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

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 as the distance between each of these categories is identical between any adjacent pair and thus translates perfectly into the numeric values with 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 $\boldsymbol{Y}_i \in [1, \dots, k]$, for $i = 1, \dots n$, and let $\boldsymbol{Y}$ be assumed to be produced by the unobserved latent continuous variable $\boldsymbol{Y^{cont}}$. $\boldsymbol{Y^{cont}}$ is continuous on R from $-\infty$ to ∞. The 'response mechanism' for the r^{th} category is $Y = r \iff \xi_{r-1} < Y^{cont} < \xi_r$. This requires there to be thresholds on R: Y_i^{cont} : $\xi_0 \underset{a=1}{\longleftrightarrow} \xi_1 \underset{a=2}{\longleftrightarrow} \xi_2 \underset{a=3}{\longleftrightarrow} \xi_3 \dots \xi_{A-1} \underset{a=A}{\longleftrightarrow} \xi_A$. The vector of (unseen) utilities across individuals in the sample, Y^{cont}, is determined by a linear model of explanatory variables: $\boldsymbol{Y^{cont}} = \boldsymbol{X\beta} + \mu$, where $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_p]$ does not depend on the ξ_j and $\mu \sim F_\mu$. For the observed vector $\boldsymbol{Y}$,

$$\begin{aligned} p(\boldsymbol{Y} \leq r|\boldsymbol{X}) = p(\boldsymbol{Y^{cont}} \leq \xi_r) &= p(\boldsymbol{X\beta} + \mu \leq \xi_r) \\ &= p(\mu \leq \xi_r + \boldsymbol{X\beta}) = F_\mu(\xi_r + \boldsymbol{X\beta}) \end{aligned} \tag{1.1}$$

is called the cumulative model because $p(\boldsymbol{Y} \leq \xi_r|\boldsymbol{X}) = p(\boldsymbol{Y} = 1|\boldsymbol{X}) + p(\boldsymbol{Y} = 2|\boldsymbol{X}) + \ldots + p(\boldsymbol{Y} = r|\boldsymbol{X})$. A logistic distributional assumption on the errors produces the ordered logit specification

$$F_\mu(\xi_r - \boldsymbol{X'\beta}) = P(\boldsymbol{Y} \leq r|\boldsymbol{X}) = [1 + \exp(-\xi_r - \boldsymbol{X'\beta})]^{-1}. \tag{1.2}$$

The likelihood function is

$$L(\boldsymbol{\beta}, \boldsymbol{\xi}|\boldsymbol{X}, \boldsymbol{Y}) = \prod_{i=1}^{n} \prod_{j=1}^{A-1} [\Lambda(\xi_j + \boldsymbol{X}_i'\boldsymbol{\beta}) - \Lambda(\xi_{j-1} + \boldsymbol{X}_i'\boldsymbol{\beta})]^{z_{ij}} \tag{1.3}$$

where $z_{ij} = 1$ if the i^{th} case is in the j^{th} category, and $z_{ij} = 0$ otherwise. The thresholds on R partition the variable into regions corresponding to the ordinal categories. The linear model, Y^{cont}, 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 model 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 political 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 simulations. Since there currently is no available tool to block online survey experiments, I create my own survey environment in `R shiny`, which is outlined in appendix A.

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

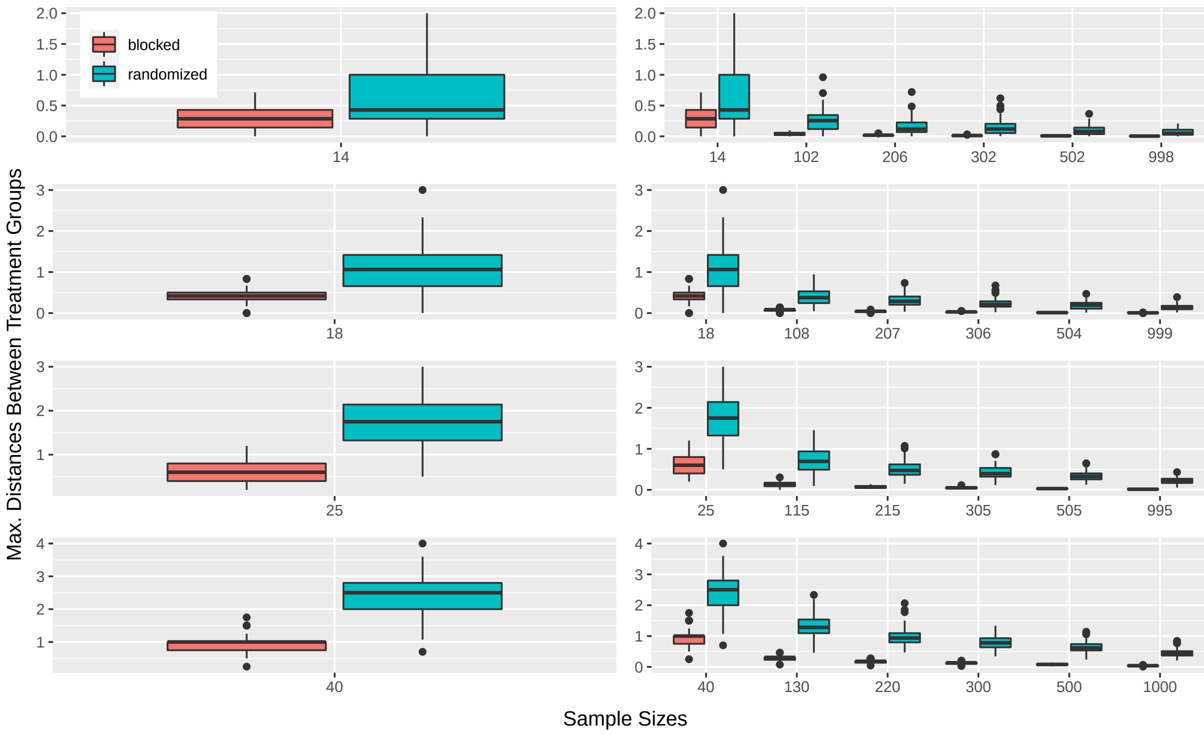

Figure 2.1. Distances between treatment group means in randomized and blocked data. Increasing sample size for 2 (top row), 3 (second row), 5 (third row), and 10 treatment groups (bottom row). Leftmost pair on right panel is exactly the pair on the left panel

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.

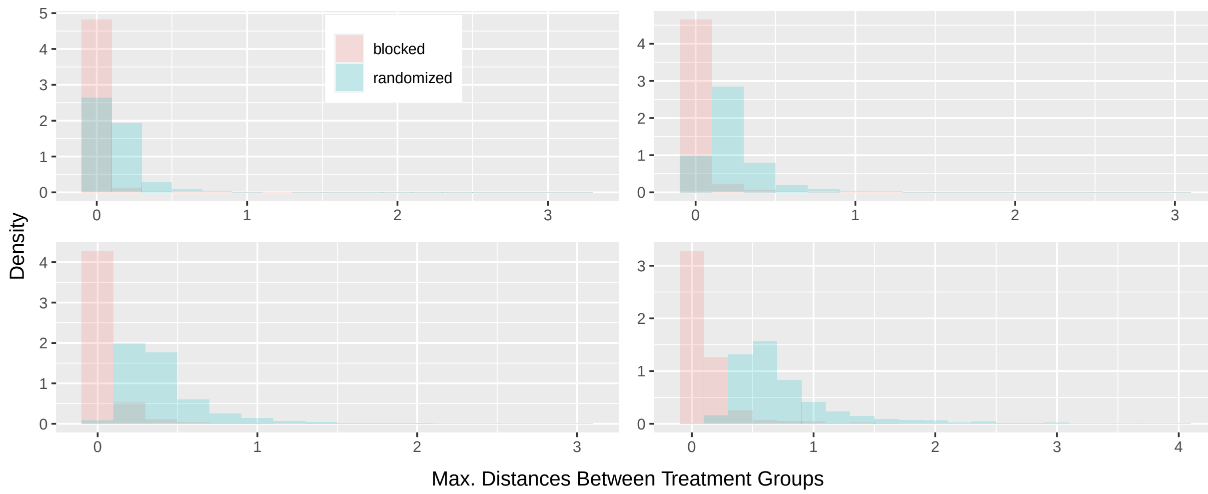

Figure 2.2. Distribution of treatment group differences in randomized and blocked data for 2 (top left), 3 (top right), 5 (bottom left), and 10 (bottom right) treatment groups

Blocking On The Go

In political science, researchers often have an already-collected data set in front of them. One example are 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 sections will apply this solution to external survey data in simulations.

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

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	−0.711	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

values with the Mahalanobis distance, which would not theoretically permitted without empirical justification. The following sections show that the new estimated categories affect analyses and results.

2.3.1 Simulations

I conduct various simulations to compare the Ordered Probit Model and its resulting re-estimated 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:

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

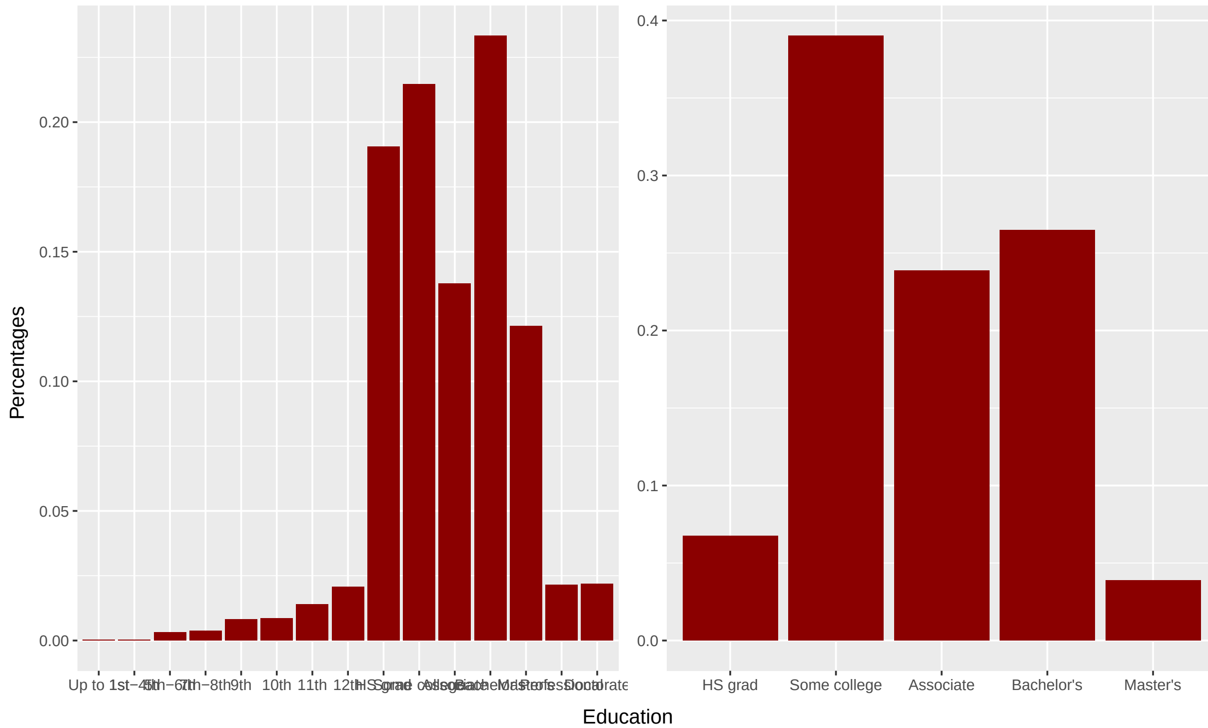

Figure 2.3. Distribution of Education Categories. Original 2016 ANES categories on the left, ordered probit estimated categories on the right

`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 be zero.

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.

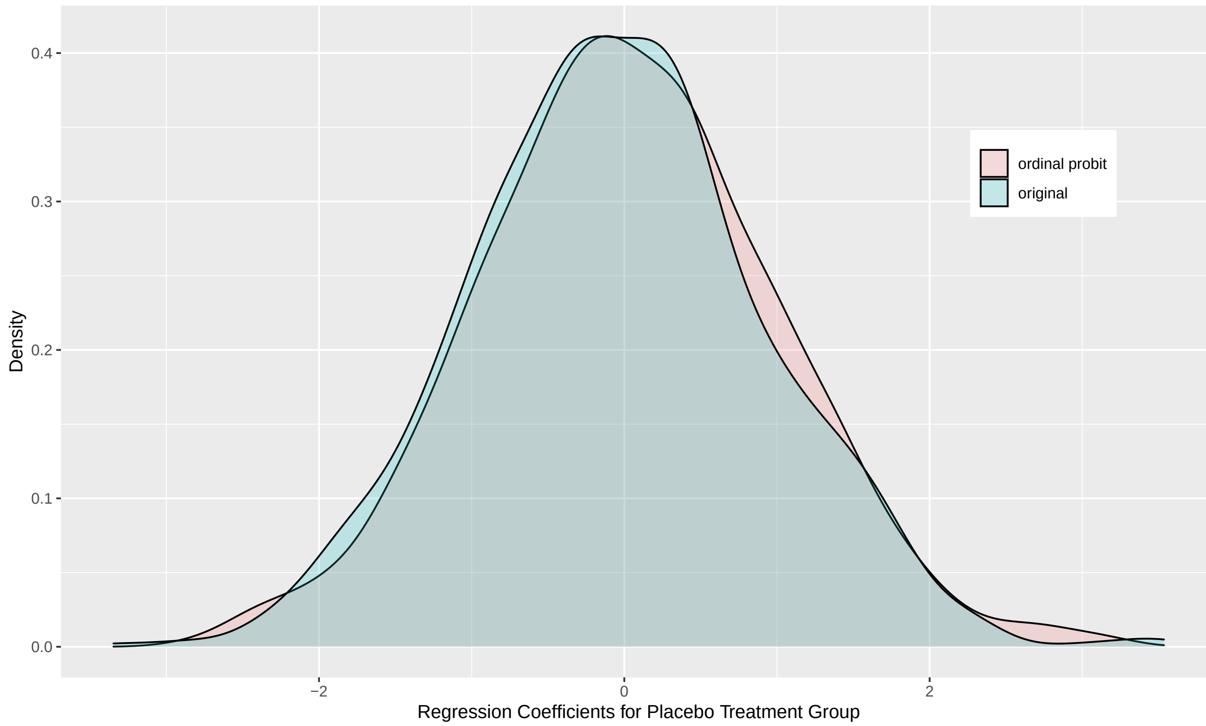

Figure 2.4. Distribution of placebo treatment coefficients by education model

Upon closer inspection, the ordered probit categories are closer to the true values than the original categories on both mean (-0.05 v. 0.019) and median (-0.07 v. -0.01). This indicates slightly superior performance by the ordered probit categories.

Table 2.2. Variable Proportions/Means After Blocking ANES Data on ANES and OP Education Categories, Differentiated by Treatment Group

	Group 1		Group 2		Group 3		Group 4		Group 5	
	AN	OP	AN	OP	AN	OP	AN	OP	AN	OP
Gender										
Male	.467	.468	.473	.481	.475	.463	.463	.468	.476	.473
Female	.533	.529	.524	.514	.522	.535	.533	.532	.521	.524
Other	.000	.003	.003	.005	.003	.002	.003	.000	.003	.003
Race										
White	.763	.751	.770	.778	.752	.743	.749	.746	.752	.770
African American	.095	.089	.090	.089	.106	.100	.094	.117	.110	.100
Asian	.032	.035	.041	.027	.027	.035	.027	.032	.030	.029
Native American	.008	.000	.003	.005	.003	.006	.006	.008	.006	.008
Hispanic	.102	.125	.095	.102	.111	.116	.124	.097	.102	.094
Income										
Under 25,000	.206	.203	.198	.219	.224	.206	.216	.211	.211	.216
25,000-49,999	.246	.237	.246	.205	.190	.225	.211	.206	.217	.238
50,000-749,999	.167	.175	.187	.178	.210	.173	.189	.208	.162	.181
75,000-99,999	.119	.135	.148	.133	.130	.151	.133	.127	.135	.119
100,000-124,999	.102	.090	.094	.095	.094	.092	.079	.087	.089	.092
125,000-149,999	.038	.040	.033	.052	.040	.035	.049	.041	.052	.044
150,000-174,999	.046	.038	.030	.035	.041	.038	.043	.040	.044	.054
175,000 or more	.076	.083	.063	.083	.071	.079	.079	.079	.089	.056
Employment										
Working	.646	.646	.649	.608	.600	.652	.624	.613	.622	.622
Unemployed	.056	.065	.057	.054	.063	.051	.051	.059	.075	.073
Retired	.183	.181	.184	.213	.203	.179	.214	.210	.192	.194
Disabled	.033	.035	.043	.046	.051	.037	.027	.033	.038	.041
Homemaker	.049	.049	.052	.054	.052	.052	.060	.063	.054	.049
Student	.033	.024	.014	.025	.030	.029	.024	.022	.019	.021
Party ID										
Democrat	.314	.378	.367	.348	.338	.341	.378	.344	.378	.363
Republican	.279	.292	.279	.294	.317	.281	.308	.321	.295	.292
Independent	.362	.300	.330	.324	.311	.333	.284	.302	.292	.321
Something else	.044	.030	.024	.035	.033	.044	.030	.033	.035	.024
President										
Approve	.551	.533	.514	.524	.502	.535	.522	.522	.544	.519
Disapprove	.449	.467	.486	.476	.498	.465	.478	.478	.456	.481
Min. Wage										
Raised	.648	.635	.621	.630	.644	.662	.619	.637	.681	.649
Kept the same	.302	.303	.302	.289	.308	.273	.310	.308	.251	.298
Lowered	.019	.021	.027	.029	.013	.027	.025	.021	.025	.013
Eliminated	.032	.041	.051	.052	.035	.038	.046	.035	.043	.040
Country										
Right direction	.290	.292	.252	.251	.249	.252	.267	.265	.268	.267
Wrong track	.710	.708	.748	.749	.751	.748	.733	.735	.732	.733
Age	48.548	48.957	48.949	49.54	48.811	47.157	49.702	49.773	49.1	49.683
Feel Trump	34.695	34.378	37.278	36.394	40.579	36.241	35.767	38.584	34.59	37.313
Education										
Up to 1st	.000		.000		.000		.000		.002	
1st-4th	.002		.000		.000		.000		.000	
5th-6th	.003		.003		.003		.003		.003	
7th-8th	.003		.005		.005		.003		.003	
9th	.008		.008		.008		.010		.008	
10th	.010		.010		.008		.008		.008	
11th	.013		.014		.014		.014		.014	
12th	.021		.021		.021		.021		.021	
HS grad	.190	.068	.190	.067	.190	.068	.190	.068	.190	.067
Some college	.214	.389	.214	.390	.216	.389	.214	.390	.214	.392
Associate	.138	.240	.138	.238	.137	.240	.138	.238	.138	.238
Bachelor's	.233	.265	.233	.265	.233	.263	.233	.265	.233	.265
Master's	.121	.038	.121	.040	.122	.040	.121	.038	.122	.038
Professional	.022		.021		.021		.024		.021	
Doctorate	.022		.022		.022		.021		.022	

Table 2.3. Summary of ANOVA Regression of Variable on ANES/OP Indicator, Differentiated by Treatment Group

	Df	Sum.Sq	Mean.Sq	F-value	p-value
Age					
T1	1	52.830	52.830	.170	.680
T1 Residuals	1,258	384,359.900	305.530		
T2	1	109.830	109.830	.360	.550
T2 Residuals	1,258	388,724.900	309		
T3	1	861.720	861.720	2.810	.090
T3 Residuals	1,258	386,400.000	307.150		
T4	1	1.610	1.610	.010	.940
T4 Residuals	1,258	387,498.400	308.030		
T5	1	106.900	106.900	.350	.550
T5 Residuals	1,258	382,695.200	304.210		
Feel Trump					
T1	1	31.750	31.750	.030	.870
T1 Residuals	1,258	1,465,958.000	1,165.310		
T2	1	246.230	246.230	.200	.660
T2 Residuals	1,258	1,576,959.000	1,253.540		
T3	1	5,928.010	5,928.010	4.870	.030
T3 Residuals	1,258	1,530,847.000	1,216.890		
T4	1	2,500.500	2,500.500	2.040	.150
T4 Residuals	1,258	1,539,114.000	1,223.460		
T5	1	2,334.310	2,334.310	1.890	.170
T5 Residuals	1,258	1,550,046.000	1,232.150		

Table 2.4. Tukey's 'Honest Significant Difference' Test of ANOVA Regression of Variable on ANES/OP Indicator, Differentiated by Treatment Group

	Diff.Means	CI Lower	CI Upper	Adj. p-value
Age				
T1 OP1-AN1	.410	-1.523	2.342	.678
T2 OP2-AN2	.590	-1.353	2.534	.551
T3 OP3-AN3	-1.654	-3.591	.283	.094
T4 OP4-AN4	.071	-1.869	2.011	.942
T5 OP5-AN5	.583	-1.345	2.510	.553
Feel Trump				
T1 OP1-AN1	-.317	-4.091	3.456	.869
T2 OP2-AN2	-.884	-4.798	3.030	.658
T3 OP3-AN3	-4.338	-8.194	-.482	.027
T4 OP4-AN4	2.817	-1.049	6.684	.153
T5 OP5-AN5	2.722	-1.158	6.602	.169

Table 2.5: Summary of GLM Regression of Variable on ANES/OP Indicator, Differentiated by Treatment Group

	Estimate	Std.Error	z-value	p-value
Gender				
T1 Intercept	.134	.080	1.672	.095
T1	-.006	.113	-.056	.955
T2 Intercept	.108	.080	1.354	.176
T2	-.032	.113	-.282	.778
T3 Intercept	.102	.080	1.274	.203
T3	.045	.113	.395	.693
T4 Intercept	.146	.080	1.831	.067
T4	-.019	.113	-.169	.865
T5 Intercept	.095	.080	1.195	.232
T5	.013	.113	.113	.910
Race				
T1 Intercept	-1.172	.094	-12.500	.000
T1	.069	.131	.526	.599
T2 Intercept	-1.207	.095	-12.757	.000
T2	-.045	.135	-.337	.736
T3 Intercept	-1.111	.092	-12.040	.000
T3	.050	.130	.389	.697
T4 Intercept	-1.094	.092	-11.907	.000
T4	.017	.130	.130	.897
T5 Intercept	-1.111	.092	-12.040	.000
T5	-.096	.132	-.727	.467
Income				
T1 Intercept	1.347	.098	13.683	.000
T1	.019	.140	.140	.889
T2 Intercept	1.396	.100	13.976	.000
T2	-.125	.139	-.901	.368
T3 Intercept	1.244	.096	13.010	.000
T3	.103	.137	.754	.451
T4 Intercept	1.290	.097	13.320	.000
T4	.028	.138	.206	.837
T5 Intercept	1.318	.098	13.503	.000
T5	-.028	.138	-.206	.837
Occupation				
T1 Intercept	-.602	.083	-7.221	.000
T1	-.000	.118	-.000	1.000
T2 Intercept	-.616	.083	-7.373	.000
T2	.177	.117	1.515	.130
T3 Intercept	-.405	.081	-4.986	.000
T3	-.224	.117	-1.920	.055
T4 Intercept	-.506	.082	-6.149	.000
T4	.047	.116	.406	.685
T5 Intercept	-.499	.082	-6.072	.000
T5	.000	.116	.000	1.000

Party ID				
T1 Intercept	.780	.086	9.090	.000
T1	-.281	.119	-2.366	.018
T2 Intercept	.547	.083	6.611	.000
T2	.083	.118	.705	.481
T3 Intercept	.672	.084	7.977	.000
T3	-.014	.119	-.119	.905
T4 Intercept	.499	.082	6.072	.000
T4	.145	.117	1.231	.218
T5 Intercept	.499	.082	6.072	.000
T5	.061	.117	.525	.600
Pres. Approval				
T1 Intercept	-.204	.080	-2.545	.011
T1	.070	.113	.622	.534
T2 Intercept	-.057	.080	-.717	.473
T2	-.038	.113	-.338	.735
T3 Intercept	-.006	.080	-.080	.936
T3	-.134	.113	-1.184	.236
T4 Intercept	-.089	.080	-1.115	.265
T4	.000	.113	.000	1.000
T5 Intercept	-.178	.080	-2.228	.026
T5	.102	.113	.903	.366
Min. Wage				
T1 Intercept	-.609	.083	-7.297	.000
T1	.055	.117	.470	.638
T2 Intercept	-.492	.082	-5.995	.000
T2	-.041	.116	-.349	.727
T3 Intercept	-.595	.083	-7.145	.000
T3	-.077	.118	-.651	.515
T4 Intercept	-.486	.082	-5.918	.000
T4	-.075	.117	-.641	.522
T5 Intercept	-.758	.085	-8.870	.000
T5	.143	.119	1.193	.233
Country Track				
T1 Intercept	.893	.088	10.176	.000
T1	-.008	.124	-.062	.951
T2 Intercept	1.086	.092	11.840	.000
T2	.008	.130	.065	.948
T3 Intercept	1.103	.092	11.974	.000
T3	-.017	.130	-.130	.897
T4 Intercept	1.012	.090	11.228	.000
T4	.008	.128	.064	.949
T5 Intercept	1.003	.090	11.159	.000
T5	.008	.127	.064	.949

Table 2.6. ANOVA Chisq Test of GLM Regression of Variable on ANES/OP Indicator, Differentiated by Treatment Group

	Df	Deviance	Resid.Df	Residual Deviance	Pr.Chi
Gender					
T1	1	.003	1,258	1,741.387	.955
T2	1	.080	1,258	1,743.981	.778
T3	1	.156	1,258	1,741.743	.693
T4	1	.029	1,258	1,740.828	.865
T5	1	.013	1,258	1,743.466	.910
Race					
T1	1	.276	1,258	1,396.688	.599
T2	1	.113	1,258	1,347.160	.736
T3	1	.151	1,258	1,423.487	.697
T4	1	.017	1,258	1,423.621	.897
T5	1	.528	1,258	1,384.957	.467
Income					
T1	1	.019	1,258	1,277.453	.889
T2	1	.812	1,258	1,290.106	.367
T3	1	.569	1,258	1,311.366	.451
T4	1	.043	1,258	1,306.695	.837
T5	1	.043	1,258	1,306.695	.837
Occupation					
T1	1	.000	1,258	1,637.669	1.000
T2	1	2.299	1,258	1,660.174	.129
T3	1	3.695	1,258	1,661.905	.055
T4	1	.165	1,258	1,675.415	.685
T5	1	.000	1,258	1,670.674	1.000
Party ID					
T1	1	5.617	1,258	1,619.670	.018
T2	1	.498	1,258	1,641.929	.480
T3	1	.014	1,258	1,614.864	.905
T4	1	1.517	1,258	1,646.701	.218
T5	1	.276	1,258	1,661.142	.600
Pres. Approval					
T1	1	.387	1,258	1,737.416	.534
T2	1	.114	1,258	1,744.787	.735
T3	1	1.402	1,258	1,743.649	.236
T4	1	.000	1,258	1,744.241	1.000
T5	1	.816	1,258	1,740.832	.366
Min. Wage					
T1	1	.221	1,258	1,644.543	.638
T2	1	.122	1,258	1,666.506	.727
T3	1	.424	1,258	1,626.133	.515
T4	1	.411	1,258	1,663.111	.521
T5	1	1.426	1,258	1,605.348	.232
Country Track					
T1	1	.004	1,258	1,520.274	.951
T2	1	.004	1,258	1,421.458	.948
T3	1	.017	1,258	1,419.261	.897
T4	1	.004	1,258	1,459.355	.949
T5	1	.004	1,258	1,463.401	.949

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 since it accounts for and incorporates uncertainty around the estimated imputations through repeated draws (Andridge & Little, 2010; Graham & Schafer, 1999; Schafer & Graham, 2002; White, Royston, & Wood, 2011). 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. Uncertainty is not reflected in the imputed values, which leads to biased standard errors and confidence intervals (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 as well as the most common single imputation methods. Section 3.2.4 explains the basics of multiple imputation and outlines major `R` packages that implement it. These include `Amelia`, `mice`, `hot.deck` (applying multiple hot deck imputation), and my self-penned multiple hot deck imputation function for ordinal variables, `hd.ord`. Since single imputation is widely condemned as a general imputation method, my focus lies on multiple imputation. Section 3.3 describes the survey data these packages/functions are tested on: A selection of the 2016 ANES and a subset of the 2016 CCES. 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 (nominal, ordinal, interval). Section 3.4 shows the results and assesses the usefulness of my proposed new method. Finally, section 3.6 will feature 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$(R = 0|Y^{obs}, Y^{miss}, \beta)$, i.e. the probability of missing data depends on the observed data, the missing data, and a vector of unknown parameters.[1] 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), or missing not at random (MNAR). It is not possible to test respective data on the mechanism underlying their 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, 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 to survey data: Survey respondents are known to withhold

[1]The literature often uses θ for missing data notation, but since my concern lies primarily with regression style models, I will be using β instead.

sensitive data (Groves et al., 2009). 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, e.g. 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$(R = 0|\beta)$, i.e. the generic probability of missing data, independent of the data themselves and only dependent on β.

As a result, a MCAR assumption in political science surveys is rare and requires tremendous justification. 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$(R = 0|Y^{obs}, \beta)$, i.e. the chances of missing data depend on the observed data and β.

Finally, missing data can be missing not at random (MNAR).[2] 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$(R = 0|Y^{obs}, Y^{miss}, \beta)$,

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

i.e. the 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 always biased, since the observed data is tilted towards respondents with characteristics that make them more likely to respond. Whether this bias is trivial or substantial depends on the research in question (Collins et al., 2001; Graham, Hofer, Donaldson, MacKinnon, & Schafer, 1997). The potential for severe 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 can be implemented 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

Mean imputation, sometimes also called unconditional mean imputation, means replacing missing values within cells 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 non-sensical 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 the `mice` package. It is also very simple to execute 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. Let there be an independent variable x in a multiple regression model. Assume 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}, thus 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 the `mice` function to conduct linear regression. The `predict` function in base `R` also applies linear regression imputation.

Hot Decking

Hot deck 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 deck 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. Contrary to regression imputation, hot decking also 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 single hot deck imputation. Nonetheless, variations 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

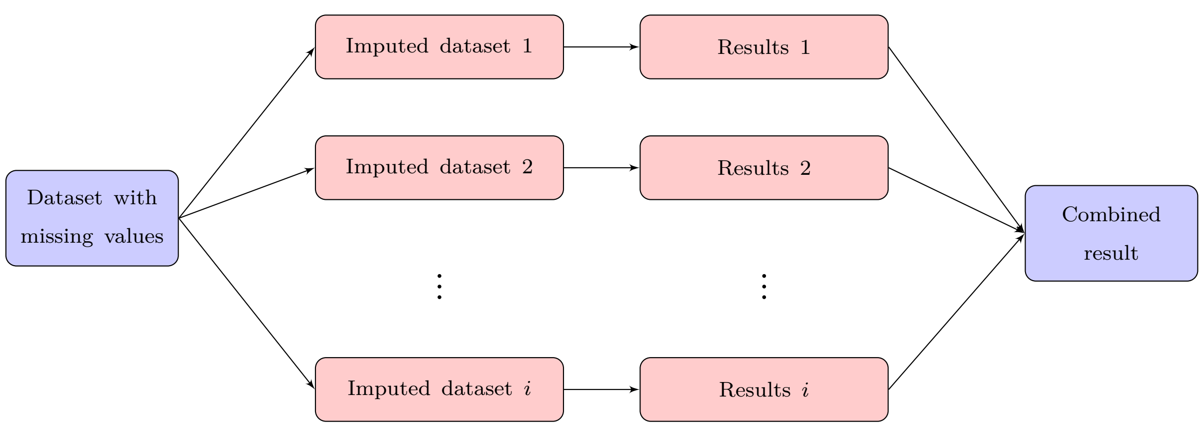

Figure 3.1. Multiple Imputation Workflow

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 $i = 1, ..., m$. Averaging τ_i then gives us the single estimate, $\overline{\tau}$. Together, this is expressed as:

$$\overline{\tau} = \frac{1}{m}\sum_{i=1}^{m}\tau_i. \tag{3.1}$$

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 both datasets, Var_A:

$$\overline{Var_W} = \frac{1}{m}\sum_{i=1}^{m} SE(\tau_i)^2 \tag{3.2}$$

$$Var_A = \sum_{i=1}^{m}\frac{(\tau_i - \overline{\tau})^2}{m-1} \tag{3.3}$$

$$Var_T = \overline{Var_W} + Var_A \tag{3.4}$$

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

$$Var_T = \overline{Var_W} + Var_A(1 + \frac{1}{m}) \tag{3.5}$$

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, $m = 20$ (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 1998 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`. My `hd.ord` analysis 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`.[3] 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 of users 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 five, and the console printing options are turned off.

`mice`: Multivariate Imputation by Chained Equations

The `R` package `mice` was 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 (Buuren, 2007; Buuren & Groothuis-Oudshoorn, 2011), a version of the imputation-posterior (IP) (King et al., 2001). Full

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

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). The initial release of `mice` 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).

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 proven 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 β. Now let Y further be a random sample from the z-variate multivariate distribution, $Z(Y|\beta)$, with β accounting for the multivariate distribution of Y. The proverbial pot of gold here is how to estimate the multivariate distribution of β. Under the chained equations model, we estimate a posterior distribution of β by sampling repeatedly from conditional distributions, i.e.:

$$\begin{aligned} &Z(Y_1|Y_{-1}, \beta_1) \\ &\quad\vdots \\ &Z(Y_z|Y_{-z}, \beta_z) \end{aligned} \tag{3.6}$$

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

$$\begin{aligned} \beta_1^{\sim(n)} &\sim Z(\beta_1|Y_1^{obs}, Y_2^{(n-1)}, ..., Y_z^{(n-1)}) \\ Y_1^{\sim(n)} &\sim Z(Y_1|Y_1^{obs}, Y_2^{(n-1)}, ..., Y_z^{(n-1)}, \beta_1^{\sim(n)}) \\ &\vdots \\ \beta_z^{\sim(n)} &\sim Z(\beta_z|Y_z^{obs}, Y_1^{(n)}, ..., Y_{z-1}^{(n)}) \\ Y_z^{\sim(n)} &\sim Z(Y_z|Y_z^{obs}, Y_1^{(n)}, ..., Y_z^{(n)}, \beta_z^{\sim(n)}) \end{aligned} \tag{3.7}$$

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

```r
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

user-friendliness. The majority of arguments are not of importance to general users. Arguments 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 `data` requires user input. As mentioned above, `m` should also be set to the percentage of missing data. The argument `defaultMethod` does not require user input but is crucial for insights 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. The argument `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 two 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

The `R` package `Amelia` was originally released in 1998 (Honaker, Joseph, King, Scheve, & Singh, 1998). 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. The `Amelia` package 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. The `Amelia II` package addresses this by mixing EM with 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 β, with $\beta = (\mu, \sum)$. Let there further be our missingness matrix$R$ and the likelihood of Y^{obs}, prob$(Y^{obs}, R|\beta)$. `Amelia II` is explicitly set up for the MAR assumption of missing data, prob$(R = 0|Y^{obs}, \beta)$. Under this assumption, the likelihood can be transformed as

$$\text{prob}(Y^{obs}, R|\beta) = \text{prob}(R|Y^{obs})\text{prob}(Y^{obs}|\beta). \tag{3.8}$$

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

$$L(\beta|Y^{obs}) \propto \text{prob}(Y^{obs}|\beta) \tag{3.9}$$

which further translates into

$$\text{prob}(Y^{obs}|\beta) = \int \text{prob}(Y|\beta)yY^{miss} \tag{3.10}$$

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

$$\text{prob}(\beta|Y^{obs}) \propto \text{prob}(Y^{obs}|\beta) = \int \text{prob}(Y|\beta)yY^{miss}. \tag{3.11}$$

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 β. The latter is a linear regression with parameters that can be estimated from β. This bootstrapped EM approach is 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 argument that requires user input is `x`, which needs to be data with missing values that can be in a variety of formats. `m`, identical 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.

General multiple imputation 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. While `amelia` thus does incorporate ordinal variables to some extent, the rounding process changes the nature of ordinal variables to continuous variables. In addition, none of its 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 section 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 non-sensical 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. Rounding thus by definition introduces at least some level of bias. The problem is exacerbated for ordinal variables, where the spacing between the discrete variable categories is unknown, since it arbitrarily reduces or lengthens distances between the categories. 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 are 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

$$\alpha_{co} = \frac{k - q_c - w_{co}}{k - q_c} \tag{3.12}$$

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, following the logic displayed in Figure 3.1.

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 requires user input for `data`. `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 for instance 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 is 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 demonstrated with a simplified example shown in Table 3.1. Respondent B

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

.

shows 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 makes sense 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 following the convention of 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: $.032 - (1.063 - 0.032) / 2 =$ -0.483. The same process is applied to estimate 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. Figure 3.5 shows my self-penned imputation function, `hd.ord`, with all

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

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 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 two different sets of survey data: Data from the 2016 ANES and the 2016 CCES. Data for all selected variables in both 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 is a difficult feat. Data that are artificially created to be 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 four different functions: `amelia`, `mice`, `hot.deck`, and `hd.ord`. 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 five, and the console printing options are turned off.

I test each function for accuracy and speed for binary, ordinal, and interval variables in both 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 explicitly 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 and survey experiment data, and the `polr` analysis from section 2.3 estimates five or six levels to best represent `Education` in a US context. Each dataset was imputed 1,000 times with each of the four imputation methods. With the exception of Table 3.14, 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`. Highly collinear variables were excluded with a cutoff of .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 both 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 both 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` in `hd.ord` by including `Interest`. Imputations are again 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 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 both 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 both datasets, while `mice` and `amelia` perform best. `hd.ord` is relatively close for CCES `Dem` (+.0001 `amelia` vs. +.0000 `hd.ord`) but further away for ANES (+.0003 `mice` vs. –.0010 `hd.ord` `Dem`; +.0001 `mice` vs. –.0013 `hd.ord` `Male`) and CCES `Male` (–.0001 `amelia` vs. –.0011 `hd.ord`). For the ordinal variable, `Interest`, `hd.ord` performs worst for both datasets, with considerable distance to `hot.deck` (–.0130 vs. –.0191 ANES; –.0125 vs. –.0196 CCES). `mice` and `amelia` perform by far best across both 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 both datasets, while the maximum difference for `hd.ord` amounts to –.0130 (ANES).

For the interval variables, `Inc` and `Age`, `hd.ord` also performs worst for both 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 –.0002 (CCES) away from the true value across both datasets, but the maximum difference for `hd.ord` is –.1068 (ANES). Similarly, `amelia`'s largest deviation from the true value for `Age` is –.0033

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

Method	Variable	ANES	CCES
true	Dem	.3420	.3770
hd.ord	Dem	−.0010	+.0000
hot.deck	Dem	−.0011	−.0004
amelia	Dem	+.0004	+.0001
mice	Dem	+.0003	+.0002
na.omit	Dem	−.0290	−.0229
true	Male	.4890	.4830
hd.ord	Male	−.0013	−.0011
hot.deck	Male	−.0013	−.0014
amelia	Male	+.0002	−.0001
mice	Male	+.0001	−.0001
na.omit	Male	−.0392	−.0414
true	Interest	2.9340	3.3290
hd.ord	Interest	−.0130	−.0125
hot.deck	Interest	−.0191	−.0196
amelia	Interest	+.0003	+.0003
mice	Interest	+.0003	+.0000
na.omit	Interest	−.0705	−.0724
true	Inc	16.6140	6.4810
hd.ord	Inc	−.1068	−.0259
hot.deck	Inc	−.1278	−.0407
amelia	Inc	+.0008	−.0004
mice	Inc	+.0003	−.0002
na.omit	Inc	−.5631	−.2468
true	Age	50.0410	52.8230
hd.ord	Age	−.3888	−.2616
hot.deck	Age	−.4597	−.3895
amelia	Age	+.0007	−.0033
mice	Age	+.0017	−.0073
na.omit	Age	−1.1875	−1.2361

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.

(CCES) as opposed to `hd.ord`'s −.3888 (ANES).

Table 3.5 shows the results of imputing both 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 both 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 an 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
true	Dem	.3420	.3770
hd.ord	Dem	−.0005	−.0003
hot.deck	Dem	−.0005	−.0004
amelia	Dem	+.0000	+.0000
mice	Dem	+.0000	+.0001
na.omit	Dem	−.0191	−.0172
true	Male	.4890	.4830
hd.ord	Male	−.0004	−.0002
hot.deck	Male	−.0001	−.0003
amelia	Male	+.0001	−.0001
mice	Male	+.0000	−.0002
na.omit	Male	−.0256	−.0364
true	Interest	2.9340	3.3290
hd.ord	Interest	−.0053	−.0041
hot.deck	Interest	−.0077	−.0067
amelia	Interest	+.0001	−.0001
mice	Interest	+.0000	−.0001
na.omit	Interest	−.0620	−.0515
true	Inc	16.6140	6.4810
hd.ord	Inc	−.0470	−.0130

hot.deck	Inc	–.0591	–.0212
amelia	Inc	–.0007	–.0005
mice	Inc	–.0013	–.0003
na.omit	Inc	–.6303	–.2860
true	Age	50.0410	52.8230
hd.ord	Age	–.1391	–.0883
hot.deck	Age	–.1835	–.1435
amelia	Age	+.0056	–.0015
mice	Age	+.0048	–.0050
na.omit	Age	–.8638	–.5974
true	Black	.0790	.0950
hd.ord	Black	+.0000	+.0000
hot.deck	Black	+.0000	+.0000
amelia	Black	+.0000	+.0001
mice	Black	+.0000	+.0001
na.omit	Black	–.0092	–.0090
true	Empl	.6610	.4370
hd.ord	Empl	+.0006	+.0000
hot.deck	Empl	+.0006	+.0001
amelia	Empl	+.0000	+.0000
mice	Empl	+.0000	–.0001
na.omit	Empl	–.0087	–.0301
true	Religious	.6460	.6420
hd.ord	Religious	–.0006	–.0003
hot.deck	Religious	–.0005	–.0003
amelia	Religious	–.0001	–.0001
mice	Religious	–.0001	–.0002
na.omit	Religious	–.0166	–.0234
true	Married	.5290	.6310
hd.ord	Married	+.0002	–.0001
hot.deck	Married	+.0002	+.0001
amelia	Married	–.0001	–.0002
mice	Married	–.0001	–.0002
na.omit	Married	–.0384	–.0326
true	OwnHome	.6820	.7010
hd.ord	OwnHome	–.0001	–.0002
hot.deck	OwnHome	+.0000	+.0000
amelia	OwnHome	+.0001	+.0000
mice	OwnHome	–.0001	–.0001
na.omit	OwnHome	–.0334	–.0304
true	Rally	.0830	—
hd.ord	Rally	–.0001	—
hot.deck	Rally	–.0002	—
amelia	Rally	+.0001	—
mice	Rally	+.0001	—
na.omit	Rally	–.0191	—
true	Donate	.1390	—
hd.ord	Donate	–.0002	—
hot.deck	Donate	–.0005	—
amelia	Donate	+.0000	—
mice	Donate	+.0001	—
na.omit	Donate	–.0320	—

true	Gay	—	.0420
hd.ord	Gay	—	+.0001
hot.deck	Gay	—	+.0000
amelia	Gay	—	+.0000
mice	Gay	—	+.0000
na.omit	Gay	—	−.0112
true	StudLoans	—	.1910
hd.ord	StudLoans	—	+.0003
hot.deck	StudLoans	—	+.0002
amelia	StudLoans	—	+.0000
mice	StudLoans	—	−.0001
na.omit	StudLoans	—	−.0117

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 `Empl`). For the ordinal variable, `hd.ord` continues to perform worst across both datasets. `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 −.0053 (ANES) for 12 variables but −.0130 (ANES) 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 both 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).

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 both datasets MNAR for five amputed variables. It is immediately noticeable that the differences between the methods' imputation results and the true values are much higher for all methods for all variables for both datasets. The results for `Dem` and `Male`, for instance, hover around .0100 and .0125 for both datasets. In the corresponding MAR analysis, however, the results for `Dem` and `Male` showed around .0005. For the binary variables, `hd.ord` performs more closely on par with `amelia` and `mice` than in the corresponding MAR analysis above, sometimes more (–.0136 `hd.ord` vs. –.0132 `amelia` ANES `Male`) and sometimes less so (–.0114 `hd.ord` vs. –.0099 `mice` ANES `Dem`).

The results for the ordinal variables confirm those of the MAR analysis: `hd.ord` represents the worst method across both 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.

The results for the interval variables paint the same picture as the ordinal ones. `hd.ord` shows the worst performance. Note that `na.omit` performs equally well as `hd.ord` for ANES `Age`.

Table 3.7 shows the results of imputing both 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 (–.0055 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
true	Dem	.3420	.3770
hd.ord	Dem	–.0049	–.0046
hot.deck	Dem	–.0049	–.0049
amelia	Dem	–.0043	–.0045
mice	Dem	–.0040	–.0044
na.omit	Dem	–.0092	–.0081

true	Male	.4890	.4830
hd.ord	Male	–.0055	–.0049
hot.deck	Male	–.0053	–.0051
amelia	Male	–.0055	–.0050
mice	Male	–.0055	–.0049
na.omit	Male	–.0093	–.0119
true	Interest	2.9340	3.3290
hd.ord	Interest	–.0113	–.0090
hot.deck	Interest	–.0134	–.0113
amelia	Interest	–.0068	–.0061
mice	Interest	–.0068	–.0061
na.omit	Interest	–.0236	–.0161
true	Inc	16.6140	6.4810
hd.ord	Inc	–.0899	–.0350
hot.deck	Inc	–.1046	–.0421
amelia	Inc	–.0495	–.0223
mice	Inc	–.0503	–.0218
na.omit	Inc	–.2088	–.0970
true	Age	50.0410	52.8230
hd.ord	Age	–.2571	–.1732
hot.deck	Age	–.3081	–.2251
amelia	Age	–.1100	–.1014
mice	Age	–.1047	–.0986
na.omit	Age	–.3397	–.1367
true	Black	.0790	.0950
hd.ord	Black	–.0035	–.0038
hot.deck	Black	–.0037	–.0038
amelia	Black	–.0037	–.0040
mice	Black	–.0034	–.0038
na.omit	Black	–.0045	–.0052
true	Empl	.6610	.4370
hd.ord	Empl	–.0034	–.0053
hot.deck	Empl	–.0033	–.0053
amelia	Empl	–.0031	–.0040
mice	Empl	–.0031	–.0040
na.omit	Empl	–.0014	–.0111
true	Religious	.6460	.6420
hd.ord	Religious	–.0045	–.0039
hot.deck	Religious	–.0043	–.0038
amelia	Religious	–.0040	–.0040
mice	Religious	–.0040	–.0040
na.omit	Religious	–.0049	–.0073
true	Married	.5290	.6310
hd.ord	Married	–.0041	–.0038
hot.deck	Married	–.0040	–.0037
amelia	Married	–.0042	–.0037
mice	Married	–.0042	–.0037
na.omit	Married	–.0122	–.0096
true	OwnHome	.6820	.7010
hd.ord	OwnHome	–.0030	–.0030
hot.deck	OwnHome	–.0028	–.0027
amelia	OwnHome	–.0027	–.0030

mice	OwnHome	–.0027	–.0030
na.omit	OwnHome	–.0111	–.0090
true	Rally	.0830	—
hd.ord	Rally	–.0043	—
hot.deck	Rally	–.0042	—
amelia	Rally	–.0040	—
mice	Rally	–.0039	—
na.omit	Rally	–.0077	—
true	Donate	.1390	—
hd.ord	Donate	–.0051	—
hot.deck	Donate	–.0054	—
amelia	Donate	–.0049	—
mice	Donate	–.0048	—
na.omit	Donate	–.0122	—
true	Gay	—	.0420
hd.ord	Gay	—	–.0024
hot.deck	Gay	—	–.0025
amelia	Gay	—	–.0026
mice	Gay	—	–.0024
na.omit	Gay	—	–.0035
true	StudLoans	—	.1910
hd.ord	StudLoans	—	–.0058
hot.deck	StudLoans	—	–.0058
amelia	StudLoans	—	–.0051
mice	StudLoans	—	–.0050
na.omit	StudLoans	—	–.0070

For the ordinal variable, `hd.ord` shows the worst performance across both datasets. `mice` and `amelia` perform far better and display virtually identical results. `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. `na.omit` performs better than `hot.deck` and `hd.ord` for CCES `Age` (–.1367 vs. –.2251 and –.1732).

3.4.3 Increased Number of Ordinal Variables

This section shows the imputation results for an increased number of ordinal variables to be treated by `polr` in `hd.ord`. Specifically, I add `Interest` to the `ord` argument in `hd.ord`. The intuition behind this is a strengthening of the underlying latent continuous variable assumption. The results of the previous analyses 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

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

Method	Variable	ANES	CCES
true	Dem	.3420	.3770
hd.ord	Dem	–.0114	–.0099
hot.deck	Dem	–.0120	–.0105
amelia	Dem	–.0106	–.0102
mice	Dem	–.0099	–.0101
na.omit	Dem	–.0176	–.0140
true	Male	.4890	.4830
hd.ord	Male	–.0136	–.0116
hot.deck	Male	–.0133	–.0124
amelia	Male	–.0132	–.0121
mice	Male	–.0132	–.0120
na.omit	Male	–.0214	–.0219
true	Interest	2.9340	3.3290
hd.ord	Interest	–.0288	–.0246
hot.deck	Interest	–.0335	–.0296
amelia	Interest	–.0167	–.0146
mice	Interest	–.0167	–.0146
na.omit	Interest	–.0379	–.0372
true	Inc	16.6140	6.4810
hd.ord	Inc	–.2299	–.0928
hot.deck	Inc	–.2554	–.1038
amelia	Inc	–.1225	–.0578
mice	Inc	–.1229	–.0566
na.omit	Inc	–.2770	–.1334
true	Age	50.0410	52.8230
hd.ord	Age	–.6319	–.4596
hot.deck	Age	–.7415	–.5929
amelia	Age	–.2450	–.2266
mice	Age	–.2369	–.2160
na.omit	Age	–.6427	–.6392

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.

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` 'moves' 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 both datasets with two `polr`-treated variables MAR for four amputed variables. `hd.ord` displays the worst results for `Dem` across both 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: −.0008, +.0002 vs. −.0010, +.0000 for `Dem` and −.0022, −.0019 vs. −.0013, −.0011 for `Male`. `hd.ord` also performs worst for both datasets across both interval variables. `amelia` and `mice` again perform best. `amelia` and `mice` perform equally well for ANES, while `mice` does better for CCES. 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: −.0830, −.0246 vs. −.1068, −.0259 for `Inc` and −.2889, −.2350 vs. −.3888, −.2616 for `Age`.

Table 3.9 shows the results of imputing both 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`: −.0002, −.0004 vs. −.0005, −.0003 for `Dem` and −.0002, −.0005 vs. −.0004, −.0002 for `Male`.

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

Method	Variable	ANES	CCES
true	Dem	.3420	.3770
hd.ord	Dem	−.0008	+.0002
hot.deck	Dem	−.0018	−.0005
amelia	Dem	+.0002	+.0001
mice	Dem	+.0001	+.0002
na.omit	Dem	−.0333	−.0294
true	Male	.4890	.4830
hd.ord	Male	−.0022	−.0019
hot.deck	Male	−.0015	−.0018
amelia	Male	+.0001	+.0000
mice	Male	+.0000	+.0000
na.omit	Male	−.0396	−.0407
true	Inc	16.6140	6.4810
hd.ord	Inc	−.0830	−.0246
hot.deck	Inc	−.1523	−.0516
amelia	Inc	+.0010	−.0006
mice	Inc	−.0008	+.0002
na.omit	Inc	−.5771	−.2564
true	Age	50.0410	52.8230
hd.ord	Age	−.2889	−.2350
hot.deck	Age	−.5431	−.4664
amelia	Age	+.0018	+.0085
mice	Age	+.0024	−.0002
na.omit	Age	−1.1521	−1.1228

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.

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

Method	Variable	ANES	CCES
true	Dem	.3420	.3770
hd.ord	Dem	−.0002	−.0004
hot.deck	Dem	−.0005	−.0005
amelia	Dem	+.0000	−.0001
mice	Dem	+.0000	+.0000
na.omit	Dem	−.0200	−.0213

true	Male	.4890	.4830
hd.ord	Male	−.0002	−.0005
hot.deck	Male	−.0001	−.0004
amelia	Male	+.0000	−.0001
mice	Male	+.0000	−.0001
na.omit	Male	−.0254	−.0350
true	Inc	16.6140	6.4810
hd.ord	Inc	−.0346	−.0114
hot.deck	Inc	−.0657	−.0241
amelia	Inc	−.0018	−.0007
mice	Inc	−.0027	−.0005
na.omit	Inc	−.6432	−.2888
true	Age	50.0410	52.8230
hd.ord	Age	−.0887	−.0704
hot.deck	Age	−.1951	−.1532
amelia	Age	+.0010	−.0021
mice	Age	−.0016	−.0055
na.omit	Age	−.8025	−.4330
true	Black	.0790	.0950
hd.ord	Black	+.0001	−.0001
hot.deck	Black	+.0000	−.0001
amelia	Black	+.0000	+.0001
mice	Black	+.0001	+.0001
na.omit	Black	−.0103	−.0122
true	Empl	.6610	.4370
hd.ord	Empl	+.0007	+.0003
hot.deck	Empl	+.0008	+.0001
amelia	Empl	+.0001	+.0000
mice	Empl	+.0001	−.0001
na.omit	Empl	−.0109	−.0328
true	Religious	.6460	.6420
hd.ord	Religious	−.0001	−.0002
hot.deck	Religious	−.0005	−.0003
amelia	Religious	+.0000	−.0001
mice	Religious	+.0000	−.0002
na.omit	Religious	−.0174	−.0241
true	Married	.5290	.6310
hd.ord	Married	−.0001	−.0002
hot.deck	Married	+.0003	+.0001
amelia	Married	−.0001	−.0002
mice	Married	−.0001	−.0002
na.omit	Married	−.0390	−.0324
true	OwnHome	.6820	.7010
hd.ord	OwnHome	−.0005	−.0001
hot.deck	OwnHome	−.0001	+.0002
amelia	OwnHome	+.0000	+.0001
mice	OwnHome	−.0001	+.0000
na.omit	OwnHome	−.0341	−.0295
true	Rally	.0830	—
hd.ord	Rally	+.0000	—
hot.deck	Rally	−.0002	—
amelia	Rally	+.0001	—

mice	Rally	+.0002	—
na.omit	Rally	−.0186	—
true	Donate	.1390	—
hd.ord	Donate	−.0003	—
hot.deck	Donate	−.0007	—
amelia	Donate	−.0001	—
mice	Donate	+.0000	—
na.omit	Donate	−.0310	—
true	Gay	—	.0420
hd.ord	Gay	—	+.0001
hot.deck	Gay	—	+.0000
amelia	Gay	—	+.0000
mice	Gay	—	+.0001
na.omit	Gay	—	−.0113
true	StudLoans	—	.1910
hd.ord	StudLoans	—	+.0001
hot.deck	StudLoans	—	+.0001
amelia	StudLoans	—	−.0001
mice	StudLoans	—	−.0001
na.omit	StudLoans	—	−.0146

The same can be observed for the interval variables, with `hd.ord` again claiming last place across both 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`: −.0346, −.0114 vs. −.0470, −.0130 for `Inc` and −.0887, −.0704 vs. −.1391, −.0883 for `Age`.

Table 3.10 shows the results of imputing both 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 (−.0133 `hd.ord` vs. −.0130 `mice` CCES `Dem`), sometimes less so (−.0142 `hd.ord` vs. −.0127 `mice` ANES `Dem`). `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: −.0142,−.0133 vs. −.0114, −.0099 for `Dem` and −.0180, −.0162 vs. −.0136, −.0116 for `Male`. The results for

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

Method	Variable	ANES	CCES
true	Dem	.3420	.3770
hd.ord	Dem	–.0142	–.0133
hot.deck	Dem	–.0155	–.0131
amelia	Dem	–.0136	–.0131
mice	Dem	–.0127	–.0130
na.omit	Dem	–.0211	–.0185
true	Male	.4890	.4830
hd.ord	Male	–.0180	–.0162
hot.deck	Male	–.0172	–.0160
amelia	Male	–.0170	–.0154
mice	Male	–.0170	–.0153
na.omit	Male	–.0233	–.0241
true	Inc	16.6140	6.4810
hd.ord	Inc	–.2481	–.1034
hot.deck	Inc	–.3174	–.1303
amelia	Inc	–.1555	–.0741
mice	Inc	–.1568	–.0730
na.omit	Inc	–.3114	–.1513
true	Age	50.0410	52.8230
hd.ord	Age	–.6020	–.4844
hot.deck	Age	–.8997	–.7259
amelia	Age	–.3103	–.2831
mice	Age	–.2994	–.2702
na.omit	Age	–.6726	–.6482

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 interval variables are consistent with the previous analyses. In addition, note that `na.omit` performs better than `hd.ord` for `Age` in both datasets and for ANES `Inc`. Once more, `hd.ord` consistently performs slightly worse with more than two ordinal variables: –.2481, –.1034 vs. –.2299, –.0928 for `Inc` and –.6020, –.4844 vs. –.6319, –.4596 for `Age`.

Table 3.11 shows the results of imputing both 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 (–.0061 `hd.ord` vs. –.0060 `mice` ANES `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: –.0050, –.0053 vs. –.0049, –.0046 for `Dem` and –.0061, –.0058 vs. –.0055, –.0049 for `Male`.

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

Method	Variable	ANES	CCES
true	Dem	.3420	.3770
hd.ord	Dem	–.0050	–.0053
hot.deck	Dem	–.0053	–.0053
amelia	Dem	–.0047	–.0049
mice	Dem	–.0044	–.0048
na.omit	Dem	–.0097	–.0097
true	Male	.4890	.4830
hd.ord	Male	–.0061	–.0058
hot.deck	Male	–.0057	–.0056
amelia	Male	–.0060	–.0054
mice	Male	–.0060	–.0054
na.omit	Male	–.0088	–.0116
true	Inc	16.6140	6.4810
hd.ord	Inc	–.0841	–.0349
hot.deck	Inc	–.1142	–.0473
amelia	Inc	–.0540	–.0250
mice	Inc	–.0549	–.0249
na.omit	Inc	–.2112	–.0982
true	Age	50.0410	52.8230
hd.ord	Age	–.2124	–.1607
hot.deck	Age	–.3329	–.2424
amelia	Age	–.1194	–.1135
mice	Age	–.1141	–.1102
na.omit	Age	–.2978	–.0974
true	Black	.0790	.0950
hd.ord	Black	–.0036	–.0045
hot.deck	Black	–.0040	–.0042
amelia	Black	–.0041	–.0044
mice	Black	–.0037	–.0041
na.omit	Black	–.0050	–.0066
true	Empl	.6610	.4370
hd.ord	Empl	–.0042	–.0057
hot.deck	Empl	–.0036	–.0058
amelia	Empl	–.0034	–.0045

mice	Empl	–.0033	–.0044
na.omit	Empl	–.0022	–.0122
true	Religious	.6460	.6420
hd.ord	Religious	–.0045	–.0044
hot.deck	Religious	–.0047	–.0042
amelia	Religious	–.0043	–.0045
mice	Religious	–.0043	–.0045
na.omit	Religious	–.0055	–.0077
true	Married	.5290	.6310
hd.ord	Married	–.0047	–.0043
hot.deck	Married	–.0042	–.0040
amelia	Married	–.0046	–.0039
mice	Married	–.0045	–.0039
na.omit	Married	–.0120	–.0095
true	OwnHome	.6820	.7010
hd.ord	OwnHome	–.0037	–.0035
hot.deck	OwnHome	–.0030	–.0030
amelia	OwnHome	–.0031	–.0033
mice	OwnHome	–.0030	–.0033
na.omit	OwnHome	–.0112	–.0088
true	Rally	.0830	—
hd.ord	Rally	–.0047	—
hot.deck	Rally	–.0047	—
amelia	Rally	–.0045	—
mice	Rally	–.0044	—
na.omit	Rally	–.0080	—
true	Donate	.1390	—
hd.ord	Donate	–.0057	—
hot.deck	Donate	–.0060	—
amelia	Donate	–.0054	—
mice	Donate	–.0053	—
na.omit	Donate	–.0122	—
true	Gay	—	.0420
hd.ord	Gay	—	–.0027
hot.deck	Gay	—	–.0026
amelia	Gay	—	–.0028
mice	Gay	—	–.0027
na.omit	Gay	—	–.0037
true	StudLoans	—	.1910
hd.ord	StudLoans	—	–.0067
hot.deck	StudLoans	—	–.0064
amelia	StudLoans	—	–.0057
mice	StudLoans	—	–.0056
na.omit	StudLoans	—	–.0082

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 both 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: –.0841, –.0349 vs. –.0899, –.0350 for `Inc` and –.2124, –.1607 vs. –.2571, –.1732 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 with the CCES data. The results are shown in Table 3.12. `hd.ord` performs comparatively well for 50 percent missing data `Dem` (–.0006 vs. +.0000 `amelia`) but falls short for `Male` (–.0020 vs. +.0000 `mice`). `hd.ord` also represents the second-worst method for both percentages for all ordinal and interval variables. `amelia` and `mice` show virtually identical results and are hardly affected by the increase in missingness for the ordinal variable. This changes somewhat for the interval variable `Age`. Nonetheless, `amelia` and `mice` far outperform the other methods. Note that `amelia` significantly outperforms `mice` for `Age` for both percentages.

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 both datasets (Table 3.13) and by the percentage of missingness for the CCES data (Table 3.14). Both analyses are conducted MAR, with the latter only for five imputed variables. All running times are given in minutes and apply to all 1,000 imputation iterations combined.

Table 3.13 shows `hd.ord` and `hot.deck` with virtually identical running times for both 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 3.4 times slower than `hd.ord` (for the ANES data with five imputed variables). The highest

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

Method	Variable	20% NA	50% NA
true	Dem	.3770	.3770
hd.ord	Dem	+.0000	–.0006
hot.deck	Dem	–.0004	–.0012
amelia	Dem	+.0001	+.0000
mice	Dem	+.0002	+.0002
na.omit	Dem	–.0229	–.0516
true	Male	.4830	.4830
hd.ord	Male	–.0011	–.0020
hot.deck	Male	–.0014	–.0036
amelia	Male	–.0001	–.0001
mice	Male	–.0001	+.0000
na.omit	Male	–.0414	–.1032
true	Interest	3.3290	3.3290
hd.ord	Interest	–.0125	–.0336
hot.deck	Interest	–.0196	–.0538
amelia	Interest	+.0003	+.0001
mice	Interest	+.0000	–.0003
na.omit	Interest	–.0724	–.2014
true	Inc	6.4810	6.4810
hd.ord	Inc	–.0259	–.0809
hot.deck	Inc	–.0407	–.1240
amelia	Inc	–.0004	+.0010
mice	Inc	–.0002	+.0019
na.omit	Inc	–.2468	–.5958
true	Age	52.8230	52.8230
hd.ord	Age	–.2616	–.7685
hot.deck	Age	–.3895	–1.1573
amelia	Age	–.0033	–.0075
mice	Age	–.0073	–.0137
na.omit	Age	–1.2361	–3.1442

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.

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 to 19.2 times slower (for the ANES data with five imputed variables). Table 3.14 shows that

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	ANES	CCES
hd.ord	2.632	2.778	2.614	2.679
hot.deck	2.628	2.786	2.640	2.690
amelia	9.052	10.611	9.038	10.345
mice	50.445	57.205	104.143	113.390

`hd.ord` and `hot.deck` remain the fastest methods across both percentages of missingness, but the gap to `amelia` and `mice` narrows as the missingness increases. `amelia` improves from 3.8 times slower than `hd.ord` for 20 percent NA to 2.2 times slower for 50 percent NA. Similarly, `mice` speeds up from 20.6 to 12.3 times slower.

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

Method	20% NA	50% NA
hd.ord	2.780	12.920
hot.deck	2.790	13.030
amelia	10.610	28.770
mice	57.210	158.840

3.5 Brief Evaluation of Shortcomings in `hd.ord`

Given `hd.ord`'s insufficient performance, we need to speculate about possible reasons. Perhaps the importance of the uneven distances between ordinal variable categories is over-emphasized in the literature, with the distances potentially being not as uneven as previously thought. The fact that `hd.ord` consistently performs worse when a second ordinal variable is added to the `polr` treatment seems to point in this direction. To further investigate this possibility, I test the influence of `polr` on regression coefficient transformations with data from previous publications.

The first column of Table 3.15 shows a replication of a linear model estimated by Bartels (1999) on the 1992 ANES data. Bartels regresses several explanatory variables on `Campaign Interest` (this estimation corresponds to the column "Panel" in Table 2 of the original publication). As we can see, the ordinal variable `Education` is part of the explanatory variables in the model. To test the transformations implemented by an ordered probit model for ordinal variables, I remove the current dependent variable (`Campaign Interest`) and replace it with `Education`. The resulting model is then estimated as a linear regression (column 2) and an ordered probit regression (column 3). To determine how statistically distinct the variables on the right-hand side in these models (`Age`, `Income`, `Black`, `Female`, `Partisan strength`, `Days before election`) are, I simulate the posterior distribution of each β coefficient from both regressions in columns 2 and 3 as a normal distribution with mean $\hat{\beta}$, standard error $SE(\hat{\beta})$, and $n = 100,000$. I then plot the overlapping distributions of each linear regression coefficient with the corresponding ordered probit regression coefficient to assess the posterior percentage of overlay. The results are shown in Figure 3.6. We observe a small posterior percentage of overlay for the distributions of `Age` and `Income` ($<$ 10 percent) and a large posterior percentage of overlay for the distributions of `Black`, `Female`, `Partisan strength`, and `Days before election` ($>$ 40 percent). Since a large percentage indicates little difference in significance between the two sets of coefficients, these results appear to provide further proof against the importance of re-estimating ordinal variable categories with ordered probit approaches.

Based on evidence from these brief investigative analyses, it thus seems that uneven

Table 3.15. lm and polr Differences in 1992 ANES Data as used by Bartels (1999)

	Dependent variable:		
	Campaign Interest	Education	Education
	OLS	*OLS*	*ordered logistic*
Education	0.023		
	(0.004)		
Age	0.002	−0.032	−0.021
	(0.001)	(0.004)	(0.003)
Income	0.071	4.092	3.133
	(0.037)	(0.245)	(0.204)
Black	−0.028	−0.865	−0.673
	(0.028)	(0.198)	(0.150)
Female	−0.055	−0.058	−0.054
	(0.018)	(0.133)	(0.102)
Partisan strength	0.214	0.290	0.196
	(0.027)	(0.198)	(0.152)
Days before election	−0.001	0.014	0.012
	(0.0005)	(0.004)	(0.003)
Constant	0.391	−0.388	
	(0.038)	(0.273)	
Observations	1,359	1,359	1,359
R^2	0.114	0.248	
Adjusted R^2	0.109	0.245	
Residual Std. Error	0.331 (df = 1351)	2.401 (df = 1352)	
F Statistic	24.750 (df = 7; 1351)	74.507 (df = 6; 1352)	

distances between ordinal variable categories might not actually be of crucial importance when it comes to missing data imputation. These are mere preliminary conclusions, however, and more research is needed to further corroborate this tentative impression.

3.6 Conclusion

I set out to improve multiple imputation results with ordinal variables by accounting for the unevenly spaced ordering contained in ordinal variables. I did so by adapting

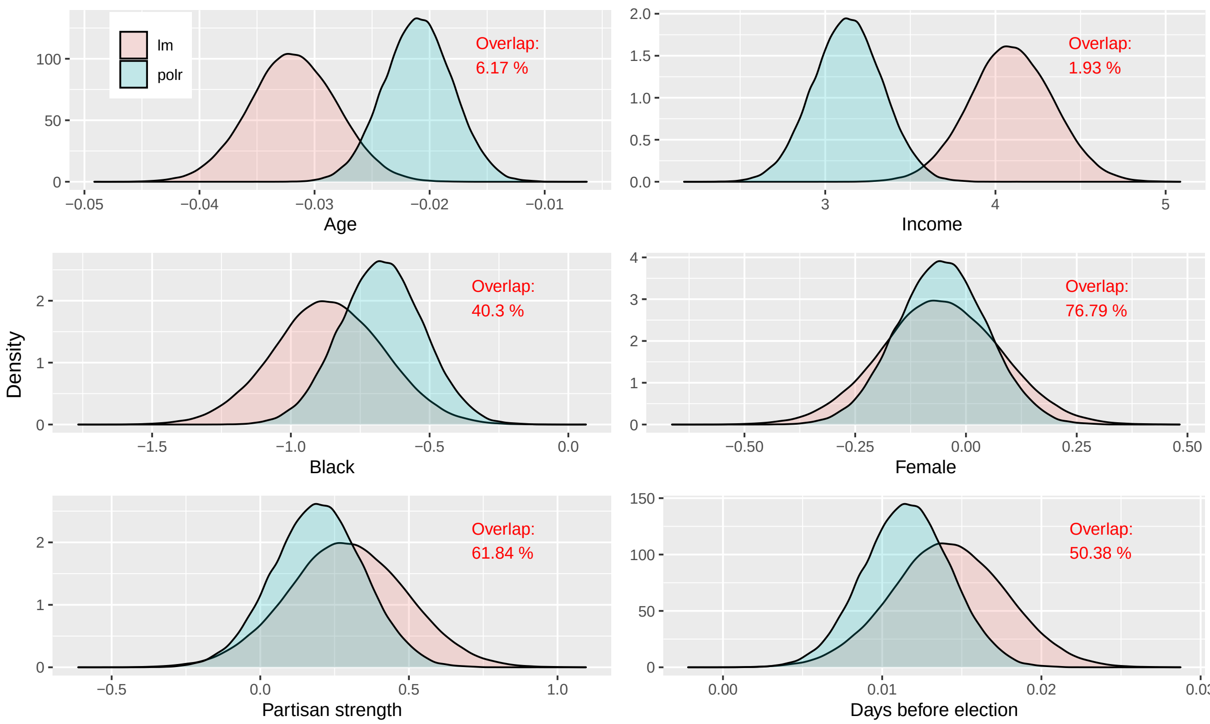

Figure 3.6. Distributions of lm and polr coefficients

the multiple hot deck imputation function `hot.deck` to treat ordinal variables with an ordered probit model in order to estimate numeric thresholds from an assumed underlying latent continuous variable. The results clearly show diverging outcomes from the different algorithms, with each algorithm behaving differently under changing circumstances. While `hd.ord` did not yield the desired improvements, my analysis provides in-depth insights into and corroborates the robustness of multiple imputation implementations like `amelia` and `mice` for a variety of variable types in survey settings.

`hd.ord` performs on par with some binary variables but overall worse than `amelia` and `mice` for data MAR with 5 and with 12 variables with missing values. The difference between the methods is somewhat less pronounced for 12 variables, which could possibly be explained by the thinner spread of missing values across a higher number of variables.

The results for the MNAR analyses paint a more mixed but overall unchanged picture. `hd.ord` performs somewhat better for binary variables but remains the least accurate method for interval and ordinal variables. Increasing the number of ordinal variables included in the `polr` treatment did not have a positive effect on the performance of `hd.ord`. In fact, `hd.ord` consistently performed slightly worse for all variables for both datasets for all mechanisms of missingness. Unlike `amelia` and `mice`, `hd.ord`'s performance also drastically worsens when the percentage of missingness in the data increases to 50 percent and above.

On the positive side, `hd.ord` performs multiple imputation much more quickly: `amelia` and `mice` are at least 3.4 and 19.2 times slower than `hd.ord` for 20 percent of missing data, respectively. This speed gain is of dubious value, however. While it is necessary to iterate multiple imputation runs many times over for simulation purposes like this one, users likely will not do so, which greatly diminishes the computing time saved. Even if users do opt to run multiple imputation for 1,000 times, the differences in terms of absolute time with `amelia` are not so great as to be impractical. `amelia` on average took a little over nine minutes to compute 1,000 iterations of multiple imputation, regardless of the dataset in question. In absolute terms, this is not a lot of time for the general user to invest in data preparation. With an average of around 110 minutes, the same cannot be said for `mice`. Nonetheless, `hd.ord`'s speed gain over `amelia` cannot be considered enough reason to choose `hd.ord` over `amelia`, unless the data consists of exclusively binary variables where the difference between `hd.ord` and `amelia` are small.

Given all the above, the result of this quality comparison of major missing data solutions is a clear endorsement of `amelia`. It performs well for all types of variables in all stages of missingness and does so in a reasonably short amount of time. The combination of EM with bootstrapping clearly represents a great improvement in terms of speed over IP used in `mice`. While it offers a wealth of sophisticated options for specialized users,`amelia`'s default out-of-the-box settings are simple and intuitive for general users. On top of that, it is notable that `amelia` produces better results than `na.omit` when data is MNAR, i.e. it performs well in a setting it was not designed for. `amelia` thus represents the best `R` solution to problems of missing data for general users.

CHAPTER 4

MORALS, SELF-INTEREST, AND FRAME STRENGTH

4.1 Introduction

In today's world, we rarely consume information directly, for instance by attending a demonstration or listening to a talk. Instead, most of the information we consume is mediated through television, radio, blogs, social media, messenger apps, and so forth. Mediated information by definition only represent a selective fragment of reality. The messenger – news anchors, pundits, influencers, politicians etc. – makes conscious and unconscious choices what information to distribute and how to present this information to us. The mediated message that eventually reaches us has been manipulated. Numerous strategies to carry out such manipulations exist. One of these strategies is framing.

Framing is the practice of presenting an issue to affect the way people see it (Chong & Druckman, 2007). It reorganizes existing information already present in people's minds and attempts to direct people's attention towards particular considerations (Druckman & Nelson, 2003). Numerous experiments have shown that frames can have substantial influence in moving people's opinions (Chong & Druckman, 2010; Sniderman & Theriault, 2004). Despite an abundance of experimental framing research, however, we don't know why some frames are successful in affecting people's opinions and others are not. In other words, we don't know what aspects make frames strong. I provide an avenue of clarification by testing the influence of morals and self-interest in frames.

For some time, scholars have postulated the 'death' of self-interest to explain issue positioning: People eschew their self-interest to defend and vote according to their morals instead, as they consider them more important (Frank, 2004; Haidt, 2008). Others, however, cast doubt on these claims: Morals and cultural values do not outweigh self-interest concerns, which in turn represent a significant factor in determining issue positioning (Bartels, 2005, 2006). Which is it? Have moral concerns displaced self-interest? Does self-interest matter more than postulated by many? I intend to shed some light on these questions with an online framing survey experiment that directly juxtaposes morals and self-interest. Morals are assessed on the basis of Moral Foundations Theory (Haidt, 2012). To measure self-interest, I widen the widely adopted definition of self-interest to include goals other than material gains, i.e. personal autonomy, health/safety, wealth, and status (Gintis, 2017; Weeden & Kurzban, 2017). Demographic information and conviction measurements are used to analyze the influence of each concept. To my knowledge, this setup has not been studied before. In addition to the substantive analysis, both methods developed in chapters II and III are applied and their performance is analyzed.

4.2 Theory

4.2.1 Framing

People usually do not have stable, consistent, and informed opinions (Converse, 1964; Zaller, 1992). It is therefore possible to influence people's opinion through communication. One of the ways to do so is framing. 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). When frames influence people's opinion about an issue, we speak of a framing effect. Two different types of framing effects have evolved in the literature: Equivalency and emphasis framing effects.

Equivalency framing effects involve phrasing the same logical content in different ways, i.e. "casting the same information in either a positive or negative light" (Druckman, 2004, p. 671). One example is Tversky & Kahneman (1981)'s death and survival experiment: Presented with a hypothetical disease outbreak in the US that is expected to kill 600 people, respondents were asked to choose between alternating programs. The first group chose between programs A and B. In program A, "200 people will be saved". In program B, "there is 1/3 probability that 600 people will be saved and 2/3 probability that no people will be saved". The second group chose between programs C and D. In program C, "400 people will die". In program D, "there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die". The factual content given to both groups is of course identical: In both groups, the first program results in the death of 400 people, and the second program has a 66 percent chance of killing all 600 people. The differences is the light cast on this information. The frames given to group 1 focus on lives saved, whereas the frames presented to group 2 center on lives lost. Rationally, both are the same. However, this difference in equalency framing matters: The majority in group 1 chose program A. The majority in group 2 chose program D. Other work focuses on policy evaluations, rather than risk preferences. For example, respondents were found to rate a proposed economic program more favorably when it is framed as resulting in 95 percent employment, instead of it's logical equivalent, 5 percent unemployment (Quattrone & Tversky, 1988). Equivalency framing effects have been likened by some to survey question wording effects (Zaller, 1992). Examples here are the alternate use of "not allow" and "forbid" in reference to Communist speech in the 1970s, which can arguably be described as substantively and logically equivalent (Bartels, 2003, p. 61).

Emphasis framing effects concern putting the stress on a particular aspect, viewpoint, or consideration about an issue in the attempt to get people to look at the issue in the proposed way. Describing the invasion of Iraq as freeing the population and gifting them democracy, as done by the Bush administration, for instance, emphasizes a very different viewpoint than a possible alternative narrative that focuses on the economic benefits of gaining access to oil fields and trade opportunities. Unlike equivalency frames, emphasis frames are not logically identical ways of phrasing the same issue. Instead, they highlight qualitatively different considerations of that issue (Druckman, 2001c, 2001b; Kinder & Sanders, 1996; Levin, Schneider, & Gaeth, 1998; Nelson & Kinder, 1996; Sniderman & Theriault, 2004), forcing people to think about the relative importance of the differing considerations suggested by the frame (Nelson, Clawson, & Oxley, 1997).

My focus lies on emphasis frames. An abundance of experiments has shown that emphasis frames elicit significant changes in issue positioning (Chong & Druckman, 2010, 2013; Druckman, 2001b; Druckman, Fein, & Leeper, 2012; Druckman & Nelson, 2003; Druckman et al., 2013; Nelson et al., 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'. However, despite the mass of experimental framing research, we still have little insight into what makes an emphasis frame strong. We don't know why some emphasis frames elicit effects and others don't, i.e. we don't know why some frames are strong and other frames are weak. We just know that strong frames are stronger than weak frames. Chong & Druckman (2007) attempt a differentiation into weak and strong frames. They ask respondents in a pre-test what arguments the respondents consider strong or weak for a variety of chosen issues. Frames containing strong arguments are then deemed strong frames, while frames containing weak arguments are considered weak frames. These weak and strong frames are then used in the subsequent framing experiment. While it is no doubt laudable to explore which arguments in selected political issues are deemed more persuasive than others, this setup does not provide insights as to what actually makes a frame strong or weak. We simply know which arguments (to then be embedded in frames) are considered to be strong and weak, but not why. A major challenge for framing research thus still "concerns the identification of factors that make a frame strong" (Chong & Druckman, 2007, p. 116).

In the attempt to address this challenge, scholars have increasingly adopted theory-driven approaches that deduce likely strong frames and subsequently empirically evaluate their theoretical expectations. (Arceneaux, 2012, p. 280) for instance theorizes that "individuals are more likely to be persuaded by political arguments that evoke cognitive biases" and asserts that messages which highlight out-group threats resonate with respondents to a greater extent. Druckman & Bolsen (2011) report that adding factual information to messages about carbon nanotubes does nothing to enhance their strength and actually makes the message weaker. Feinberg & Willer (2013) assess whether frames are stronger when they cohere with an individual's personal value system. They frame environmental issues as a matter of moral 'purity', a theme that supposedly correlates with conservative ideology, and find this approach leads to increased conservative support of environmental policies. I attempt to provide a further avenue of clarification by testing the theory that morals contribute to frame strength.

4.2.2 Morals

Two of the main ingredients of public opinion are commitment to principles (i.e. morals) and interests that citizens see at stake (Kinder, 1998). Morals play a role in establishing one's self-concept and identity. Moral arguments can achieve a much higher emotional connection with people because they invoke people's values and feelings (Bauman & Skitka, 2009; Haidt, 2003b; Skitka, Bauman, & Sargis, 2005; Skitka & Wisneski, 2011; Tatalovich & Daynes, 2011). Emotions play an important role in how people conceptualize moral stimuli since people feel emotionally committed to their values (Ryan, 2014b; Suhay, 2008; Tetlock, Peterson, & Lerner, 1996). Appeals to values/morals tend to be more successful than non-moral appeals "especially when the moral principles invoked resonate with the individuals targeted by the appeal" (Feinberg & Willer, 2013, p. 57). 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). 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 widely argued that people rely to a disproportionate extent on moral arguments to form their opinions (Ryan, 2014a, 2014b; Smith, 2002).

Moral arguments are ubiquitous in political issues because they are essential to how people perceive and make sense of the world around them (Cervone, 2004; Frank, 2004; Mooney, 2001; Skitka, Bauman, & Mullen, 2008; Tatalovich, Smith, & Bobic, 1994). A lot of scholarly work tries to establish a distinction between moral and non-moral issues, yet agreement on the definition of a 'moral issue' remains elusive. To some, an issue is moral when at least one side sees the issue as threatening a core value/principle and/or uses moral arguments to support their position (Haider-Markel & Meier, 1996; Mooney, 2001). To others, moral issues are based on values rooted deeply within people's belief systems (Biggers, 2011; Glick & Hutchinson, 2001). Yet another strain sees morality as intrinsic to some issues, i.e. nontechnical issues that are easy to understand or issues that concern fundamental aspects of life and death (Studlar, 2001; Tavits, 2007). Yet others again assert that economic issues are not to be considered moral (Abramowitz, 1995; Engeli, Green-Pedersen, & Larsen, 2012; Mooney & Lee, 1995; Tatalovich & Daynes, 2011) despite evidence to the contrary: Ryan (2014b) finds evidence that some people perceive distinctly economic issues such as labor relations laws or social security reform in moral ways. In addition, studies on abortion also show that, despite the media's contrary depiction, not everyone conceives of it as a moral issue (Mullen & Skitka, 2006; Skitka, 2002; Skitka et al., 2005). Assertions of moral and non-moral issues are, in my eyes, misleading and unnecessary. Instead, we should concern ourselves with the moral and non-moral content of each respective issue. I thus eschew moral and non-moral issue definitions and focus on moral and self-interest frames within issues, rather than the moral nature of the issues themselves.

In normative terms, we differentiate between prescriptive and proscriptive morality, which are based on approach-avoidance differences in self-regulation (Carnes & Janoff-Bulman, 2012). Prescriptive morality is sensitive to positive outcomes and focused on what we should do. Proscriptive morality is sensitive to negative outcomes and focused on what we should not do (Janoff-Bulman et al., 2009). Based on this categorization, Haidt and his colleagues developed Moral Foundations Theory (MFT), which encompasses cognitive foundations of moral matrices and has become widely adopted in moralization literature (Clifford, Iyengar, Cabeza, & Sinnott-Armstrong, 2015; Clifford & Jerit, 2013; Feinberg & Willer, 2013; Graham, Haidt, & Nosek, 2009; Haidt, 2001, 2003a, 2003b, 2007, 2008, 2012; Haidt & Graham, 2007; Haidt & Joseph, 2004, 2007; Haidt & Kesebir, 2010; Hofmann, Wisneski, Brandt, & Skitka, 2014; Koleva, Graham, Iyer, Ditto, & Haidt, 2012;

Schein & Gray, 2017). These foundations are shown in Table 4.1. Moral intuitions are

Table 4.1. Foundations of Moral Intuitions

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
Authority	Submitting to tradition/authority	*Subversion*	Resisting tradition/authority
Sanctity	Repulsion at disgust	*Degradation*	Enjoyment of disgust
Liberty	Acting without constraint	*Oppression*	Dominate/Constrain others

Positive and negative foundations are conceptual opposites.

a powerful psychological mechanism that underlies issue positions (Koleva et al., 2012). Haidt and his colleagues argue that they intuitions are rooted in evolved mechanisms. They propose that Care/Harm developed as the response to the adaptive challenge of protecting and caring for children. It makes us sensitive to signs of suffering and activates us to care for those who are being harmed. Fairness/Cheating developed as the response to acts of cooperation or selfishness that people show towards us. It makes us trust fair actors and punish cheaters. Loyalty/Betrayal developed to meet the adaptive challenge of forming and maintaining cohesive coalitions. It makes us reward loyalty and ostracize those who betray us. Authority/Subversion developed as the response to forging beneficial relationships within social hierarchies. It makes us sensitive to signs that people do not behave in accordance with their social position. Sanctity/Degradation developed as a response to the challenge of living in world full of pathogens and parasites. It binds us into moral communities by shared notions of sacredness. Liberty/Oppression developed as a response to the challenge of living in small groups whose members can dominate and constrain us. It makes us value freedom and reject repression.

Because of the deep connection between morals and emotions, people have a tendency to internalize and defend their values. This "makes it very difficult for people to consider the possibility that there might really be more than one form of moral truth, or more than one valid framework for judging people or running a society" (Haidt, 2012, p. 111). Haidt distinguishes the use of these foundations by ideology. According to him, liberals tend to utilize a three-foundation morality, whereas conservatives allegedly lean towards using all six foundations. Additionally, both sides are said to focus on different aspects in the three foundations they share (Care/Harm, Fairness/Cheating, Liberty/Oppression): Conservatives are more likely to emphasize society's protection and security and aim to prevent losses and generally negative outcomes, while liberals tend to emphasize the provision of welfare for others and aim to advance gains and generally positive outcomes. Both seek normatively 'good' outcomes for society, but with different orientations (Janoff-Bulman, 2009; Janoff-Bulman & Carnes, 2013; Janoff-Bulman et al., 2009).

While the liberal/conservative divide of course plays a role in opinion formation, it is not central to moral analyses. While liberals and conservatives are said to display differing tendencies towards each individual foundation, the majority of people value all foundations. The foundations themselves, not party ID, are what matter most. I use the foundations as the building block of moral arguments in my juxtaposition of morality and self-interest.

4.2.3 Self-Interest

Two of the main ingredients of public opinion are commitment to principles (i.e. morals) and interests that citizens see at stake (Kinder, 1998). Section 4.2.2 covers morals. This section addresses self-interest.

Self-interest refers to actions or behaviors that elicit personal benefit. When acting in self-interest, individuals look out for themselves in life and act on the grounds of personal gains (Arrow, 1967; Coleman, 1986; Downs, 1957; Ferejohn & Fiorina, 1974; Fiorina, 1977; Olson, 1965; Riker & Ordeshook, 1968; Sears & Funk, 1991). In neoclassical economics, this claim is extended to the overall postulation of human beings as *homines* *economici* (Coleman, 1986); creatures who act strictly rationally with the intent to maximize utility at all times. In political psychology, scholars instead focus on the importance of self-interest as a motivation for opinions on political issues.

In the 1960s, self-interest was considered a major determinant of opinions on political issues. Campbell, Converse, Miller, & Stokes (1960) for instance report that people with less income and education are more likely to support a strong government role in the provision of social welfare, stating that “people presented with certain policy alternatives can do a reasonable job of selecting responses that appear to further their self-interest” (p. 208). Empirical research from the 1970s and 1980s appears to confirm this (Batson, 1991; Etzioni, 1988; Feldman, 1982, 1984; Friedland & Robertson, 1989; Hirschman, 1985; Kohn, 1990; Lau & Sears, 1981; Lerner, 1980; Mansbridge, 1990b; Miller & Ratner, 1996; Schwartz, 1986; Sears & Citrin, 1985; Weatherford, 1983). Hawthorne & Jackson (1987) find that income level and tax bracket significantly predicts support for the tax cuts in the 1978 federal Tax Revenue Act. Lau, Coulam, & Sears (1983) identify a similar relationship for high taxpayers in their Massachusetts income tax study. Coughlin (1990) discover that higher income citizens oppose redistribution slightly more than lower income ones. Sears, Lau, Tyler, & Allen (1980) and Sears & Lau (1983) find evidence for pocketbook voting in their studies on government-guaranteed full employment. Courant, Gramlich, & Rubinfeld (1980) discover modest but consistent effects for the perceived property tax burden in Michigan. Lau, Sears, & Jessor (1990) find a significant relationship between perceived financial situation and anti-incumbent voting. The analysis by Sears & Lau (1983) reveals that perceived personal federal tax burden predicts support for the tax cuts specified in the 1979 Kemp-Roth bill. Bartels (2005) identifies a high subjective tax burden as the source of support for the 2001 and 2003 Bush tax cuts. Sears & Citrin (1985) similarly find that citizens with a high subjective tax burden act out of self-interest in the so-called California Tax Revolt at the end of the 1970s.

Despite this evidence, an abundance of criticism has claimed the inadequacy of self-interest as a major predictor of human behavior. Miller & Ratner (1996) call "homo economicus (…) a social construction, not a biological entity" (p. 45). Human behavior is argued to be too complex and much more than self-interest is claimed to come into play when people think about what they want in life (Batson, 1991; Friedland & Robertson, 1989; Kohn, 1990; Lerner, 1980; Mansbridge, 1990b, 1990a; Schwartz, 1986; Sen, 1977; Tyler, 1990b). The study of social movements, for instance, provides clear evidence that motivations other than self-interest also majorly account for human behavior (Mansbridge, 1990c). The overall most important reasons for humans' choices are affective and normative (Etzioni, 1988) and we need to shift our focus away from obsession with self-interest and towards moral values (Hirschman, 1985). The dominant theories of motivation are fundamentally based on personal gains and miss the influence that values and ideologies have on people's attitudes towards policies (Sears & Funk, 1990). People are more interested in the fairness of procedures and policies than the self-beneficial outcomes of these processes (Tyler, 1990a). It is possible and plausible that self-interest plays a role in opinion formation, but even then it does not play the only role (Chong, Citrin, & Conley, 2001; Citrin & Green, 1990; Miller, 1999; Sears & Funk, 1990). It is argued that self-interest is limited in explanatory scope (Chong et al., 2001; Huddy, 2013; Kinder, 1998; Lau & Heldman, 2009; Lewis-Beck, Jacoby, Norpoth, & Weisberg, 2008; Sears & Funk, 1990) and only applies to short-term material gains (Sears et al., 1980). As a result, self-interest is argued to be negligible: "Many political scientists used to assume that people vote selfishly, choosing the candidate or policy that will benefit them the most. But decades of research on public opinion have led to the conclusion that self-interest is a weak predictor of policy preferences" (Haidt, 2012, p. 85). Furthermore, standard demographic measures such as education, income, race, and gender are not deemed adequate indicators of self-interest, thus discounting findings from the 1960s

(Kinder, 1998; Sears & Funk, 1990).

These assertions strike me as odd. For one, empirical evidence appears to indicate that self-interest at least plays some role in economic aspects. For another, arguing that self-interest does not matter because of its exclusive focus on material gains after having defined self-interest in precisely this way has the feel of a snake biting its own tail. Of course self-interest does not account for many policy preferences when only material gains count as self-interest. After all, material gains are not all that humans desire. But should the definition of self-interest be so narrow? It seems perfectly reasonable that people have all kinds of self-serving preferences, from having more money to gaining prestige, having sex, or wishing success for one's children, among many others (Becker, 1996; Owens & Pedulla, 2014; Weeden, 2015; Weeden & Kurzban, 2016). Instead of a restricted categorization as purely economic gains, self-interest can be defined as "advancing any of a range of people's typical goals, whether directly involving material gain or not, whether involving immediate gain or something more subtle that advances someone's progress over the longer term" (Weeden & Kurzban, 2014, p. 38). By restricting self-interest to financial benefits and declaring ordinary demographic effects uninterpretable, scholars defined away any chance of self-interest being a major determinant of opinions on political issues. I see no reason for such restrictions.

Following evolutionary psychology, humans have self-interests that reach beyond material gains and include cultural as well as social aspects (Buss, 2015; Petersen, 2016). Self-interest thus applies not only to economic aspects but also to various social and cultural domains as well (Kurzban et al., 2010; Nteta, 2013). Self-interest is here defined as related to personal autonomy, health/safety, wealth, and status (Gintis, 2017). Demographics are argued to provide clues to particular outcomes that help or harm people's self-interests, which enables us to predict which opinion they have. By linking demographic measures with answers from the General Social Survey, Weeden & Kurzban (2014) find strong links between self-interest and meritocracy, discrimination, sexual lifestyles, religiosity, abortion, and marijuana legalization. Crucially, not all demographics provide clues about all kinds of issues. Religion can provide insights into people's views on abortion, but is unlikely to do so for immigration. Household income can predict views on taxes but probably not on legalizing marijuana (Weeden & Kurzban, 2017).

From this perspective, self-interest arguably matters most for issues that greatly affect people's everyday interests (DeScioli, Cho, Bokemper, & Delton, 2020; Kurzban, 2010; Sznycer et al., 2017). Rather than focusing on material gains – such as people's perceived tax burden (Bobo, 1983) – Weeden & Kurzban (2014) identify issues related to sex/reproduction (pre-marital sex, pornography, abortion, birth control access) and government spending on health care as prime candidates for a major influence of self-interest. Sexually conservative, married people with high regular church attendance can be predicted to be pro-life and against pre-marital sex because it aligns with their chosen lifestyle. Wealthy, white, heterosexual, Christian men can be predicted to oppose government redistribution of wealth as such a policy would endanger their own position. In both cases, it is in these individuals' self-interest to take these issue positionings because these issues directly relate to their lives. Such a connection is said to be less pronounced for issues such as the environment or the military, as they are deemed too remote for too many people to affect everyday lives.

4.3 Design

Each of the previous areas has been studied, but they have not been applied together. An abundance of framing research exists and we know that emphasis frames can be successful in directing people's attention towards particular considerations of an issue, but we still don't know why some emphasis frames succeed while others fail. We know that morals play a large role in people's issue positioning and that self-interest accounts for some human behavior, but we don't know how much self-interest and morals matter in direct comparison when people choose issue positions. I conduct a framing experiment that explores what we don't know. I investigate how morals and self-interest contribute to persuasive strength in emphasis frames.

Juxtaposing morals and self-interest in frames follows the recognition that people don't just want to satisfy their interests, but neither do they set aside their self-regarding preferences in favor of some non-self-regarding ones (Stoker, 1992). Both concepts are in constant movement, depending on the situation. People deal with the 'central problem of ethics': "how the lives, interests, and welfare of others make claims on us, and how these claims, of various forms, are to be reconciled with the aim of living our own lives" (Nagel, 1970, p. 142). Support for issues, for instance the welfare state, is based on "generalized and reciprocal self-interest" (Baldwin, 1990, p. 299), i.e. it depends on both value-based attitudes and immediate self-interest. Opinions can be influenced by self-interest or alterations in empathy towards others (Hacker, Rehm, & Schlesinger, 2013; Rehm, Hacker, & Schlesinger, 2012). It is often difficult to tell which is which: Frank (2004) claims Republicans have duped rural and working-class Americans into voting against their self-interest by supporting Republican tax cuts that negatively affect them financially. Bartels (2005) argues that they instead vote in unenlightened self-interest because they perceive their personal tax burden to be too high and (falsely) believe tax cuts can lower them. Haidt (2012), finally, states that they vote for their moral interest, as they don't want the country to devote itself primarily to the pursuit of social justice. It is difficult to assess which is correct.

Previous studies and experiments on moral framing that utilize MFT focus on partisan division (Clifford & Jerit, 2013; Feinberg & Willer, 2013; Koleva et al., 2012; Slothuus & Vreese, 2010; Wolsko, Ariceage, & Seiden, 2016). While party ID of course matters in any political analysis, partisanship is not central to analyses of moral concerns.

Instead, I utilize MFT to focus on the moral vs. self-interest juxtaposition within framing. I combine emphasis framing with aspects of MFT and the literature on self-interest.

Issues are selected from previous research and based on suitability to assess self-interest. In order to assess self-interest, all issues need to apply directly to respondents' daily lives. Following Weeden & Kurzban (2014), I choose government spending on healthcare and abortion. Everyone is affected by healthcare in their lives, so everyone has a vested self-interest in this issue. At the same time, how much the government should be involved in the provision of healthcare coverage is the subject of great debate. Similarly, the vast majority of people are affected by the issue of abortion, resulting in viable cause for self-interest, while the topic is also of great controversy in moral terms.

Each respondent is presented with one of five randomly assigned frames: An issue-supporting moral frame, an issue-opposing moral frame, an issue-supporting self-interest frame, an issue-opposing self-interest frame, or the control frame. The design consists of issue-supporting and issue-opposing frames in order to control for both signs in determining frame strength. The self-interest frames aim to activate each respondent's self-interest as identified by Weeden & Kurzban (2014). The moral frames are designed on the basis of Moral Foundations Theory. For simplicity and to avoid cluttering, I only use the foundation that most applies to each issue and to the majority of people. For both issues, that is Care/Harm. Support/Opposition for/to the frame's content is assessed on a 5-point Likert scale.

To measure self-interest, respondents answer a series of demographic questions related to each issue (questions Q12 to Q22 in the questionnaire). As described above, these demographics provide a match for each issue to determine underlying self-interest. For government spending on healthcare, for instance, those with less wealth and income, those with little education, minorities, women, and those with young children are likely to have a strong self-interest in the issue. Abortion is said to have major relationships with personal lifestyles as well as with religion and church attendance. All question wordings are based the ANES, CCES, and GSS surveys. I also measure the strength of Care/Harm for each respondent with questions taken from the Moral Foundations Questionnaire.

4.3.1 Expectations

Moral arguments are said to relate more strongly than non-moral ones since they are deeply connected to emotions. Non-moral arguments, on the other hand, are said to be more loosely connection to emotions and thus deemed less powerful (Haugtvedt & Wegener, 1994; Johnson & Eagly, 1989). This would suggest that moral frames exert a stronger influence on respondents' issue positioning than self-interest frames.

H1. Moral frames move people more than self-interest frames.

<!– Respondents with strongly held moral or self-interested attitudes are not moved by any frames.

H2. Respondents without strongly held attitudes are moved by self-interest and moral frames.

H3. Respondents without strongly held attitudes are moved further by moral frames than self-interest frames.

–>

4.4 Data

The designed frames are pre-tested for their moral and self-interest content on MTurk. The confirmed frames are subsequently applied in a survey experiment fielded online with Lucid.

4.4.1 Pre-Tests

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). Following this line of argument, pre-testing frames with participants who are not part of the main survey experiment is crucial in frame analysis. These participants are exposed to the designed frames and asked whether they reflect the core ideas behind morals and self-interest. 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. 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). This pre-test represents a thorough, robust safety net to ensure that the designed frames represent arguments relating to self-interest/morals and connect with participants, which in turn ensures meaningful survey results.

4.4.2 Survey Experiment

The survey experiment builds on the insights from the pre-test. It is fielded online for a random sample of U.S. adults with Lucid. Lucid has been shown to be equally reliable and performs well on a national scale in survey experiments (Coppock & McClellan, 2019). The survey collects demographic information on standard political science variables and applies all successfully pre-tested frames on all chosen issues. Both methods from the previous chapters are applied. Respondents are blocked on ordinal variables with the method from chapter I. The results are evaluated substantively to reveal the effect of morals and self-interest in respondents' issue positions. The ordinal variable imputation method from chapter II is also applied to artificially inserted missing data. This allows us to assess the performance of ordinal variable imputation method by comparing the imputation results with the true values of the completely observed data.

4.5 Results

4.5.1 Pre-Tests

4.5.2 Survey Experiment

Imputation

4.6 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

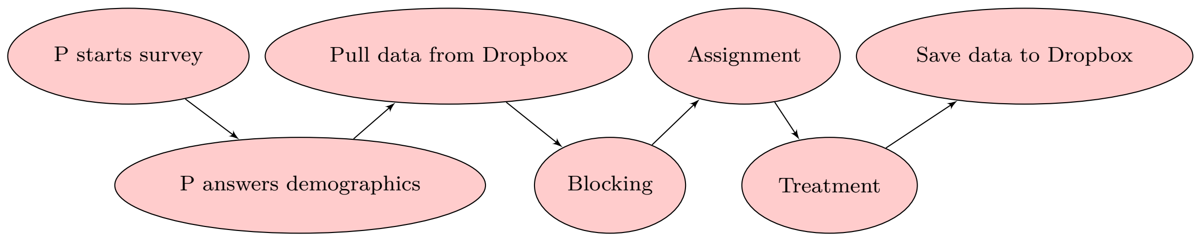

Figure A.1. Online survey experiment workflow

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 or Lucid.

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 percent

B.3 Speed

CCES 10,000 iterations +++ 5 variables: Dem, Male, Interest, Inc, Age MAR 20, 50 percent As can be seen in Table B.4, `hd.ord` and `hot.deck` show virtually identical

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.

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.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.orig	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.orig	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.orig	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.orig	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.orig	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.orig	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.orig	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.orig	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 .

`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. (1995). It's Abortion, Stupid: Policy Voting in the 1992 Presidential Election. *The Journal of Politics*, *57*(1), 176–186.

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.

Arrow, K. J. (1967). Values and Collective Decision-Making. In P. Laslett & W. G. Runciman (Eds.), *Philosophy, Politics, and Society* (pp. 215–232). New York, NY: Basic Books.

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.

Baldwin, P. (1990). *The Politics of Social Solidarity: Class Bases of the European Welfare State, 1875–1975.* Cambridge, MA: Cambridge University Press.

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.

Bartels, L. (2003). Democracy with Attitudes. In M. B. MacKuen & G. Rabinowitz (Eds.), *Electoral Democracy* (pp. 48–82). Ann Arbor, MI: University of Michigan Press.

Bartels, L. (2005). Homer Gets a Tax Cut: Inequality and Public Policy in the American Mind. *Perspectives on Politics*, *3*(1), 15–31.

Bartels, L. M. (1999). Panel Effects in the American National Election Studies. *Political Analysis*, *8*(1), 1–8.

Bartels, L. M. (2006). What's the Matter with 'What's the Matter with Kansas'? *Quarterly Journal of Political Science*, *1*, 201–226.

Batson, C. D. (1991). *The Altruism Question: Toward a Social Psychological Answer*. Hillsdale, NJ: Erlbaum.

Bauman, C. W., & Skitka, L. J. (2009). In the Mind of the Perceiver: Psychological Implications of Moral Conviction. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *The Psychology of Learning and Motivation* (Vol. 50, pp. 339–362). Cambridge, MA: Academic Press.

Becker, G. S. (1996). *Accounting for tastes*. Cambridge, MA: Harvard University Press.

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.

Biggers, D. R. (2011). When Ballot Issues Matter: Social Issue Ballot Measures and Their Impact on Turnout. *Political Behavior*, *33*(1), 3–25.

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

Bobo, L. (1983). Whites' Opposition to Busing: Symbolic Racism or Realistic Group Conflict? *Journal of Personality and Social Psychology*, *45*, 1196–1210.

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.

Buss, D. M. (2015). *The Handbook of Evolutionary Psychology*. Hoboken, NJ: Wiley-Interscience.

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.

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The American Voter*. Chicago, IL: Chicago University Press.

Carnes, N., & Janoff-Bulman, R. (2012). Harm, Help, and the Nature of (Im)Moral (In)Action. *Psychological Inquiry*, *23*, 137–142.

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

Cervone, D. (2004). The Architecture of Personality. *Psychological Review*, *111*, 183–204.

Chong, D., Citrin, J., & Conley, P. (2001). When Self-Interest Matters. *Political Psychology*, *22*(3), 541–557.

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

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

Chong, D., & Druckman, J. N. (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.

Citrin, J., & Green, D. P. (1990). The Self-Interest Motive in American Public Opinion. *Research in Micropolitics*, *3*, 1–28.

Clifford, S., Iyengar, V., Cabeza, R., & Sinnott-Armstrong, W. (2015). Moral Foundations Vignettes: A Standardized Stimulus Database of Scenarios Based on Moral Foundations Theory. *Behavior Research Methods*, *47*, 1178–1198.

Clifford, S., & Jerit, J. (2013). How Words Do the Work of Politics: Moral Foundations Theory and the Debate over Stem Cell Research. *The Journal of Politics*, *75*(3), 659–671.

Coleman, J. S. (1986). *Individual Interests and Collective Action*. Cambridge, MA: Cambridge University Press.

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.

Coughlin, R. (1990). The Economic Person in Sociological Context: Case Studies in the Mediation of Self-Interest. *Journal of Behavioral Economics*, *19*(2), 181–207.

Courant, P., Gramlich, E., & Rubinfeld, D. (1980). Why Voters Support Tax Limitation Amendments: The Michigan Case. *National Tax Journal*, *33*, 1–22.

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.

DeScioli, P., Cho, B., Bokemper, S., & Delton, A. W. (2020). Selfish and Cooperative Voting: Can the Majority Restrain Themselves? *Political Behavior*, *42*, 261–283.

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.

Downs, A. (1957). *An Economic Theory of Democracy.* New York, NY: Harper & Row.

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

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

Druckman, J. N. (2001c). The Implications of Framing Effects for Citizen Competence. *Political Behavior*, *23*(3), 225–256.

Druckman, J. N. (2004). Political Preference Formation: Competition, Deliberation, and the (Ir)relevance of Framing Effects. *American Political Science Review*, *98*, 671–686.

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

Druckman, J. N., 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. N., & Nelson, K. R. (2003). Framing and Deliberation. *American Journal of Political Science*, *47*(4), 729–745.

Druckman, J. N., 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.

Engeli, I., Green-Pedersen, C., & Larsen, L. T. (2012). Theoretical Perspectives on Morality Issues. In I. Engeli, C. Green-Pedersen, & L. T. Larsen (Eds.), *Morality Politics in Western Europe: Parties, Agendas, and Policy Choices* (pp. 5–26). New York, NY: Palgrave Macmillan.

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*.

Etzioni, A. (1988). *The Moral Dimension: Toward a New Economics*. New York, NY: The Free Press.

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. (2013). The moral roots of environmental attitudes. *Psychological Science*, *24*, 56–62.

Feldman, S. (1982). Economic Self-Interest and Political Behavior. *American Journal of Political Science*, *26*, 446–466.

Feldman, S. (1984). Economic Self-Interest and the Vote: Evidence and Meaning. *Political Behavior*, *6*, 229–252.

Ferejohn, J. A., & Fiorina, M. P. (1974). The Paradox of Not Voting: A Decision

Theoretic Analysis. *American Political Science Review*, *68*(2), 525–536. http://doi.org/10.2307/1959502

Fiorina, M. P. (1977). *Congress: Keystone of the Washington Establishment*. New Haven, CT: Yale University Press.

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. (2004). *What's the Matter with Kansas? How Conservatives Won the Heart of America*. New York, NY: Henry Holt.

Friedland, R., & Robertson, A. (1989). *Beyond the Marketplace: Rethinking Economy and Society*. New York, NY: Aldine; de Gruyter.

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.

Gintis, H. (2017). *Individuality and Entanglement: The Moral and Material Bases of Social Life*. Princeton, NJ: Princeton University Press.

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.

Glick, H. R., & Hutchinson, A. (2001). Physician-Assisted Suicide: Agenda Setting and the Elements of Morality Policy. In C. Z. Mooney (Ed.), *The Public Clash of Private*

Values: The Politics of Morality Policy (pp. 55–72). New York, NY: Chatham House Publishers.

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., Haidt, J., & Nosek, B. A. (2009). Liberals and Conservatives Rely on Different Sets of Moral Foundations. *Journal of Personality and Social Psychology*, *96*, 1029–1046.

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.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

Hacker, J. S., Rehm, P., & Schlesinger, M. (2013). The Insecure American: Economic Experiences, Financial Worries, and Policy Attitudes. *Perspectives on Politics*, *11*(1), 23–49.

Haider-Markel, D. P., & Meier, K. J. (1996). Politics of Gay and Lesbian Rights: Expanding the Scope of the Conflict. *The Journal of Politics*, *58*(2), 332–349.

Haidt, J. (2001). The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review*, *108*(4), 814–834.

Haidt, J. (2003a). The Emotional Dog Does Learn New Tricks: A Reply to Pizarro and Bloom. *Psychological Review*, *110*(1), 197–198.

Haidt, J. (2003b). 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. (2007). The New Synthesis in Moral Psychology. *Science*, *316*, 998–1002.

Haidt, J. (2008). Morality. *Perspectives on Psychological Science*, *3*, 65–72.

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

Haidt, J., & Graham, J. (2007). When Morality Opposes Justice: Conservatives Have Moral Intuitions that Liberals May Not Recognize. *Social Justice Research*, *20*, 98–116.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues. *Daedalus*, *133*(4), 55–66.

Haidt, J., & Joseph, C. (2007). The Moral Mind: How 5 Sets of Innate Moral Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules. In S. Laurence & S. Stich (Eds.), *The Innate Mind* (Vol. 3, pp. 367–391). New York, NY: Oxford University Press.

Haidt, J., & Kesebir, S. (2010). Morality. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of Social Psychology* (5th ed., pp. 797–832). Hoboken, NJ: Wiley-Blackwell.

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.

Haugtvedt, C. P., & Wegener, D. T. (1994). Message Order Effects in Persuasion: An Attitude Strength Perspective. *Journal of Consumer Research*, *21*(1), 205–218.

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.

Hawthorne, M., & Jackson, J. (1987). The Individual Political Economy of Federal Tax Policy. *American Political Science Review*, *81*, 757–774.

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

Hirschman, A. (1985). Against Parsimony: Three Easy Ways of Complicating some Categories of Economic Discourse. *Economics and Philosophy*, *1*(1), 7–21.

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in Everyday Life. *Science*, *345*(6202), 1340–1343.

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. (1998). *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.

Huddy, L. (2013). From Group Identity to Political Cohesion and Commitment. In

L. Huddy, D. O. Sears, & J. S. Levy (Eds.), *The Oxford Handbook of Political Psychology* (2nd ed., pp. 737–773). Oxford, UK: Oxford University Press.

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*.

Janoff-Bulman, R. (2009). To Provide or Protect: Motivational Bases of Political Liberalism and Conservatism. *Psychological Inquiry*, *20*(2-3), 120–128.

Janoff-Bulman, R., & Carnes, N. (2013). Surveying the Moral Landscape: Moral Motives and Group-Based Moralities. *Personality and Social Psychology Review*, *17*(3), 219–236.

Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive Versus Prescriptive Morality: Two Faces of Moral Regulation. *Journal of Personality and Social Psychology*, *3*(521-537).

Johnson, B. T., & Eagly, A. H. (1989). Effects of Involvement on Persuasion: A Meta-Analysis. *Psychological Bulletin*, *106*(2), 290–314.

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.

Kinder, D. R. (1998). Opinion and Action in the Realm of Politics. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (pp. 778–867). New York, NY: McGraw-Hill.

Kinder, D. R., & Sanders, L. M. (1996). *Divided By Color: Racial Politics and Democratic Ideals*. Chicago, IL: Chicago University Press.

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.

Kohn, A. (1990). *The Brighter Side of Human Nature*. New York, NY: Basic Books.

Koleva, S. P., Graham, J., Iyer, R., Ditto, P. H., & Haidt, J. (2012). Tracing the Threads: How Five Moral Concerns (Especially Purity) Help Explain Culture War Attitudes. *Journal of Research in Personality*, *46*, 184–194.

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

Kurzban, R. (2010). *Why Everyone (Else) Is a Hypocrite: Evolution and the Modular mind*. Princeton, NJ: Princeton University Press.

Kurzban, R., Dukes, A., & Weeden, J. (2010). Sex, Drugs and Moral Goals: Reproductive Strategies and Views About Recreational Drugs. *Proceedings of the Royal Society B*, *277*, 3501–3508.

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.

Lau, R., Coulam, R., & Sears, D. O. (1983). *Proposition 2.5 in Massachusetts: Self-Interest, Anti-Government Attitudes, and Political Schemas*.

Lau, R. R., & Heldman, C. (2009). Self-Interest, Symbolic Attitudes, and Support for Public Policy: A Multilevel Analysis. *Political Psychology*, *30*, 513–537.

Lau, R., & Sears, D. O. (1981). Cognitive Links Between Economic Grievances and Political Responses. *Political Behavior*, *3*(279-302).

Lau, R., Sears, D. O., & Jessor, T. (1990). Fact or Artifact Revisited: Survey Instrument Effects and Pocketbook Politics. *Political Behavior*, *12*, 217–242.

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

Lerner, M. J. (1980). *The Belief in a Just World*. New York, NY: Plenum Press.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All Frames Are Not Created Equal. *Organizational Behavior and Human Decision Processes*, *76*, 149–188.

Lewis-Beck, M. S., Jacoby, W. G., Norpoth, H., & Weisberg, H. F. (2008). *The American Voter Revisited*. Ann Arbor, MI: University of Michigan 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.

Mansbridge, J. (1990a). On the Relation of Altruism and Self-Interest. In *Beyond Self-Interest*. Chicago, IL: Chicago University Press.

Mansbridge, J. (1990b). Preface. In *Beyond Self-Interest* (pp. ix–3). Chicago, IL: The University of Chicago Press.

Mansbridge, J. (1990c). The Rise and Fall of Self-Interest in the Explanation of Political Life. In *Beyond Self-Interest* (pp. 3–25). Chicago, IL: The University of Chicago Press.

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.

Miller, D. T. (1999). The Norm of Self-Interest. *American Psychologist*, *54*(12), 1053–1060.

Miller, D. T., & Ratner, R. K. (1996). The Power of the Myth of Self-Interest. In *Current Societal Concerns about Justice*. New York, NY: Plenum Press.

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.

Mooney, C. Z., & Lee, M.-H. (1995). Legislative Morality in the American States: The Case of pre-Roe Abortion Regulation Reform. *American Journal of Political Science*, *39*(3), 599–627.

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.

Mullen, E., & Skitka, L. J. (2006). Exploring the Psychological Underpinnings of Moral Mandate Effect: Motivated Reasoning, Identification, or Affect? *Journal of Personality and Social Psychology*, *90*, 629–643.

Nagel, T. (1970). *The Possibility of Altruism*. Oxford, UK: Clarendon.

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.

Nelson, T. E., & Kinder, D. R. (1996). Issue frames and group-centrism in American public opinion. *Journal of Politics*, *58*, 1055–1078.

Nteta, T. (2013). United We Stand? African Americans, Self-Interest, and Immigration Reform. *American Politics Research*, *41*, 147–172.

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.

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.

Owens, L. A., & Pedulla, D. S. (2014). Material Welfare and Changing Political Preferences: The Case of Support for Redistributive Social Policies. *Social Forces*, *92*, 1087–1113.

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

Petersen, M. B. (2016). Evolutionary political psychology. In D. Buss (Ed.), *The Handbook of Evolutionary Psychology* (2nd ed., Vol. 2). Hoboken, NJ: Wiley-Interscience.

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.

Quattrone, G., & Tversky, A. (1988). Contrasting Rational and Psychological Analyses of Political Choice. *American Political Science Review*, *82*, 719–736.

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.

Rehm, P., Hacker, J. S., & Schlesinger, M. (2012). Insecure Alliances: Risk, Inequality, and Support for the Welfare State. *American Political Science Review*, *106*(2), 386–406.

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

Riker, W., & Ordeshook, P. (1968). A Theory of the Calculus of Voting. *American Political Science Review*, *62*(1), 25–42.

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 non-randomized 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.

Schein, C., & Gray, K. (2017). The Theory of Dyadic Morality: Reinventing Moral Judgment by Redefining Harm. *Personality and Social Psychology Review*, *22*(1), 32–70.

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.

Schwartz, B. (1986). *The Battle for Human Nature*. New York, NY: W. W. Norton.

Sears, D. O., & Citrin, J. (1985). *Tax Revolt: Something for Nothing in California*. Cambridge, MA: Harvard University Press.

Sears, D. O., & Funk, C. L. (1990). Self-Interest in Americans' Political Opinions. In *Beyond Self-Interest* (pp. 147–171). Chicago, IL: The University of Chicago Press.

Sears, D. O., & Funk, C. L. (1991). The Role of Self-Interest in Social and Political Attitudes. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology, Volume 24*. New York, NY: Academic Press.

Sears, D. O., & Lau, R. (1983). Inducing Apparently Self-Interested Political Preferences. *American Journal of Political Science*, *74*, 670–684.

Sears, D. O., Lau, R. R., Tyler, T. R., & Allen, H. M. (1980). Self-Interest vs. Symbolic Politics in Policy Attitudes and Presidential Voting. *American Political Science Review*, *74*(3), 670–684.

Sen, A. K. (1977). Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philosophy and Public Affairs*, *6*, 317–344.

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

Skitka, L. J. (2002). Do the Means Always Justify the Ends or Do the Ends Sometimes Justify the Means? A Value Protection Model of Justice Reasoning. *Personality and Social Psychology Bulletin*, *28*, 588–597.

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

Skitka, L. J., Bauman, C. W., & Mullen, E. (2008). Morality and Justice: An Expanded Theoretical Perspective and Empirical Review. In K. A. Hedgvedt & J. Clay-Warner (Eds.), *Advances in Group Processes* (Vol. 25, pp. 1–27). New York, NY: Emerald Publishing.

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.

Stoker, L. (1992). Interest and Ethics in Politics. *American Political Science Review*, *86*(2), 369–380.

Studlar, D. T. (2001). What Constitutes Morality Policy? A Cross-National Analysis. In C. Z. Mooney (Ed.), *The Public Clash of Private Values: The Politics of Morality Policy* (pp. 37–54). New York, NY: Chatham House Publishers.

Suhay, E. (2008). *Group Influence and American Ideals: How Social Identity and Emotion Shape Our Political Values and Attitudes* (PhD thesis). American University.

Sznycer, D., Seal, M. F., Sell, A., Lim, J., Porat, R., Shalvi, S., … Tooby, J. (2017). Support for Redistribution Is Shaped by Compassion, Envy, and Self-Interest, but Not a Taste for Fairness. *Proceedings of the National Academy of Sciences*, *114*(31), 8420–8425.

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.

Tavits, M. (2007). Principle vs. Pragmatism: Policy Shifts and Political Competition. *American Journal of Political Science*, *51*(1), 151–165.

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.

Tetlock, P. E., Peterson, R. S., & Lerner, J. S. (1996). Revising the Value Pluralism Model: Incorporating Social Content and Context Postulates. In C. Seligman, J. M. Olson, & M. P. Zanna (Eds.), *The Psychology of Values: The Ontario Symposium* (Vol. 8, pp. 25–51). Mahwah, NJ: Lawrence Erlbaum.

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

Tyler, T. R. (1990a). Justice, Self-Interest, and the Legitimacy of Legal and Political Authority. In *Beyond Self-Interest* (pp. 171–183). Chicago, IL: The University of Chicago Press.

Tyler, T. R. (1990b). *Why People Obey the Law*. Chicago, IL: Chicago University Press.

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.

Weatherford, S. M. (1983). Economic Voting and the 'Symbolic Politics' Argument: A Reinterpretation and Synthesis. *American Political Science Review*, *77*(158-174).

Weeden, J. (2015). Losing My Religion: An Analysis of the Decline in Religious Attendance from Childhood to Adulthood. In D. J. Slone & J. A. V. Slyke (Eds.), *The Attraction of Religion: A New Evolutionary Psychology of Religion* (pp. 73–91). London, UK: Bloomsbury Academic.

Weeden, J., & Kurzban, R. (2014). *The Hidden Agenda of the Political Mind: How*

Self-Interest Shapes Our Opinions and Why We Won't Admit It. Princeton, NJ: Princeton University Press.

Weeden, J., & Kurzban, R. (2016). Do People Naturally Cluster into Liberals and Conservatives? *Evolutionary Psychological Science*, *2*, 47–57.

Weeden, J., & Kurzban, R. (2017). Self-Interest Is Often A Major Determinant of Issue Attitudes. *Advances in Political Psychology*, *38*(1), 67–90.

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.

Wolsko, C., Ariceage, H., & Seiden, J. (2016). Red, White, and Blue Enough to Be Greeen: Effects of Moral Framing on Climate Change Attitudes and Conservation Behaviors. *Journal of Experimental Social Psychology*, *65*, 7–19.

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.