ONLINE SUPPLEMENT

for

Consider Your Origins: Parental Social Class and Preferences for Redistribution in the United States from 1977 to 2018

Alternative Question Wording of EQWLTH

In 1978, a subsample of respondents was assigned to the question item with an alternative wording as part of an experiment on question-wording effects. These wording alterations are referred as the "terse" alternatives and are labeled as "version Y" (Smith 2006). For example, the terse-alternative version of this variable is labeled EQWLTHY. The exact wording of the alternative is as follows:

> *"Some people think that the income differences between the rich and the poor ought to be reduced, perhaps by raising the taxes of wealthy families or by giving income assistance to the poor. Others think that the government should not concern itself with reducing this income difference between the rich and the poor. What score between 1 and 7 comes closest to the way you feel?"*

The list of response categories was the same with the non-altered question. I include all respondents who were assigned the terse alternative in the study sample unless they had to be excluded for other reasons. In the analysis, I model the wording-difference effect by including an indicator variable for whether the respondent was assigned the terse alternative. I do not find any substantial wording-difference effects for this item.

Use of Sampling Weights

All the analyses in this article apply the appropriate sampling weights provided for the GSS. The sampling weight, WTSSALL, is applied to all respondents prior to 2004. Beginning in 2004, the GSS adopted a two-stage sub-sampling design for non-response. To weight accordingly to this adjusted sample scheme, the GSS provides sampling weight, WTSSNR, from 2004 onward. In the analyses, I apply this weight for all respondents in 2004 and later. In 1982 and 1987, the GSS included an oversample of black respondents. To ensure racial balance of sampled respondents in these years, I re-weight these respondents using the weight OVERSAMP which weights down the black oversamples to match the racial distribution of the regular cross-section of the GSS (see Smith, Davern, Freese, and Morgan 2019: 3175).

Finally, after this weight has been constructed, I re-scale these weights to adjust for the difference in relative sample sizes of each survey period. The overall mean of the weight in each year is inversely scaled to its relative sample size. This procedure ensures that each survey period is accounted equivalently in the pooled-year analyses.

Imputation for Missing Information

In imputing for missing information on the variables used in the analyses, the emphasis was put on constructing a model that most precisely imputes missing information on class origins because the issue of missing information is most severe for this variable. Therefore, I use a rich set of variables that I consider to be good predictors of class origins. A full list of variables utilized for imputation is provided in Table S2.

Before any model-based attempts at imputation, I employ a logical imputation for respondents first interviewed in 2006, 2008, and 2010. Such respondents were part of the panel data sample of the GSS in which these respondents were subject to two follow-up surveys two and four years after the first interview. Using the panel datasets, I impute from later waves parents' occupation as well as items that are unlikely to vary: age (adjusting for time in between surveys), sex, race, years of education, spouse's years of education, U.S-born status, parents' U.S-born status, family income when the respondent was 16 (for consistency in response patterns of the panel respondents, see Hout and Hastings 2016).

After logical imputation using panel data, I impute missing values employing an iterative imputation algorithm using random forests proposed by Stekhoven and Bühlmann (2011).[1] When imputing current class and class origins, I allow imputation into only six class groups (salariat class, intermediate class, farming class, service working class, manual working class, and military). To check the sensitivity of the findings to the treatment of missing information, I replicated all the analyses in the paper, excluding all respondents initially missing information on current class and class origins. I find that the results are the same with only minor fluctuations.

Descriptives for Primary Measures

---

[1]In practice, the imputation was implemented using an **R** package **missForest** (Stekhoven 2013)

Table S2. Coding of Variables and Variables Used to for Imputation of Missing Values

Variable Label	Description	Availability
YEAR	Year of survey	GSS 77-18
AGE	Age at the time of the survey	GSS 77-18
SEX	Self-reported gender (Male; Female)	GSS 77-18
RACE	Self-reported race (White; Black; Other)	GSS 77-18
REGION	Respondent's region of residence (Based on nine Census divisions)	GSS 77-16
EDUC	Highest completed school level in years	GSS 77-18
REALINC	Family income in constant dollars	GSS 77-18
BORN	U.S born status (U.S born; Foreign born)	GSS 77-18*
PARBORN	Parents' U.S born status (Both U.S born; Only one parent U.S born; Both foreign born)	GSS 77-18*
MARITAL	Marital status at the time of the survey (Married Widowed; Divorced; Separated; Never Married)	GSS 77-18*
CLASS	Subjective-class identification (Lower-class; Working-class; Middle-class; Upper-class)	GSS 77-18
POLVIEWS	Political orientation of the respondent (7-point scale ranging from "extremely liberal" to "extremely conservative")	GSS 77-18
INCOM16	Family income when 18 years old (Far below average; Below average; Average; Above average; Far above average)	GSS 77-18 (except GSS 96-00)
FAMILY16	Family structure when 16 years old (Two biological parents; Two-parent & one-biological; Single-parent; Living with Relatives; Other)	GSS 77-18*
RES16	Type of place lived in when 16 years old (Non-farm country; Farm; Town less than 50,000; Town of 50,000 to 250,000; Big-city Suburb; City greater than 250,000)	GSS 77-18*
REG16	Region of residence when 16 years old (Based on nine Census divisions with an additional category for a foreign country)	GSS 77-18*
RELIG16	Religion in which the respondent was raised in (Protestant; Catholic; Jewish; None; Other)	GSS 77-18*
PAEDUC	Father's highest completed school level in years	GSS 77-18
MAEDUC	Mother' highest completed school level in years	GSS 77-18
SPEDUC	Father's highest completed school level in years	GSS 77-18

Note: The super-scripted asterisks indicate that the relevant variable is not available for respondents assigned Ballot D in the 2006 GSS. The list in parentheses indicate all unique categories respondents could have been classified into.

Table S3: Sample Descriptive by Class Origins

			Class Origins			
	All	Salariat	Intermediate	Farming	Working	Military
EQWLTH	4.2	3.8	4.0	4.4	4.4	4.1
	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.1)
TAX						
"Too high"	0.67	0.61	0.67	0.67	0.70	0.65
"Too low" or "About right"	0.33	0.39	0.33	0.33	0.30	0.35
Age	44.3	43.3	44.0	47.2	44.3	43.2
	(9.3)	(9.1)	(9.3)	(9.2)	(9.3)	(8.9)
Gender						
Male	0.53	0.53	0.52	0.56	0.52	0.45
Female	0.47	0.47	0.48	0.44	0.48	0.55
Race						
White	0.82	0.86	0.86	0.77	0.79	0.82
Black	0.12	0.07	0.08	0.13	0.16	0.10
Other	0.06	0.06	0.06	0.10	0.05	0.08
Parent's Place of Birth						
Both in the US	0.82	0.81	0.81	0.79	0.84	0.82
One born in the US	0.06	0.06	0.06	0.05	0.05	0.08
Both born outside the US	0.13	0.14	0.13	0.16	0.11	0.09
Place of Residence at 16						
Foreign	0.08	0.10	0.09	0.13	0.04	0.14
New England	0.05	0.06	0.06	0.01	0.06	0.04
Middle Atlantic	0.16	0.17	0.17	0.04	0.18	0.05
East North Central	0.19	0.17	0.18	0.13	0.22	0.05
West North Central	0.08	0.07	0.07	0.15	0.06	0.06
South Atlantic	0.15	0.13	0.15	0.17	0.16	0.28
East South Central	0.07	0.04	0.05	0.13	0.07	0.05
West South Central	0.08	0.07	0.08	0.13	0.08	0.09
Mountain	0.04	0.05	0.05	0.05	0.03	0.07
Pacific	0.10	0.13	0.11	0.05	0.09	0.18
Type of Residence at 16						
Non-farm country	0.12	0.08	0.11	0.10	0.14	0.07
Farm	0.13	0.03	0.06	0.70	0.07	0.02
Town less than 50,000	0.32	0.30	0.34	0.15	0.35	0.37
Town of 50,000 to 250,000	0.16	0.19	0.18	0.02	0.17	0.25
Big-city Suburb	0.12	0.22	0.14	0.01	0.10	0.15
City greater than 250,000	0.16	0.17	0.18	0.02	0.17	0.13
Religion at 16						
Protestant	0.59	0.54	0.55	0.69	0.61	0.59
Catholic	0.31	0.30	0.34	0.25	0.31	0.31
Jewish	0.02	0.05	0.03	0.00	0.01	0.00
None	0.05	0.06	0.05	0.04	0.05	0.06
Other	0.03	0.05	0.03	0.02	0.02	0.03
Years of Education	13.5	15.5	14.0	11.6	12.8	14.4
	(3.0)	(2.6)	(2.8)	(3.6)	(2.7)	(2.7)
Father's Years of Education	10.6	14.6	11.0	7.4	9.4	13.0
	(3.9)	(3.3)	(3.2)	(3.6)	(3.0)	(3.2)
Mother's Years of Education	10.9	13.7	11.3	8.3	10.0	12.2
	(3.5)	(2.8)	(3.0)	(3.8)	(3.1)	(2.5)
Family Income (in 1,000 USD)	41.9	54.1	45.1	34.2	36.3	45.2
	(31.6)	(37.7)	(32.3)	(27.0)	(27.0)	(34.4)
Current Class						
Salariat	0.25	0.45	0.28	0.14	0.17	0.32
Intermediate	0.37	0.34	0.42	0.32	0.35	0.36
Farming	0.01	0.00	0.01	0.07	0.01	0.01
Working	0.36	0.20	0.28	0.46	0.46	0.28
Military	0.00	0.01	0.01	0.00	0.00	0.03
Subjective Class						
Lower Class	0.03	0.02	0.02	0.04	0.04	0.03
Working Class	0.49	0.30	0.44	0.61	0.58	0.45
Middle Class	0.45	0.61	0.50	0.33	0.37	0.49
Upper Class	0.03	0.06	0.04	0.02	0.02	0.04
n (unweighted)	25,031	5,185	6,703	2,476	10,266	441

Note: Standard deviations reported in parentheses. Sampling weights applied.

References

Hout, Michael and Orestes P Hastings. 2016. "Reliability of the core items in the General Social Survey: Estimates from the three-Wave panels, 2006–2014." *Sociological Science* 3:971–1002.

Smith, Tom William. 2006. *Wording Effects on the National Spending Priority Items across Time, 1984-2004.* Chicago, IL: GSS Methodological Report No. 107, NORC.

Smith, Tom W., Michael Davern, Jeremy Freese, and Stephen L. Morgan. 2019. *General Social Surveys, 1972-2018: Cumulative codebook.* Chicago, IL: NORC.

Stekhoven, Daniel J. 2013. *missForeset: Nonparametric Missing Value Imputation using Random Forest.* Retrieved August 7, 2019 (https://cran.r-project.org/web/packages/missForest/missForest.pdf (R package version 1.4)).

Stekhoven, Daniel J and Peter Bühlmann. 2011. "MissForest—non-parametric missing value imputation for mixed-type data." *Bioinformatics* 28:112–118.