WHAT THE PEOPLE THINK: ADVANCES IN PUBLIC OPINION MEASUREMENT USING MODERN STATISTICAL METHODS

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 MODERN STATISTICAL METHODS

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 MODERN STATISTICAL METHODS

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

Table Page

LIST OF FIGURES

TABLE OF CONTENTS

APPENDIX

CHAPTER 1

INTRODUCTION

This is the introduction.

CHAPTER 2

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

2.1 Introduction

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

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

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

2.2 Theory

2.2.1 Preliminary Notations on Survey Experiments

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

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

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

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

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

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

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

2.2.2 Randomization

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

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

2.2.3 Blocking

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

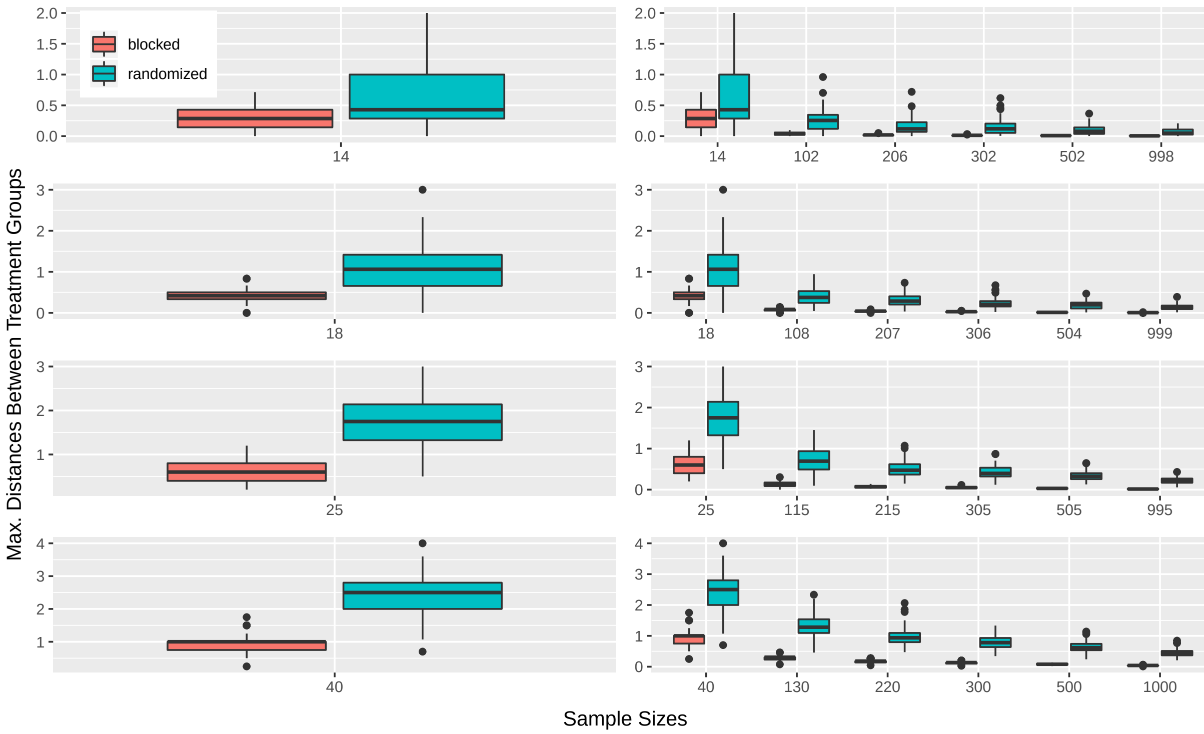

Figure 2.1. Distances between treatment group means in randomized and blocked data. Increasing sample size for 2 (top row), 3 (second row), 5 (third row), and 10 treatment groups (bottom row). Leftmost pair on right panel is exactly the pair on the left panel

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

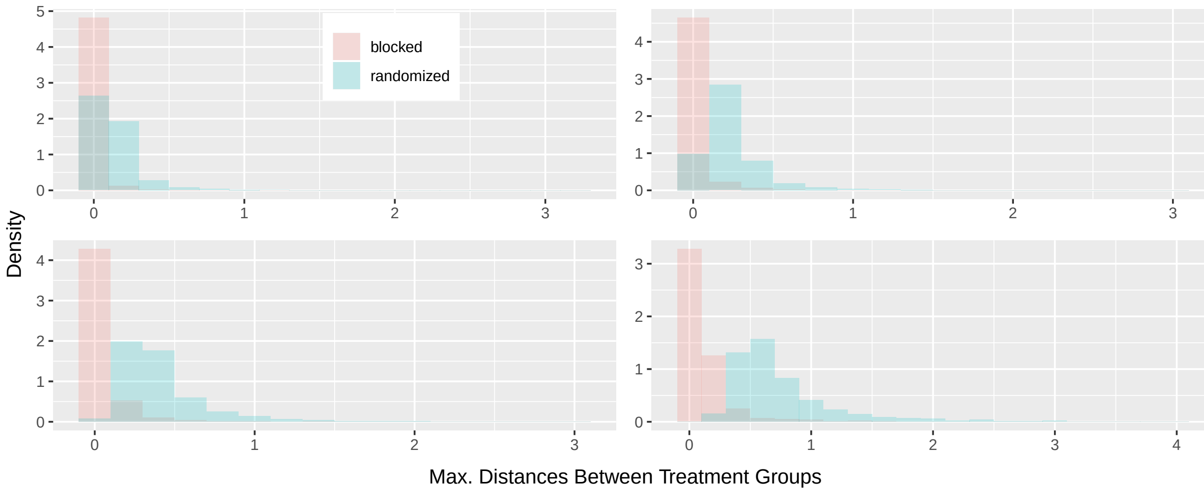

Figure 2.2. Distribution of treatment group differences in randomized and blocked data for 2 (top left), 3 (top right), 5 (bottom left), and 10 (bottom right) treatment groups

Blocking On The Go

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

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

Blocking is thus possible when all covariate information is known at the time of assignment and when this information 'trickles in' over time. Covariate information, however, is only one side of the coin. Researchers also need to take into consideration the characteristics of the variable to block on. Not all types of variables can and should be used the same way to be blocked on. Specifically, the current use of ordinal variables as blocking variables is somewhat problematic.

2.2.4 Ordinal Variables

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 blocking, the categories of nominal variables are often turned into binary variables. This manipulation does not impose any unnatural ordering onto the variable and thus does not require any theoretical assumptions. Interval variables are often made numeric, which is statistically sound. It makes sense to assign numeric values such as 1, 2, 3, and 4 to `income` categories of $20,000, $40,000, $60,000, and $80,000. The distance between each of these categories is identical between any adjacent pair and thus translates perfectly into the numeric values with equally identical distances. The distance between \$20,000 and \$40,000 is the same as the distance between 1 and 2. Ordinal variables are also often made numeric for blocking. 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.

2.2.5 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 a predictive variable.

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

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

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

To use this ordered probit model for blocking, 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 block on the resulting categories.

2.3 Data

ANES Model Training

One of the most common ordinal variables in political science 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). 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

Table 2.1. Ordered Probit Threshold Estimates

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

model uses the ordinal information with unevenly spaced distances provided and returns categories that do fit the data. We can now use these estimated education categories as the basis for blocking. Assigning numeric values to the new categories is now justifiable because they are based on data-driven estimations. This allows us to block on numerical values with the Mahalanobis distance, which would not be possible without empirical justification. The following sections show that the new estimated categories significantly affect analyses and results.

2.3.1 Simulations

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

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

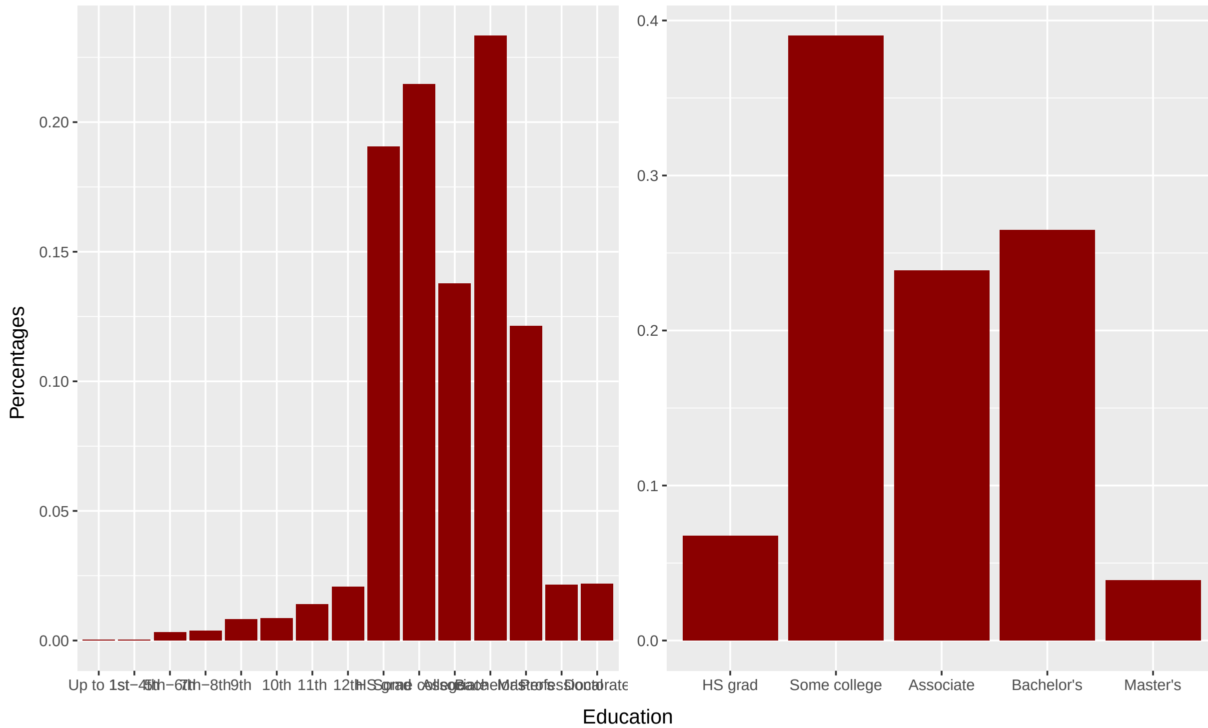

Figure 2.3. Distribution of Education Categories. Original 2016 ANES categories on the left, ordered probit estimated categories on the right

on an interval response variable, a feeling thermometer towards Donald Trump as the Republican presidential candidate:

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

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

2.3.2 Framing Survey

2.4 Results

2.4.1 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.

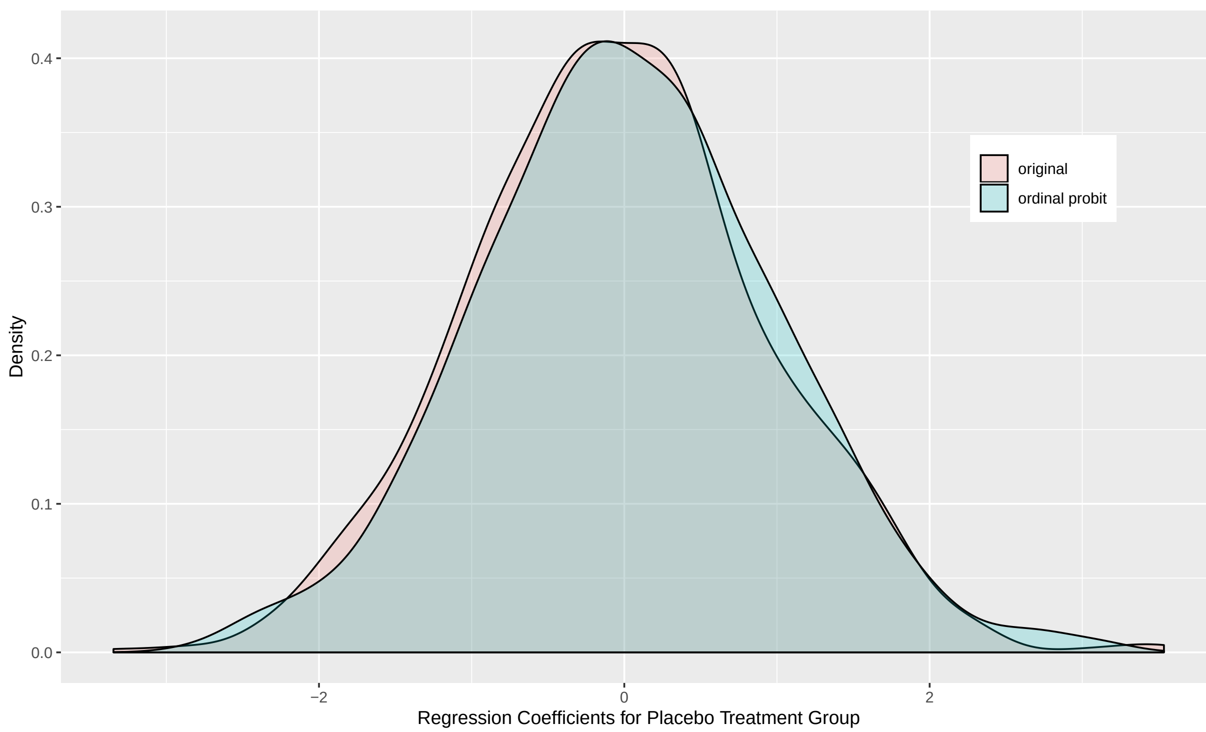

Figure 2.4. Distribution of placebo treatment coefficients by education model

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

2.4.2 Framing Survey

Put text here

CHAPTER 3

PRECISION IN SURVEY MEASUREMENT – A NEW METHOD TO IMPUTE MISSING DATA FROM 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). Missing data pose 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 general ways to treat missing data. These range from deleting all observations with missing data (listwise deletion) over randomly drawing a 'similar' respondent to provide a fill-in value for a missing slot (hot decking) to estimating missing values from conditional distributions (multiple imputation) (Fay, 1996; King, Honaker, Joseph, & Scheve, 2001; Rubin, 1976). Listwise deletion has been shown to induce bias with political data, and hot decking does not reflect statistical uncertainty in the filled-in values since there is only one draw (Gill & Witko, 2013; Kroh, 2006; Rees & Duke-Williams, 1997). While multiple implementation has become and remains the state of the art in missing data management, it is not necessarily always suitable for all types of variables. Multiple hot deck imputation, an improvement over generic multiple imputation, solves this for non-granular discrete data (Gill & Cranmer, 2012; Reilly, 1993). However, the underlying algorithm assumes even distances between categories in discrete data, which makes it unsuitable for ordinal variables. I propose a method designed to impute missing data specifically from ordinal variables that fills this gap in multiple hot deck imputation.

Multiple hot deck imputation uses draws of values from the variable with the missing values (hot decking) to impute them distributionally (multiple imputation) and estimate affinity scores. This score measures how close other respondents are to the one with the missing value. 'Closeness' is measured as the distance between respondents in the variables that do not contain missing values. This is best illustrated with simplified data shown in Table 3.1. Respondent B shows missing data for party ID. To impute a fill-in

Respondent	Age	Party ID	Education	Income	Gender
A	25	Republican	High School Graduate	$40-50,000	Male
B	40	NA	Some High School	$30-40,000	Female
C	30	Democrat	Bachelor's Degree	$60-70,000	Female

Table 3.1. Illustrative Data

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. Multiple hot deck imputation measures these distances and estimates affinity scores for respondents A and C. B then receives the party ID fill-in value from whichever respondent has the higher score. The algorithm building the affinity score, however, assumes evenly spaced distances between categories. This is the case for age, income, and gender, but not for education, since education is an ordinal variable. Applying multiple hot deck imputation here would misrepresent the data.

Instead, I propose a weighted distance solution with the estimated numeric thresholds from the ordered probit model approach in chapter II to measure the distances between the categories and calculate the affinity score. I demonstrate the benefits of this method with several Monte Carlo simulations. As in chapter II, simulations are crucial as they allow comparison to the 'true' results, which is not possible with actual data. They show that weighted distance multiple hot decking imputation outperforms current general missing data techniques for ordinal variables.

3.2 Theory

3.2.1 Missing Data Mechanisms

MCAR MAR NMAR

3.2.2 Deletion

Deletion of incomplete observations, listwise deletion

3.2.3 Imputation

Mean

Mean is used to substitute values

Regression

Missing variables for a unit are estimated by predicted values from the regression on the known variables for that unit

Hot Decking

Recorded units in the sample are used to substitute values

Hot decking (Marker, Juddkins, Winglee (2002), Ernst (198), Kalton and Kish (1981), Ford (1983), David et al. (1986)) (chapter 4, (Little & Rubin, 2002)) – by simple random sampling with replacement – within adjustment cells – nearest neighbor – sequential ordered by a covariate

3.2.4 Multiple Imputation

`Amelia`

`mice`

Multiple Hot Decking

Multiple hot deck imputation (all COPIED OVER from Cranmer, Gill): – A nonparametric alternative to multiple imputation – A variation of hot deck imputation combined with the repeated imputation and estimation method typical of parametric multiple imputation – Designed to work well in situations where (traditional) parametric multiple imputation falls short – when a discrete variable with a small number of categories has missing values. This can produce nonsensical imputations, biased results and artificially smaller standard errors – Maintains the integrity of the data by using draws of actual values from the variable with the missing values to impute missing items – Maintains the discrete nature of discrete data and produces more accurate imputations than parametric multiple imputation in a majority of social science applications where the data are discrete – Since many political science applications rely on highly discrete measures, multiple hot deck imputation provides the researcher with more accuracy in imputations than parametric multiple imputation, while requiring none of parametric multiple imputation's standard assumptions

Concrete estimation steps: (1) Create several copies of the dataset. (2) Search down columns of the data sequentially looking for missing observations a) When a missing value is found, compute a vector of affinity scores, for that missing value. This vector is as long as the number of rows in the dataset, minus any rows with missing values for the same variable b) Create the imputation cell of best donors for this missing value and draw randomly from it to produce a vector of imputations c) Impute one of these values into the appropriate cell of each duplicate dataset for this missing value (3) Repeat Step 2 until no missing observations remain (4) Fit the statistic of interest for each dataset (5) Combine the estimates of the statistic into a single estimate using the combination rules of parametric multiple imputation

Ordinal Variable Multiple Hot Decking

Imputation techniques rely on continuous distributional assumptions. I use my ordered probit estimated latent underlying continuous variable.

Ordinal hot deck imputation: – An extension of multiple hot deck imputation – Designed specifically to implement multiple hot deck imputation with ordinal variables – Fully utilizes the unevenly spaced yet ordered information provided in ordinal variables – A key variable in political science surveys is ordinal: Education. Ordinal hot deck imputation enables researchers to impute missing data using the full information provided in this most important predictor variable

3.2.5 Amputation

`ampute()` from `mice` and my own functions. Explain all here.

3.3 Data

3.4 Results

3.5 Accuracy of Imputation Methods

We test the accuracy of five multiple imputation methods (`hd.ord`, `hd.norm`, `amelia`, `mice`, `na.omit`) on several data sets (old framing experiment, ANES, framing survey). `hd.ord` is my ordinal multiple hot decking function. `hd.norm` is Gill and Cranmer's multiple hot decking function. `amelia` and `mice` are used in its default settings unless specified otherwise. The 'old framing experiment' consists of data from a framing experiment I ran on MTurk on 2017. ANES refers to the 2016 ANES data. The framing survey data will come from the eventual framing experiment I will run with Lucid (and is thus obviously not included in this write-up right now). NAs are inserted with three amputation methods (`ampute`, `own.NA`, `own.NA.rows`) and some variations of the options within these methods. `ampute` is included in the `Amelia` package. `own.NA` and and `own.NA.rows` are functions I have written, the details of which will be explained below.

The tables below provide an overview of the accuracy of each imputation method. All difference values are absolute values, since we care about the closeness of each method to the truth, regardless of whether closeness has a positive or negative value. The sign thus does not matter here. The analyses use differing data sets, differing missing data mechanisms, differing amputation functions, and differing settings for the respective amputation functions.

Table 3.2 uses `ampute`, the old framing data (n = 1,003), and inserts NAs MAR for 5 variables at 20, 50, and 80 percent for 12,500, 12,462, and 10,566 iterations.

Table 3.3 uses `ampute`, the 2016 ANES data (n = 3,223), and inserts NAs MAR for 5 variables at 20 percent for 2,396 iterations.

Table 3.4 uses `ampute` with the options `bycases=FALSE` and `cont=FALSE`, the old framing data (n = 1,003), and inserts NAs MAR for 5 variables at 20 percent for 9,644 iterations.

Table 3.5 uses `ampute` and inserts NAs MNAR for 5 variables at 20 percent. This is done for the old framing data (n = 1,003) for 12,499 iterations and for the 2016 ANES data (n = 3,223) for 2,351 iterations.

Table 3.6 uses my own function `own.NA`, the old framing data (n = 1,003), and inserts NAs MAR for 3 variables at 20 percent for 12,324 iterations.

Table 3.7 uses my own function `own.NA.rows`, the old framing (n = 1,003) data, and inserts NAs MAR for 17 variables at 20 percent for 10,000 iterations.

The number of iterations differs due to the percentage of missingness, the amputation function used, the `ampute` options, the missingness mechanism, and the levels of `education` in each data set. `education` is crucial to these analyes as the point of conducting them is to test the performance of `hd.ord`; a multiple hot decking function designed specifically for ordinal variables. `hd.ord` in turn depends on another specifically designed function called `OPMOrd`. `OPMOrd` applies `polr()` to an ordinal variable (here `education`) to estimate the underlying latent continuous variable. `polr()` can only be run on data without missing variables, which means the amputation process sometimes leads to fewer levels of `education` levels in some amputation iterations. Since a comparison of data with differing variable levels makes little sense, these iterations are discarded. That in turn leads to differing numbers of iterations across analyses.

3.5.1 With Standard `ampute()`

Table 3.2 uses ampute, the old framing data (n = 1,003), and inserts NAs at 20, 50, and 80 percent MAR. `amelia` performs best, followed by `mice`. `hd.ord` outperforms `hd.norm`. `Amelia` is better than `mice` for `inc`, `age`, and `interest` (except for 20 percent NAs for the latter). For `Dem`, `Female`, and `White`, `Amelia` is also ahead, though `mice` sometimes edges it by one unit or so in the fourth decimal. The differences between the

Table 3.2. Accuracy of Multiple Imputation Methods (ampute, old framing data (n = 1,003))

Method	Variable	NAs					
		20 Percent		50 Percent		80 Percent	
		Value	Diff	Value	Diff	Value	Diff
true	Dem	.4666	0	.4666	0	.4666	0
hd.ord	Dem	.4667	.0001	.4668	.0002	.4675	.0009
hd.norm	Dem	.4671	.0005	.4678	.0012	.4686	.0020
amelia	Dem	.4665	.0001	.4665	.0001	.4667	.0001
mice	Dem	.4665	.0001	.4666	0	.4669	.0003
na.omit	Dem	.4339	.0327	.3898	.0768	.3438	.1228
true	inc	3.0927	0	3.0927	0	3.0927	0
hd.ord	inc	3.0864	.0063	3.0703	.0224	3.0531	.0396
hd.norm	inc	3.0798	.0129	3.0524	.0403	3.0319	.0608
amelia	inc	3.0925	.0002	3.0928	.0001	3.0933	.0006
mice	inc	3.0937	.0010	3.0949	.0022	3.0956	.0029
na.omit	inc	2.9773	.1154	2.8260	.2667	2.6674	.4253
true	age	37.9252	0	37.9252	0	37.9252	0
hd.ord	age	37.6382	.2870	37.0970	.8282	36.6639	1.2613
hd.norm	age	37.5650	.3602	36.9376	.9876	36.5279	1.3973
amelia	age	37.9238	.0014	37.9234	.0018	37.9278	.0026
mice	age	37.9281	.0029	37.9305	.0053	37.9345	.0093
na.omit	age	36.6743	1.2509	35.0811	2.8441	33.5601	4.3651
true	White	.7717	0	.7717	0	.7717	0
hd.ord	White	.7738	.0021	.7772	.0055	.7815	.0098
hd.norm	White	.7742	.0025	.7778	.0061	.7821	.0104
amelia	White	.7717	0	.7717	0	.7716	.0001
mice	White	.7716	.0001	.7712	.0005	.7709	.0008
na.omit	White	.7453	.0264	.6980	.0737	.6321	.1396
true	Female	.4666	0	.4666	0	.4666	0
hd.ord	Female	.4662	.0004	.4650	.0016	.4641	.0025
hd.norm	Female	.4658	.0008	.4638	.0028	.4624	.0042
amelia	Female	.4666	0	.4666	0	.4666	0
mice	Female	.4667	.0001	.4668	.0002	.4669	.0003
na.omit	Female	.4293	.0373	.3774	.0892	.3231	.1435
true	interest	3.2164	0	3.2164	0	3.2164	0
hd.ord	interest	3.1990	.0174	3.1712	.0452	3.1462	.0702
hd.norm	interest	3.1970	.0194	3.1623	.0541	3.1353	.0811
amelia	interest	3.2167	.0003	3.2165	.0001	3.2162	.0002
mice	interest	3.2164	0	3.2162	.0002	3.2158	.0006
na.omit	interest	3.1490	.0674	3.0386	.1778	2.8969	.3195

methods are less pronounced for the binary variables and most visible in the nominal ones. It is worst for `age`, which has the most unique values. As the percentage of NAs increases, the estimates are further off from the true means. This is to be expected for all methods, but it affects `hd.ord` and `hd.norm` more than `amelia` and `mice`.

These results are overall confirmed when `ampute()` is run on the 2016 ANES data (n = 3,223), as shown in Table 3.3. It is noticeable, however, that the differences in method performance are somewhat reduced with the ANES data (n = 3,223). This is likely due to the increased number of observations in the data. For reasons of brevity, this analysis and all further analyses insert 20 percent NAs only.

Table 3.3. Accuracy of Multiple Imputation Methods (ampute, ANES 2016 (n = 3,223))

Method	Variable	Value	Diff
true	Dem	.3549	.0000
hd.ord	Dem	.3552	.0003
hd.norm	Dem	.3552	.0003
amelia	Dem	.3549	.0000
mice	Dem	.3549	.0000
na.omit	Dem	.3388	.0161
true	inc	15.5740	.0000
hd.ord	inc	15.5762	.0022
hd.norm	inc	15.5779	.0039
amelia	inc	15.5753	.0013
mice	inc	15.5755	.0015
na.omit	inc	15.2306	.3434
true	age	49.0400	.0000
hd.ord	age	49.0075	.0325
hd.norm	age	49.0157	.0243
amelia	age	49.0384	.0016
mice	age	49.0385	.0015
na.omit	age	48.3780	.6620
true	Female	.4682	.0000
hd.ord	Female	.4684	.0002
hd.norm	Female	.4684	.0002
amelia	Female	.4683	.0001
mice	Female	.4682	.0000
na.omit	Female	.4470	.0212
true	interest	2.2197	.0000
hd.ord	interest	2.2213	.0016
hd.norm	interest	2.2213	.0016
amelia	interest	2.2198	.0001
mice	interest	2.2198	.0001
na.omit	interest	2.2009	.0188

3.5.2 With Modified `ampute()`

Table 3.4 shows the results of using `ampute()` with the options `bycases=FALSE` and `cont=FALSE` on the old framing data (n = 1,003). `hd.ord` overall performs worse than the analyses presented in Tables 3.2 and 3.3 above.

Table 3.4. Accuracy of Multiple Imputation Methods (ampute with bycases, old framing data (n = 1,003))

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

Table 3.5 shows the results of using `ampute()` with MNAR as the missing data mechanism on the old framing (n = 1,003) and the 2016 ANES data (n = 3,223). The results are in line with those presented in Tables 3.2 and 3.3. It is noteworthy, however, that `na.omit` is overall much closer to the other methods. It actually performs better than `hd.norm` for `age` (.7236 vs. .7499) in the old framing data (n = 1,003) and is the best method for `interest` for the ANES data (n = 3,223) (.0159).

Table 3.5. Accuracy of Multiple Imputation Methods (ampute with MNAR)

Method	Variable	Data			
		Old Framing (n = 1,003)		ANES (n = 3,223)	
		Value	Diff	Value	Diff
true	Dem	.4666	0	.3549	0
hd.ord	Dem	.4630	.0036	.3527	.0022
hd.norm	Dem	.4629	.0037	.3527	.0022
amelia	Dem	.4619	.0047	.3526	.0023
mice	Dem	.4629	.0037	.3526	.0023
na.omit	Dem	.4456	.0210	.3395	.0154
true	inc	3.0927	0	15.5740	0
hd.ord	inc	3.0357	.0570	15.3948	.1792
hd.norm	inc	3.0298	.0629	15.3922	.1818
amelia	inc	3.0469	.0458	15.4136	.1604
mice	inc	3.0486	.0441	15.4137	.1603
na.omit	inc	3.0210	.0717	15.3494	.2246
true	age	37.9252	0	49.0400	0
hd.ord	age	37.2342	.6910	48.7292	.3108
hd.norm	age	37.1753	.7499	48.7378	.3022
amelia	age	37.6363	.2889	48.7809	.2591
mice	age	37.6411	.2841	48.7818	.2582
na.omit	age	37.2016	.7236	48.5130	.5270
true	Female	.4666	0	.4682	0
hd.ord	Female	.4520	.0146	.4539	.0143
hd.norm	Female	.4520	.0146	.4539	.0143
amelia	Female	.4524	.0142	.4541	.0141
mice	Female	.4526	.0140	.4541	.0141
na.omit	Female	.4445	.0221	.4521	.0161
true	interest	3.2164	0	2.2197	0
hd.ord	interest	3.1821	.0343	2.1956	.0241
hd.norm	interest	3.1797	.0367	2.1954	.0243
amelia	interest	3.2028	.0136	2.1965	.0232
mice	interest	3.2025	.0139	2.1965	.0232
na.omit	interest	3.1768	.0396	2.2038	.0159

3.5.3 With My Own Amputation Methods

Table 3.6 shows the results of using my own function to insert NAs MAR into the old framing data (n = 1,003), `own.NA()`. 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 3.6 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.

Table 3.6. Accuracy of Multiple Imputation Methods (own.NA, old framing data (n = 1,003))

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

`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 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 3.7 shows the results, with the missingess percentage set to 20.

Table 3.7: *Accuracy of Multiple Imputation Methods (own.NA.rows, old framing data (n = 1,003))*

Method	Variable	Value	Diff
true	Dem	.4666	0
hd.ord	Dem	.4666	0
hd.norm	Dem	.4667	.0001
amelia	Dem	.4667	.0001
mice	Dem	.4670	.0004
na.omit	Dem	.4667	.0001
true	Ind	.2802	0
hd.ord	Ind	.2795	.0007
hd.norm	Ind	.2801	.0001
amelia	Ind	.2801	.0001
mice	Ind	.2817	.0015
na.omit	Ind	.2801	.0001
true	Cons	.2832	0
hd.ord	Cons	.2830	.0002
hd.norm	Cons	.2831	.0001
amelia	Cons	.2831	.0001
mice	Cons	.2823	.0009
na.omit	Cons	.2831	.0001
true	Lib	.5174	0
hd.ord	Lib	.5182	.0008
hd.norm	Lib	.5176	.0002
amelia	Lib	.5176	.0002
mice	Lib	.5173	.0001
na.omit	Lib	.5176	.0002
true	Black	.0698	0
hd.ord	Black	.0702	.0004
hd.norm	Black	.0698	0
amelia	Black	.0698	0

mice	Black	.0731	.0033
na.omit	Black	.0698	0
true	Hisp	.0548	0
hd.ord	Hisp	.0546	.0002
hd.norm	Hisp	.0548	0
amelia	Hisp	.0548	0
mice	Hisp	.0589	.0041
na.omit	Hisp	.0548	0
true	White	.7717	0
hd.ord	White	.7712	.0005
hd.norm	White	.7717	0
amelia	White	.7717	0
mice	White	.7613	.0104
na.omit	White	.7717	0
true	Asian	.0808	0
hd.ord	Asian	.0812	.0004
hd.norm	Asian	.0808	0
amelia	Asian	.0809	.0001
mice	Asian	.0835	.0027
na.omit	Asian	.0808	0
true	Female	.4666	0
hd.ord	Female	.4665	.0001
hd.norm	Female	.4665	.0001
amelia	Female	.4665	.0001
mice	Female	.4673	.0007
na.omit	Female	.4665	.0001
true	Unempl	.1615	0
hd.ord	Unempl	.1610	.0005
hd.norm	Unempl	.1616	.0001
amelia	Unempl	.1616	.0001
mice	Unempl	.1624	.0009
na.omit	Unempl	.1616	.0001
true	Ret	.0508	0
hd.ord	Ret	.0506	.0002
hd.norm	Ret	.0508	0
amelia	Ret	.0509	.0001
mice	Ret	.0505	.0003
na.omit	Ret	.0509	.0001
true	Stud	.0439	0
hd.ord	Stud	.0446	.0007
hd.norm	Stud	.0439	0
amelia	Stud	.0439	0
mice	Stud	.0466	.0027

na.omit	Stud	.0439	0
true	interest	3.2164	0
hd.ord	interest	3.2161	.0003
hd.norm	interest	3.2163	.0001
amelia	interest	3.2164	0
mice	interest	3.2122	.0042
na.omit	interest	3.2163	.0001
true	media	1.7268	0
hd.ord	media	1.7294	.0026
hd.norm	media	1.7270	.0002
amelia	media	1.7270	.0002
mice	media	1.7259	.0009
na.omit	media	1.7269	.0001
true	part	.9561	0
hd.ord	part	.9575	.0014
hd.norm	part	.9560	.0001
amelia	part	.9561	0
mice	part	.9546	.0015
na.omit	part	.9561	0
true	inc	3.0927	0
hd.ord	inc	3.0906	.0021
hd.norm	inc	3.0926	.0001
amelia	inc	3.0926	.0001
mice	inc	3.0949	.0022
na.omit	inc	3.0926	.0001
true	age	37.9252	0
hd.ord	age	37.9000	.0252
hd.norm	age	37.9250	.0002
amelia	age	37.9243	.0009
mice	age	37.8789	.0463
na.omit	age	37.9250	.0002

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

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

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

It is also noteable how often the difference amounts to zero and how well `na.omit` performs overall.

3.5.4 Overall Table Analysis So Far

This is the summary of all analyses conducted so far, except for the ones with my own functions `own.NA()` and `own.NA.rows()`. The results for these functions showed such a strong performance of `na.omit` that the functions' design needs to be questioned and reworked. `na.omit` should not perform so strongly for a missingness mechanism that is supposed to be MAR, so the functions likely represent a version of MCAR. This puts the usefulness of their imputation results in doubt.

For all the other analyses, I am setting a 'threshold' of the fourth decimal to determine 'closeness'. This means the performance of `hd.ord` is only considered close to the best-performing method (`amelia` or `mice`) when the difference between the best-performing method and `hd.ord` is no higher than a value in the fourth decimal.

For `Dem`, `hd.ord` is close for all combinations: Old framing (n = 1,003) ampute standard (identical values), ANES (n = 3,223) MNAR (.0001), framing (n = 1,003) MNAR (.0001), ANES (n = 3,223) ampute standard (.0003), and framing (n = 1,003) bycases (.0007).

For `Female`, `hd.ord` is close for ANES (n = 3,223) MNAR (.0002), ANES (n = 3,223) ampute standard (.0002), framing (n = 1,003) ampute standard (.0002), and framing (n = 1,003) MNAR (.0006). It is not close for framing (n = 1,003) bycases (.0047).

For `inc`, `hd.ord` is close for ANES (n = 3,223) ampute standard (.0009). It is not close framing (n = 1,003) ampute standard (.0061), framing (n = 1,003) MNAR (.0129), and ANES (n = 3,223) MNAR (.0189). It is by far the worst for framing (n = 1,003) bycases (.0447).

For `age`, `hd.ord` is not close for any combination. It is closest for ANES (n =

3,223) ampute standard (.0310) and ANES (n = 3,223) MNAR (.0526). Further away are framing (n = 1,003) ampute standard (.2856) and framing (n = 1,003) MNAR (.4069). It is by far the worst for framing (n = 1,003) bycases (1.4154).

For `interest`, `hd.ord` is close for ANES (n = 3,223) MNAR (.0009). It is not close for ANES (n = 3,223) ampute standard (.0015), framing (n = 1,003) ampute standard (.0174), and framing (n = 1,003) MNAR (.0207). It is by far the worst for framing (n = 1,003) bycases (.0788).

It thus appears that `hd.ord` is very close to `amelia` and `mice` for binary variables (`Dem`, `Female`) for both missing data mechanisms (MAR, MNAR). All combinations for `Dem` and all but one combination for `Female` (framing (n = 1,003) bycases, .0047) fall within the closeness threshold.

`hd.ord` performs worse for the nominal variables `age` and `inc` as well as the ordinal variable `interest`. Only two combinations overall (ANES (n = 3,223) ampute standard for `inc`, .0009; ANES (n = 3,223) MNAR for `interest`, .0009) fall within the closeness threshold. However, there is a slight tendency that `hd.ord` performs slightly better when applied to the ANES data (n = 3,223): The two closest `interest` values (ANES (n = 3,223) MNAR, .0009; ANES (n = 3,223) ampute standard, .0015), the closest `inc` value (ANES (n = 3,223) ampute standard, .0009), and the two closest `age` values (ANES (n = 3,223) ampute standard, .0310; ANES (n = 3,223) MNAR, .0526) all concern the ANES (n = 3,223) data. This slight tendency could be due to the higher number of observations and, in the case of `age` and `inc`, the higher number of unique values: `age` contains 73 and `inc` 27 unique values in the ANES (n = 3,223) data, compared to 58 and 7 in the old framing data (n = 1,003) (`interest` consists of the same number of unique values (4) in both data sets). In addition, the ANES (n = 3,223) data was run for fewer iterations. Nonetheless, the distances are still rather big in some cases (particularly for `age`).

`hd.ord` overall performs worse with the `ampute()` options `bycases=FALSE` and `cont=FALSE`, except for `Dem`.

Running `ampute()` with MNAR as the missing data mechanism does not yield systematically different results than its default setting of MAR.

With increased NA percentages, imputation methods are bound to get worse. `amelia` and `mice` get worse only slightly as the percentage increases. This impact is much more pronounced for `hd.ord` and `hd.norm`. This is to be expected, since the latter methods are based on hot decking with replacement. Missingness levels of 50 and 80 percent are extremely rare in practice, however, so these results do not carry high importance.

3.6 Runtimes

Run in R 3.6 on a Code Ocean AWS EC2 instance with 16 cores and 120 GB of memory. Amputed with `ampute`. As can be seen in Table 3.8, `hd.ord` and `hot.deck`

Table 3.8. 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

show virtually identical imputation times for the old framing data (n = 1,003), with `hd.ord` being 12 seconds faster. `amelia`, however, is 2.3 times slower than `hd.ord`. `mice` is 17.9 times slower than `hd.ord`. This is a dramatic speed gain. The ANES (n = 3,223) data show only a reduced speed gain. `mice` is still drastically slower than `hd.ord` (by a magnitude of 9), but the speed gain is cut in half. The previous speed gain over `amelia` is no longer observable. `hd.ord` and `amelia` now show virtually identical runtimes (with a difference of 25 seconds in favor of `hd.ord`).

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 3.9, 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 3.9. 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

3.6.1 What doesn't work

– Increasing percentage of NAs == worse `hd.ord` performance – Ordinal and nominal variables == worse `hd.ord` performance – High number of observations == reduced number of iterations (crashes and/or RAM maxing out) – High number of observations == makes `amelia` faster relative to `hd.ord` – 17 ANES education levels == increases needed number of iterations – 17 ANES education levels == causes `amelia` to stop on CO and Jeff – 10,000 iterations == results with 1,000 iterations are just as good – `ampute()` with `bycases=FALSE` and `cont=FALSE` == worse `hd.ord` performance, good `na.omit` performance – `own.NA()` == `na.omit` performs best – `own.NA.rows()`== pretty much everything is zero, incl. `na.omit`; `mice` is awful – Running `hd.ord` with `method = p.draw` == worse `hd.ord` performance – Running `hd.ord` with `method = p.draw` == only works with only binary vars in the data – Increasing `sdCutoff` == only does something with only binary vars in the data – Only binary vars in data == no gain in `hd.ord` performance

3.6.2 What works

– Results for binary variables == equal performance of `hd.ord`, `mice`, `amelia` – Increasing number of variables with NAs (all, not just binary) == better `hd.ord` performance – 1000 observations in data sets == increases `amelia` running time – MNAR == MAR in terms of `hd.ord` performance

Table 3.10: *CCES 1000 Observations, Increasing Variables*

Method	Variable	Data			
		7 vars		10 vars	
		Value	Diff	Value	Diff
true	Dem	.4320	.0000	.4320	.0000
hd.ord	Dem	.4330	.0010	.4326	.0006
hd.norm	Dem	.4332	.0012	.4326	.0006
amelia	Dem	.4321	.0001	.4320	.0000
mice	Dem	.4320	.0000	.4320	.0000
na.omit	Dem	.4085	.0235	.4163	.0157
true	Ind	.2820	.0000	.2820	.0000
hd.ord	Ind	.2823	.0003	.2821	.0001
hd.norm	Ind	.2821	.0001	.2821	.0001
amelia	Ind	.2820	.0000	.2820	.0000
mice	Ind	.2820	.0000	.2819	.0001
na.omit	Ind	.2716	.0104	.2719	.0101
true	Black	.1070	.0000	.1070	.0000
hd.ord	Black	.1074	.0004	.1072	.0002
hd.norm	Black	.1073	.0003	.1071	.0001
amelia	Black	.1070	.0000	.1070	.0000
mice	Black	.1068	.0002	.1070	.0000
na.omit	Black	.0837	.0233	.0926	.0144
true	Hisp	.0910	.0000	.0910	.0000
hd.ord	Hisp	.0913	.0003	.0911	.0001

hd.norm	Hisp	.0912	.0002	.0910	.0000
amelia	Hisp	.0910	.0000	.0910	.0000
mice	Hisp	.0909	.0001	.0910	.0000
na.omit	Hisp	.0667	.0243	.0756	.0154
true	Empl	.4260	.0000	.4260	.0000
hd.ord	Empl	.4274	.0014	.4263	.0003
hd.norm	Empl	.4276	.0016	.4265	.0005
amelia	Empl	.4262	.0002	.4261	.0001
mice	Empl	.4263	.0003	.4261	.0001
na.omit	Empl	.3956	.0304	.3958	.0302
true	Stud	.0490	.0000	.0490	.0000
hd.ord	Stud	.0492	.0002	.0493	.0003
hd.norm	Stud	.0491	.0001	.0491	.0001
amelia	Stud	.0491	.0001	.0489	.0001
mice	Stud	.0491	.0001	.0491	.0001
na.omit	Stud	.0343	.0147	.0448	.0042
true	Male	.4630	.0000	.4630	.0000
hd.ord	Male	.4631	.0001	.4630	.0000
hd.norm	Male	.4630	.0000	.4627	.0003
amelia	Male	.4631	.0001	.4629	.0001
mice	Male	.4632	.0002	.4630	.0000
na.omit	Male	.4328	.0302	.4273	.0357
true	inc	—	—	6.3840	.0000
hd.ord	inc	—	—	6.3795	.0045
hd.norm	inc	—	—	6.3723	.0117
amelia	inc	—	—	6.3832	.0008
mice	inc	—	—	6.3836	.0004
na.omit	inc	—	—	6.1499	.2341
true	age	—	—	49.4010	.0000
hd.ord	age	—	—	49.3646	.0364
hd.norm	age	—	—	49.3082	.0928
amelia	age	—	—	49.3981	.0029
mice	age	—	—	49.3979	.0031
na.omit	age	—	—	48.9935	.4075
true	interest	—	—	3.1920	.0000
hd.ord	interest	—	—	3.1903	.0017
hd.norm	interest	—	—	3.1877	.0043
amelia	interest	—	—	3.1923	.0003
mice	interest	—	—	3.1922	.0002
na.omit	interest	—	—	3.1297	.0623

Table 3.11: *Influence of Number of Variables (CCES 1,000)*

method	variable	CCES7Var	CCES10Var	CCES16Var	CCES15VarMult
true	Rep	.0000	.0000	.0000	.0000
hd.ord	Rep	.0010	.0006	.0003	.0003
hd.norm.orig	Rep	.0012	.0006	.0005	.0006
amelia	Rep	.0001	.0000	.0001	.0001
mice	Rep	.0000	.0000	.0000	.0000
na.omit	Rep	.0235	.0157	.0039	.0049
true	Ind	.0000	.0000	.0000	.0000
hd.ord	Ind	.0003	.0001	.0003	.0003
hd.norm.orig	Ind	.0001	.0001	.0003	.0004
amelia	Ind	.0000	.0000	.0001	.0001
mice	Ind	.0000	.0001	.0002	.0001
na.omit	Ind	.0104	.0101	.0158	.0163
true	Black	.0000	.0000	.0000	.0000
hd.ord	Black	.0004	.0002	.0001	.0001
hd.norm.orig	Black	.0003	.0001	.0002	.0002
amelia	Black	.0000	.0000	.0000	.0000
mice	Black	.0002	.0000	.0001	.0001
na.omit	Black	.0233	.0144	.0028	.0049
true	Hisp	.0000	.0000	.0000	.0000
hd.ord	Hisp	.0003	.0001	.0000	.0002
hd.norm.orig	Hisp	.0002	.0000	.0002	.0001
amelia	Hisp	.0000	.0000	.0000	.0001
mice	Hisp	.0001	.0000	.0001	.0002
na.omit	Hisp	.0243	.0154	.0076	.0097
true	Empl	.0000	.0000	.0000	.0000
hd.ord	Empl	.0014	.0003	.0000	.0000
hd.norm.orig	Empl	.0016	.0005	.0000	.0000
amelia	Empl	.0002	.0001	.0000	.0001
mice	Empl	.0003	.0001	.0000	.0001
na.omit	Empl	.0304	.0302	.0341	.0354
true	Stud	.0000	.0000	.0000	.0000
hd.ord	Stud	.0002	.0003	.0001	.0001
hd.norm.orig	Stud	.0001	.0001	.0001	.0001
amelia	Stud	.0001	.0001	.0000	.0000
mice	Stud	.0001	.0001	.0001	.0001
na.omit	Stud	.0147	.0042	.0052	.0059
true	Male	.0000	.0000	.0000	.0000
hd.ord	Male	.0001	.0000	.0000	.0003
hd.norm.orig	Male	.0000	.0003	.0000	.0001

amelia	Male	.0001	.0001	.0000	.0001
mice	Male	.0002	.0000	.0001	.0001
na.omit	Male	.0302	.0357	.0310	.0283
true	inc		.0000	.0000	.0000
hd.ord	inc		.0045	.0104	.0101
hd.norm.orig	inc		.0117	.0146	.0148
amelia	inc		.0008	.0006	.0005
mice	inc		.0004	.0005	.0003
na.omit	inc		.2341	.2446	.2255
true	interest		.0000	.0000	
hd.ord	interest		.0017	.0041	
hd.norm.orig	interest		.0043	.0057	
amelia	interest		.0003	.0000	
mice	interest		.0002	.0001	
na.omit	interest		.0623	.0334	
true	age		.0000		
hd.ord	age		.0364		
hd.norm.orig	age		.0928		
amelia	age		.0029		
mice	age		.0031		
na.omit	age		.4075		
true	Gay			.0000	.0000
hd.ord	Gay			.0000	.0001
hd.norm.orig	Gay			.0001	.0001
amelia	Gay			.0000	.0000
mice	Gay			.0001	.0001
na.omit	Gay			.0098	.0102
true	StudLoans			.0000	.0000
hd.ord	StudLoans			.0003	.0004
hd.norm.orig	StudLoans			.0001	.0002
amelia	StudLoans			.0000	.0001
mice	StudLoans			.0000	.0001
na.omit	StudLoans			.0213	.0231
true	InternetHome			.0000	.0000
hd.ord	InternetHome			.0001	.0000
hd.norm.orig	InternetHome			.0001	.0001
amelia	InternetHome			.0001	.0000
mice	InternetHome			.0000	.0000
na.omit	InternetHome			.0038	.0038
true	Moderate			.0000	.0000
hd.ord	Moderate			.0002	.0001
hd.norm.orig	Moderate			.0003	.0005
amelia	Moderate			.0000	.0001

mice	Moderate	.0001	.0001
na.omit	Moderate	.0178	.0200
true	NotReligious	.0000	.0000
hd.ord	NotReligious	.0001	.0001
hd.norm.orig	NotReligious	.0002	.0002
amelia	NotReligious	.0000	.0001
mice	NotReligious	.0000	.0001
na.omit	NotReligious	.0215	.0237
true	RentHome	.0000	.0000
hd.ord	RentHome	.0000	.0000
hd.norm.orig	RentHome	.0002	.0002
amelia	RentHome	.0001	.0000
mice	RentHome	.0000	.0000
na.omit	RentHome	.0107	.0146
true	Separated	.0000	.0000
hd.ord	Separated	.0004	.0003
hd.norm.orig	Separated	.0003	.0003
amelia	Separated	.0001	.0001
mice	Separated	.0001	.0001
na.omit	Separated	.0002	.0002

CHAPTER 4

MORAL ARGUMENTS AS A SOURCE OF FRAME STRENGTH

4.1 Introduction

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

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

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

4.2 Theory

4.2.1 Framing

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

4.2.2 Moralization

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

Table 4.1. Foundations of Moral Arguments

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

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

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

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

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

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

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

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

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

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

4.3 Data

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

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

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

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

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

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

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

4.3.3 Survey Experiment

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

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

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

4.4 Results

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

SURVEY ENVIRONMENT

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

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

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

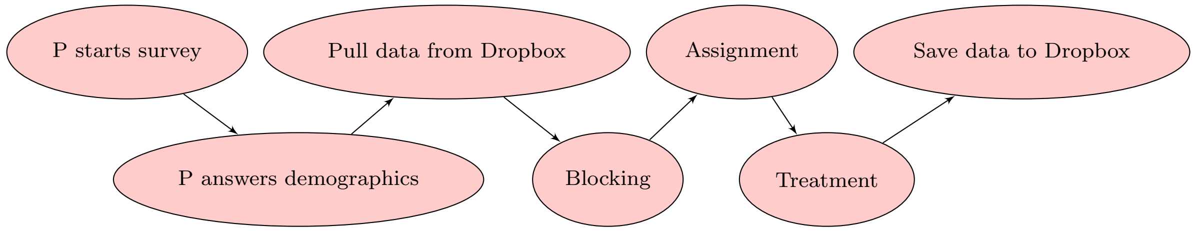

Figure A.1. Online survey experiment workflow

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

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

REFERENCES

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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.

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

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.

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

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

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

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

King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. *American Political Science Review*, *95*, 49–69.

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

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

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

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

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

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

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

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

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.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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