, J., & Cranmer, S. J. (2012). We Have to Be Discrete About This: A Non-Parametric Imputation Technique for Missing Categorical Data. *British Journal of Political Science*, *43*, 425–449.

Gill, J., & Witko, C. (2013). Bayesian Analytical Methods: A Methodological Prescription for Public Administration. *Journal of Public Administration Research and Theory*, *23*, 457–494.

Ginn, J. (2018). *Package "qualtRics"*. https://cran.r-project.org/web/packages/qualtRics/qualtRic package version 3.0.

Giorgi, R., Belot, A., Gaudart, J., & Launoy, G. (2008). The Performance of Multiple Imputation for Missing Covariate Data within the Context of Regression Relative Survival Analysis. *Statistics in Medicine*, *27*(30), 6310–6331.

Glynn, R. J., Laird, N., & Rubin, D. (1993). Multiple Imputation in Mixture Models for Nonignorable Nonresponse with Followups. *Journal of the American Statistical Association*, *88*, 984–993.

Graham, J. W., Hofer, S. M., Donaldson, S. I., MacKinnon, D. P., & Schafer, J. L. (1997). Analysis with Missing Data in Prevention Research. In K. Bryant, M. Windle, & S. West (Eds.), *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research* (pp. 325–366). American Psychological Association.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the Usefulness of Data Obtained with Planned Missing Value Patterns: An Application of Maximum Likelihood Procedures. *Multivariate Behavioral Research*, *31*, 197–218.

Graham, J. W., & Schafer, J. L. (1999). On the Performance of Multiple Imputation for Multivariate Data with Small Sample Size. In R. Hoyle (Ed.), *Statistical Strategies for Small Sample Research* (pp. 1–33). Thousand Oaks, CA: Sage.

Grimmer, J. (2015). We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together. *PS: Political Science & Politics*, *48*(1), 80–83.

Gross, K. (2008). Framing Persuasive Appeals: Episodic and Thematic Framing, Emotional Response, and Policy Opinion. *Political Psychology*, *29*(2), 169–192.

Haidt, J. (2003). The Moral Emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences* (pp. 852–870). New York, NY: Oxford University Press.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Pantheon Books.

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, *22*(1), 1–30.

Haitovsky, Y. (1968). Missing Data in Regression Analysis. *Royal Statistical Society*, *30*(1), 67–82.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk Participations Perform Better On Online Attention Checks Than Do Subject Pool Participants. *Behavior Research Methods*, *48*(1), 400–407.

Heitjan, D. F., & Rubin, D. (1991). Ignorability and Coarse Data. *Annals of Statistics*, *19*, 2244–2253.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960.

Honaker, J., Joseph, A., King, G., Scheve, K., & Singh, N. (n.d.). *AMELIA: A Program for Missing Data*. http://gking.harvard.edu/amelia.

Honaker, J., & King, G. (2010). What to Do About Missing Values in Time-Series Cross-Section Data. *American Journal of Political Science*, *54*(2), 561–581.

Honaker, J., King, G., & Blackwell, M. (2012). *Amelia II: A Program for Missing Data.*

Horton, N. J., & Kleinman, K. P. (2007). Much Ado About Nothing: A Comparison of Missing Data Methods and Software to Fit Incomplete Data. *The American Statistician*, *61*(1), 79–90.

Horton, N. J., & Lipsitz, S. R. (2001). Multiple Imputation in Practice: Comparison of Software Packages for Regression Models with Missing Variables. *American Statistician*, *55*, 244–254.

Imai, K. (2018). *Quantitative Social Science.* Princeton, NJ: Princeton University Press.

Imai, K., King, G., & Elizabeth A. Stuart. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society, Series A*, *171*, 481–502.

Imai, K., King, G., & Nall, C. (2009). The essential role of pair-matching in cluster-randomized experiments, with application to the mexican universal health insurance evaluation. *Statistical Science*, *24*, 29–53.

Iyengar, S. (1996). Framing Responsibility for Political Issues. *The ANNALS of the American Academy of Political and Social Science*, *546*(59-70).

Jackman, S. (2000). Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo. *American Journal of Political Science*, *44*(2), 375–404.

Jackman, S., & Spahn, B. (2018). Why does the american national election study overestimate voter turnout? *Political Analysis*, *forthcoming.*

Kaiser, J. (1983). The Effectiveness of Hot-Deck Procedures in Small Samples. *Meeting of the American Statistical Association.*

Kalton, G., & Kish, L. (1981). Two Efficient Random Imputation Procedures. *Proceedings of the Survey Research Methods Section*, 146–151.

Kam, C. D., & Simas, E. N. (2010). Risk Orientations and Policy Frames. *Journal of Politics*, *72*(2), 381–396.

Kim, J. K. (2004). Finite Sample Properties of Multiple Imputation Estimators. *Annals of Statistics*, *32*(2), 766–783.

Kim, J. K., & Fuller, W. (2004). Fractional Hot Deck Imputation. *Biometrika*, *91*(3), 559–578.

King, D. C. (1997). The Polarization of American Parties and Mistrust of Government. In J. S. Nye, P. D. Zelikow, & D. C. King (Eds.), *Why People Don't Trust Government.* Cambridge, MA: Harvard University Press.

King, G., Gakidou, E., Ravishankar, N., Moore, R. T., Lakin, J., Vargas, M., ... Llamas, H. H. (2007). A "politically robust" experimental design for public policy evaluation, with application to the mexican universal health insurance program. *Journal of Policy Analysis and Management*, *26*, 479–509.

King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing Incomplete

Political Science Data: An Alternative Algorithm for Multiple Imputation. *American Political Science Review*, *95*, 49–69.

King, G., Keohane, R., & Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

Kroh, M. (2006). Taking 'Don't Knows' as Valid Responses: A Multiple Complete Random Imputation of Missing Data. *Quality and Quantity*, *40*, 225–244.

Lachin, J. M. (1988). Properties of simple randomization in clinical trials. *Controlled Clinical Trials*, *9*, 312–326.

Lahav, G., & Courtemanche, M. (2012). The Ideological Effects of Framing Threat on Immigration and Civil Liberties. *Political Behavior*, *34*, 477–505.

Lahlrl, P. (2003). On the Impact of Bootstrapping in Survey Sampling and Small Area Estimation. *Statistical Science*, *18*(2), 199–210.

Leighley, J., & Nagler, J. (2014). *Who Votes Now? Demographics, Issues, Inequality, and Turnout in the United States*. Princeton, NJ: Princeton University Press.

Little, R. J., & Rubin, D. (2002). *Statistical Analysis with Missing Data* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

Lucadamoa, A., & Amenta, P. (2014). A new scaling proposal for handling ordinal categorical variables in co-inertia (-pls) analysis. *Procedia Economics and Finance*, *17*, 10–19.

Markaryan, T., & Rosenberger, W. F. (2010). Exact properties of efron's biased coin randomization procedure. *The Annals of Statistics*, *38*(3), 1546–1567.

Marker, D. A., Judkins, D. R., & Winglee, M. (2002). Large-Scale Imputation for Complex Surveys. In R. Groves & D. A. Dillman (Eds.), *Survey Nonresponse* (pp. 329–342). New York, NY: Wiley-Blackwell.

McLachlan, G. J., & Krishan, T. (1997). *The EM Algorithm and Extensions*. New York, NY: Wiley-Blackwell.

Molenberghs, G., & Kenward, M. G. (2007). *Missing Data in Clinical Studies*. New York, NY: Wiley-Interscience.

Mooney, C. Z. (2001). *The Public Clash of Private Values: The Politics of Morality Policy*. Washington, DC: CQ Press.

Moons, K. G., Donders, R., Stijnen, T., & Harrell, F. (2006). Using the Outcome for Imputation of Missing Predictor Values Was Preferred. *Journal of Clinical Epidemiology*, *59*(10), 1092–1101.

Moore, R. T. (2012). Multivariate continuous blocking to improve political science experiments. *Political Analysis*, *20*, 460–479.

Moore, R. T., & Moore, S. A. (2013). Blocking for Sequential Political Experiments. *Political Analysis*, *21*(4), 507–523.

NCES. (2002). *NCES Statistical Standards*. National Center for Education Statistics.

Nelson, T., Clawson, R., & Oxley, Z. (1997). Media Framing of a Civil Liberties Conflict and Its Effect on Tolerance. *American Political Science Review*, *91*(3), 567–583.

O'Brien, R. M. (1981). Using Rank Category Variables to Represent Continuous Variables: Defects of Common Practice. *Soc. Forces*, *59*, 1149–1162.

O'Keefe, D. J. (2002). *Persuasion: Theory & research* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Palin, S. (2009). Statement on the Current Health Care Debate. Retrieved from `https://www.facebook.com/notes/sarah-palin/statement-on-the-current-health-care-debate/113851103434/`

Pan, Z., & Kosicki, G. M. (1993). Framing Analysis: An Approach to News Discourse. *Political Communication*, *10*(1), 55–75.

Pigott, T. (2001). A Review of Methods for Missing Data. *Research and Evaluation*, *7*(4), 353–383.

Pocock, S. J., & Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, *31*, 103–115.

Price, V., Tewksbury, D., & Powers, E. (1997). Switching Trains of Thought: The Impact of News Frames on Readers' Cognitive Responses. *Communication Research*, *24*(5), 481–506.

Raghunathan, T. (2016). *Missing Data Analysis in Practice*. Boca Raton, FL: CRC Press.

Rees, P. H., & Duke-Williams, O. (1997). Methods for Estimating Missing Data on Migrants in the 1991 British Census. *International Journal of Population Geography*, *3*, 323–68.

Reilly, M. (1993). Data Analysis Using Hot Deck Multiple Imputation. *The Statistician*, *42*, 307–13.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2020). *Package "MASS"*. https://cran.r-project.org/web/packages/MASS/MASS.pdf, R package version 7.3-51.6.

Robins, J. M., Rotznitzky, A., & Scharfstein, D. O. (1998). Semiparametric Regression for Repeated Outcomes with Non-Ignorable Non-Response. *Journal of the American Statistical Association*, *93*, 1321–1339.

Rockwell, R. C. (1975). An Investigation of Imputation and Differential Quality of Data in the 1970 Census. *Journal of the American Statistical Association*, *70*, 39–42.

Rosenberger, W. F., & Lachin, J. (2002). *Randomization in clinical trials*. New York, NY: John Wiley & Sons.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701.

Rubin, D. (1976). Inference and Missing Data. *Biometrika*, *63*(3), 581–592.

Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York, NY: Wiley-Blackwell.

Rubin, D. (1994). Missing Data, Imputation, and the Bootstrap: Comment. *Journal of the American Statistical Association*, *89*(426), 475–478.

Rubin, D. (1996). Multiple Imputation after 18+ Years. *Journal of the American Statistical Association*, *91*, 473–489.

Rubin, D., & Schenker, N. (1986). Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Non-Response. *Journal of the American Statistical Association*, *81*(394), 366–374.

Ryan, T. J. (2014a). *No Compromise: The Politics of Moral Conviction* (Ph.D. Thesis). University of Michigan.

Ryan, T. J. (2014b). Reconsidering Moral Issues in Politics. *Journal of Politics*, *76*(2), 380–397.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London, UK: Chapman; Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods*, *7*(2), 147–177.

Schouten, R. M., Lugtig, P., & Vink, G. (2018). Generating Missing Values For Simulation Purposes: A Multivariate Amputation Procedure. *Journal of Statistical Computation and Simulation*, *88*(15), 2909–2930.

Schunk, D. (2008). A Markov Chain Monte Carlo Algorithm for Multiple Imputation in Large Surveys. *Advances in Statistical Analysis*, *92*(1), 101–114.

Shao, J., & Sitter, R. R. (1996). Bootstrap for Imputed Survey Data. *Journal of the American Statistical Association*, *91*(435), 1278–1288.

Skitka, L. J. (2010). The Psychology of Moral Conviction. *Social and Personality Psychology Compass*, *4*, 267–281.

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral Conviction: Another Contributor to Attitude Strength or Something More? *Journal of Personality and Social Psychology*, *88*(6), 895–917.

Skitka, L. J., & Wisneski, D. C. (2011). Moral Conviction and Emotion. *Emotion Review*, *3*(3), 328–330.

Slothuus, R. (2008). More Than Weighting Cognitive Importance: A Dual-Process Model of Issue Framing Effects. *Political Psychology*, *29*(1), 1–28.

Slothuus, R., & Vreese, C. de. (2010). Political Parties, Motivated Reasoning, and Issue Framing Effects. *Journal of Politics*, *72*(3), 630–645.

Smith, K. B. (2002). Typologies, Taxonomies, and the Benefits of Policy Classification. *Policy Studies Journal*, *30*(3), 379–395.

Sniderman, P. M., & Theriault, S. M. (2004). The Structure of Political Argument and the Logic of Issue Framing. In W. E. Saris & P. M. Sniderman (Eds.), *Studies in Public Opinion: Attitudes, Non-Attitudes, Measurement Error, and Change* (pp. 133–165). Princeton, NJ: Princeton University Press.

Stanley, L. (2016). Using focus groups in political science and international relations. *Politics*, *36*(3), 236–249.

Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions* (3rd ed.). New York, NY: Springer.

Tatalovich, R., & Daynes, B. W. (2011). *Moral Controversies in American Politics* (4th ed.). Armonk, NY: M.E. Sharpe.

Tatalovich, R., Smith, A., & Bobic, M. P. (1994). Moral Conflict and the Policy Process. *Policy Currents*, *4*, 1–7.

Testa, C. (2017). *QualtricsTools*. https://github.com/emmamorgan-tufts/QualtricsTools/, package version 0.1.0.

Tetlock, P. E. (2003). Correspondence and Coherence Indicators of Good Judgment in World Politics. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision-Making* (pp. 233–250). New York, NY: Cambridge University Press.

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, *211*, 453–458.

Urdan, T. C. (2010). *Statistics in Plain English*. New York, NY: Routledge.

USBC. (2002). *Current Population Survey: Design and Methodology* (Vol. 63). U.S. Bureau of the Census.

Vreese, C. de. (2004). The Effects of Frames in Political Television News on Issue Interpretation and Frame Salience. *Journalism and Mass Communication Quarterly*, *81*(1), 36–52.

Wei, G. C., & Tanner, M. A. (1990). A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms. *Journal of the American Statistical Association*, *85*, 699–704.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple Imputation Using Chained Equations: Issues and Guidance for Practice. *Statistics in Medicine*, *30*, 377–399.

Winship, C., & Mare, R. D. (1984). Regression models with ordinal variables. *American Sociological Review*, *49*(4), 512–525.

Wothke, W. (2000). Longitudinal and Multi-Group Modeling with Missing Data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling Longitudinal and Multiple-Group Data: Practical Issues, Applied Approaches, and Specific Examples* (pp. 219–240). Hillsdale, NJ: Erlbaum.

Young, W., Weckman, G., & Holland, W. (2011). A Survey of Methodologies for the Treatment of Missing Values within Datasets: Limitations and Benefits. *Theoretical Issues in Ergonomics Science*, *12*(1), 15–43.

Yu, L. M., Burton, A., & Rivero-Arias, O. (2007). Evaluation of Software for Multiple Imputation of Semi-Continuous Data. *Statistical Methods in Medical Research*, *16*, 243–258.

Zaller, J. (1992). *The Nature and Origins of Mass Opinion*. New York, NY: Cambridge University Press.