APPENDIX

Hooghe, M., & Oser, J. (2017). Partisan strength, political trust and generalized trust in the United States: An analysis of the General Social Survey, 1972-2014. *Social Science Research, 68*, 132-146. https://doi.org/10.1016/j.ssresearch.2017.08.005

The replication code for all analyses in article and appendix can be found in the following file archived in the Harvard Dataverse:

Hooghe, Marc; Oser, Jennifer, 2017, "Replication for 'Partisan strength, political trust and generalized trust in the United States: An analysis of the General Social Survey, 1972-2014'", doi:10.7910/DVN/DEUQRY, Harvard Dataverse.

Table A1. Descriptive Statistics and Missing Data

Variable	n	Mean	Std. Dev.	Min	Max
Political trust	40,007	1.95	0.51	1	3
Generalized trust	40,116	2.00	0.76	1	3
Partisan strength	58,323	1.74	1.01	0	3
Incumbent party	59,599	0.59	0.49	0	1
Year of Survey	59,599	1993.02	12.30	1972	2014
Age	59,388	45.84	17.48	18	89
Gender	59,599	0.56	0.50	0	1
Education	59,434	12.79	3.18	0	20
Income	53,546	44683.68	36296.54	369.5	180386
TV watching	35,524	1.90	1.21	0	4
Newspaper reading	37,364	3.87	1.34	1	5
Religious attendance	59,037	0.49	0.50	0	1
Black	59,599	0.14	0.35	0	1
Other race	59,599	0.05	0.22	0	1
Size location	57,956	3.50	2.15	0	9.00884

Note: See Table 1 for variable coding and value ranges

Summary of missing data issues for GSS cumulative file

Basic sociodemographic variables noted in the descriptive statistics that are missing 1% or less of data are "permanent items" that were included for all cases in all years (namely age, gender, education, religious attendance, race and size location). As evident from Table A1 there are a number of variables central to the analyses in the article that have substantial missing cases, namely the dependent variables of political trust and generalized trust, as well as the control variables of TV watching, newspaper reading, and income.

There are three distinct reasons for missing data on these variables in the GSS cumulative file, as documented in the GSS codebook (Smith et al. 2016), and in our analysis replication file:

1. **Item non-response:** This is an issue only for the income variable, which is missing 10% cases for the cumulative data file, ranging from a low of 6% missing cases in 1975 to a high of 14% missing cases in 2006. Missing data on questions related to income is a common problem for this type of survey, and the GSS proportion of missing cases on income are fairly low in comparison to other high quality surveys.

2. **Rotation design:** From 1972 through 1987, the GSS used a "rotation design" that included many items in two out of every three surveys waves (Smith et al. 2016). Table A1a below documents the years in which variables in our analyses were omitted from the GSS survey due to this sampling design.

Table A1a. GSS rotation design - omitted questions in specific years, 1972-1987 (O= omitted in that specific year)

Year	Political Trust	Generalized Trust	Television	Newspaper
1972	O		O	
1973			O	O
1974		O	O	O
1975				
1976			O	O
1977		O		
1978				
1980				O
1982		O		
1983				
1984			O	O
1985	O	O		
1986				
1987			O	

3. **Split-ballot design**: Beginning in 1988, the GSS implemented a split-ballot design that conducts different versions of the survey for different random sub-samples of respondents. For the same four variables documented in Table A1a, the split ballot design implemented beginning in 1988 randomly assigned the questions to different subsets of the sample. The proportion of missing data due to the split ballot design was 1/3 or less for most survey waves from 1988 through 2014, and the sample size was sufficient for valid statistical inference throughout this period. The years in which more than 1/3 of the data on these four variables are missing data due to the split ballot design was in 2002, 2004 and 2006 when missing data due to split ballot design ranged between 1/2 and 2/3 of the sample. Notably, the total sample size during these years was large enough to ensure robust statistical inference even with this relatively high proportion of missing data due to sample design (for 2002, n=2765; for 2004, n=2812; for 2006, n=4510). Missing data due to non-response is negligible (less than 5%) for variables in our analyses that were subject to the split ballot design.

Analytic strategy for addressing missing data

Based on this summary of the missing data issues with the GSS cumulative data, we implemented the following strategies for addressing missing data:

a. **Listwise deletion due to rotation sample and split ballot design:** For the years in which the GSS rotation sample omitted a variable from our analysis, this year is omitted from all regression analyses through listwise deletion. Thus, in accordance with Table A1a, analyses are not conducted for the years 1972, 1973, 1974, 1976, 1977, 1980, 1982, 1984, 1985 and 1987. In addition, as the split-ballot design was implemented for random subsamples (Smith et al. 2016), we use listwise deletion as the missing data due to this design qualifies as “missing completely at random," and therefore listwise deletion does not introduce bias.

b. **Non-response for income**: Multiple imputation or maximum likelihood can produce approximately unbiased estimates when data are missing at random, but prior research indicates that income data are often not missing at random, with lower response rates for those with very low and very high socio-economic status. In contrast, listwise deletion produces unbiased estimates even if the data are not missing at random. In addition, because income is a predictor variable in the regression, listwise deletion is a less biased approach than multiple imputation or maximum likelihood (Allison 2001, 2009). We therefore use listwise deletion for the income variable in the models reported in the manuscript, with robustness tests performed with and without income as a control variable which showed no substantive difference in the findings.

c. **Full information maximum likelihood (FIML)**: In order to address the possible effects of any missing data that was not completely at random, we repeated the analysis using FIML. The FIML estimator, unlike the usual least squares or maximum likelihood estimators under listwise deletion, uses all available information from respondents. FIML is also preferable over multiple implementation, which is sensitive to mis-specified imputation models (Enders 2001, 2010). This estimation was conducted using the sem

command in Stata 14 with the method(mlmv) option for maximum likelihood with missing values. In the Appendix Tables A6 and A7, we report on models that parallel the main findings tables of the manuscript (Tables A2 and A3) but with a FIML estimator. Tables A6 and A7 have a larger number of observations, as expected, (between about 33,000 and 39,000 depending on the model) and all support the substantive interpretations and conclusions in the manuscript. We can safely conclude, therefore, that these missing data do not pose a challenge for the validity of our findings.

Table A2. Determinants of Political Trust

Parallel to Table 2, with clustered standard errors by year for Models I and II

	Clustered standard errors		*Fixed effects*	
	Model I	Model II	Model III	Model IV
Partisan strength, leaner		0.063***		0.065***
		(0.012)		(0.016)
Partisan strength, weak	0.053**	0.110***	0.048**	0.109***
	(0.014)	(0.015)	(0.016)	(0.015)
Partisan strength, strong	0.057*	0.113***	0.056**	0.113***
	(0.020)	(0.016)	(0.018)	(0.016)
Specific party ID, Republican	0.029		0.024	
	(0.033)		(0.020)	
Party ID weak * Republican	-0.012		-0.007	
	(0.024)		(0.024)	
Party ID strong * Republican	-0.014		-0.014	
	(0.033)		(0.028)	
Year of survey	-0.003*	-0.004*		
	(0.001)	(0.001)		
TV watching, very low	0.065	0.064	0.061*	0.060*
	(0.033)	(0.031)	(0.026)	(0.024)
TV watching, low	0.064	0.061	0.059*	0.057*
	(0.033)	(0.032)	(0.028)	(0.025)
TV watching, high	0.065	0.055	0.063*	0.053*
	(0.034)	(0.034)	(0.029)	(0.027)
TV watching, very high	0.029	0.028	0.026	0.024
	(0.035)	(0.035)	(0.029)	(0.026)
Newspaper reading	0.019**	0.014*	0.017***	0.013**
	(0.007)	(0.006)	(0.004)	(0.004)
Age	-0.003***	-0.002***	-0.003***	-0.002***
	(0.001)	(0.001)	(0.000)	(0.000)
Gender, female	0.004	0.007	0.001	0.005
	(0.008)	(0.008)	(0.010)	(0.009)
Education, years of schooling	0.003	0.005	0.003	0.005**
	(0.003)	(0.003)	(0.002)	(0.002)
Income, household	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Religious attendance, 1/month+	0.041**	0.043**	0.041***	0.042***
	(0.012)	(0.012)	(0.010)	(0.010)
Race, black	-0.038	-0.044	-0.042*	-0.048**
	(0.020)	(0.028)	(0.017)	(0.015)
Race, other	0.130***	0.146***	0.131***	0.148***
	(0.029)	(0.026)	(0.027)	(0.024)
Size location	0.003	0.004	0.004	0.005*
	(0.002)	(0.003)	(0.003)	(0.002)
Incumbency	0.083*	0.083*	0.000	0.000
	(0.029)	(0.029)	(.)	(.)
Constant	8.471**	8.904**	1.824***	1.748***
	(2.839)	(2.777)	(0.047)	(0.041)
Observations	12452	14526	12452	14526
Adjusted R^2	0.039	0.044	0.052	0.057

Source: General Social Survey, 1972-2014. Entries are coefficients of ordinary least squares regression, followed by standard errors. For Models I and II, standard errors are clustered on year; robust standard errors in parentheses. For Models III and IV, standard errors are not clustered; fixed effects specification is used with year dummy (coefficients not shown).
Sig: *** p<0.001, ** p<0.001, * p<0.05.

Table A3. Determinants of Generalized Trust

Parallel to Table 3, with clustered standard errors by year for Models I and II

	Clustered standard errors		*Fixed effects*	
	Model I	Model II	Model III	Model IV
Partisan strength, leaner		0.063**		0.062**
		(0.020)		(0.022)
Partisan strength, weak	-0.039*	0.028	-0.038	0.027
	(0.018)	(0.018)	(0.023)	(0.020)
Partisan strength, strong	-0.062*	0.003	-0.063*	0.003
	(0.025)	(0.015)	(0.025)	(0.021)
Specific party ID, Republican	0.009		0.010	
	(0.027)		(0.027)	
Party ID weak * Republican	0.005		0.007	
	(0.028)		(0.034)	
Party ID strong * Republican	0.010		0.015	
	(0.036)		(0.038)	
Year of survey	-0.009***	-0.009***		
	(0.001)	(0.001)		
TV watching, very low	-0.010	0.005	-0.003	0.014
	(0.037)	(0.034)	(0.037)	(0.033)
TV watching, low	-0.015	-0.005	-0.010	0.002
	(0.050)	(0.046)	(0.039)	(0.035)
TV watching, high	-0.056	-0.038	-0.050	-0.030
	(0.044)	(0.043)	(0.041)	(0.037)
TV watching, very high	-0.089*	-0.077	-0.082*	-0.068
	(0.042)	(0.039)	(0.040)	(0.036)
Newspaper reading	0.023**	0.028***	0.024***	0.030***
	(0.007)	(0.006)	(0.006)	(0.005)
Age	0.008***	0.008***	0.008***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)
Gender, female	0.064***	0.061***	0.064***	0.061***
	(0.015)	(0.014)	(0.014)	(0.013)
Education, years of schooling	0.061***	0.060***	0.062***	0.060***
	(0.003)	(0.002)	(0.003)	(0.002)
Income, household	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Religious attendance, 1/month+	0.059**	0.060***	0.058***	0.060***
	(0.016)	(0.015)	(0.014)	(0.013)
Race, black	-0.360***	-0.341***	-0.359***	-0.341***
	(0.027)	(0.027)	(0.022)	(0.020)
Race, other	-0.174***	-0.139***	-0.169***	-0.134***
	(0.028)	(0.018)	(0.033)	(0.029)
Size location	-0.009*	-0.010**	-0.009**	-0.010**
	(0.003)	(0.003)	(0.003)	(0.003)
Incumbency	0.002	0.004	0.000	0.000
	(0.023)	(0.022)	(.)	(.)
Constant	18.249***	18.035***	0.909***	0.850***
	(2.152)	(1.942)	(0.062)	(0.054)
Observations	13191	15462	13191	15462
Adjusted R^2	0.174	0.170	0.176	0.172

Source: General Social Survey, 1972-2014. Entries are coefficients of ordinary least squares regression, followed by standard errors. For Models I and II, standard errors are clustered on year; robust standard errors in parentheses. For Models III and IV, standard errors are not clustered; fixed effects specification is used with year dummy (coefficients not shown).
Sig: *** p<0.001, ** p<0.001, * p<0.05

Table A4. Partisan Strength and Political Trust Over Time

Parallel to Table 4, with clustered standard errors

Year centered at mean	-0.004*
	(0.001)
Partisan strength	0.025**
	(0.008)
Year*Partisan strength	-0.000
	(0.000)
Democrat	0.041*
	(0.017)
Republican	0.061
	(0.031)
TV watching, very low	0.064*
	(0.030)
TV watching, low	0.062
	(0.031)
TV watching, high	0.055
	(0.034)
TV watching, very high	0.028
	(0.035)
Newspaper reading	0.014*
	(0.006)
Age	-0.002***
	(0.000)
Gender, female	0.009
	(0.007)
Education, years of schooling	0.005
	(0.003)
Income, household	0.000
	(0.000)
Religious attendance, 1/month+	0.041***
	(0.010)
Race, black	-0.040
	(0.021)
Race, other	0.149***
	(0.026)
Size location	0.004
	(0.003)
Incumbency	0.083*
	(0.029)
Constant	1.713***
	(0.066)
Observations	14526
Adjusted R^2	0.044

Source: General Social Survey, 1972-2014. Entries are coefficients of ordinary least squares regression with standard errors clustered on year; robust standard errors in parentheses. Sig: *** p<0.001, ** p<0.01, * p<0.05.

Table A5. Partisan Strength and Generalized Trust Over Time

Parallel to Table 5, with clustered standard errors

Year of the survey	-0.009***
	(0.001)
Partisan strength	-0.030**
	(0.009)
Year*Partisan strength	0.001*
	(0.000)
Democrat Identifier	0.081**
	(0.027)
Republican Identifier	0.101**
	(0.030)
TV watching, very low	0.004
	(0.034)
TV watching, low	-0.007
	(0.046)
TV watching, high	-0.040
	(0.044)
TV watching, very high	-0.078
	(0.039)
Newspaper reading	0.028***
	(0.006)
Age	0.008***
	(0.000)
Gender, female	0.062***
	(0.014)
Education, years of schooling	0.059***
	(0.002)
Income, household	0.000***
	(0.000)
Religious attendance, 1/mo.+	0.058**
	(0.015)
Race, black	-0.335***
	(0.027)
Race, other	-0.135***
	(0.019)
Size location	-0.010**
	(0.003)
Incumbency	0.004
	(0.022)
Constant	0.668***
	(0.060)
Observations	15462
Adjusted R^2	0.170

Source: General Social Survey, 1972-2014. Entries are coefficients of ordinary least squares regression with standard errors clustered on year; robust standard errors in parentheses. Sig: *** p<0.001, ** p<0.01, * p<0.05. Note that in contrast to Table 5 in the article, the findings in this table show that when clustered standard errors are added to the model specification, the interaction between year and partisan strength becomes marginally significant (p<.05), but is substantively inconsequential in size.

Table A6. Determinants of Political Trust, FIML Estimator
Parallel to Table 2, with full information maximum likelihood (FIML) estimator

	OLS		*Fixed effects approximation*	
	Model I	Model II	Model III	Model IV
Partisan strength, leaner		0.058***		0.058***
		(0.013)		(0.013)
Partisan strength, weak	0.047***	0.107***	0.045***	0.105***
	(0.013)	(0.011)	(0.013)	(0.012)
Partisan strength, strong	0.052***	0.106***	0.053***	0.105***
	(0.014)	(0.013)	(0.014)	(0.013)
Specific party ID, Republican	0.009		0.007	
	(0.016)		(0.016)	
Party ID weak * Republican	0.005		0.005	
	(0.019)		(0.019)	
Party ID strong * Republican	-0.010		-0.013	
	(0.022)		(0.022)	
Year of survey	-0.003***	-0.003***		
	(0.000)	(0.000)		
TV watching, very low	0.073**	0.073**	0.071**	0.071**
	(0.025)	(0.022)	(0.025)	(0.022)
TV watching, low	0.073**	0.074**	0.071**	0.072**
	(0.026)	(0.024)	(0.026)	(0.024)
TV watching, high	0.074**	0.064*	0.071*	0.062*
	(0.028)	(0.025)	(0.028)	(0.025)
TV watching, very high	0.044	0.042	0.041	0.039
	(0.027)	(0.025)	(0.028)	(0.025)
Newspaper reading	0.021***	0.016***	0.019***	0.014***
	(0.004)	(0.004)	(0.004)	(0.004)
Age	-0.003***	-0.003***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
Gender, female	0.001	0.006	0.002	0.006
	(0.008)	(0.007)	(0.008)	(0.007)
Education, years of schooling	0.004**	0.005***	0.004*	0.005***
	(0.002)	(0.001)	(0.002)	(0.001)
Income, household	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Religious attendance, 1/month+	0.045***	0.045***	0.045***	0.044***
	(0.008)	(0.007)	(0.008)	(0.007)
Race, black	-0.022	-0.021	-0.022	-0.021
	(0.013)	(0.012)	(0.013)	(0.012)
Race, other	0.120***	0.135***	0.120***	0.135***
	(0.019)	(0.016)	(0.019)	(0.016)
Size location	0.005*	0.006**	0.005*	0.006**
	(0.002)	(0.002)	(0.002)	(0.002)
Incumbency	0.097***	0.095***	0.097***	0.095***
	(0.008)	(0.008)	(0.008)	(0.008)
Year (Mean Centered)			-0.004***	-0.004***
			(0.001)	(0.001)
Year Squared			-0.000***	-0.000***
			(0.000)	(0.000)
Year Cubed			-0.000	0.000
			(0.000)	(0.000)
Constant	7.039***	7.066***	1.752***	1.693***
	(0.851)	(0.793)	(0.039)	(0.034)
Observations	32929	38958	32929	38958

General Social Survey, 1972-2014. Entries are regression coefficients, followed by standard errors in parentheses. Models I and II, OLS; Models III and IV, year as cubic spline to approximate fixed effects.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table A7. Determinants of Generalized Trust, FIML Estimator
Parallel to Table 3, with full information maximum likelihood (FIML) estimator

	OLS		*Fixed effects approximation*	
	Model I	Model II	Model III	Model IV
PID: Lean		0.047**		0.045**
		(0.016)		(0.016)
PID: Weak	-0.001	0.055***	-0.001	0.054***
	(0.017)	(0.015)	(0.017)	(0.015)
PID: Strong	-0.016	0.031	-0.016	0.030
	(0.019)	(0.016)	(0.019)	(0.016)
PID: Republican	-0.006		-0.005	
	(0.021)		(0.021)	
PID Republican X PID Weak	0.019		0.020	
	(0.026)		(0.026)	
PID Republican X PID Strong	-0.000		0.004	
	(0.029)		(0.029)	
Year of survey	-0.009***	-0.009***		
	(0.001)	(0.001)		
TV View: 1st Quartile	-0.014	-0.004	-0.011	-0.000
	(0.035)	(0.032)	(0.035)	(0.031)
TV View: 2nd Quartile	-0.016	-0.012	-0.015	-0.010
	(0.037)	(0.033)	(0.037)	(0.033)
TV View: 3rd Quartile	-0.072	-0.059	-0.070	-0.055
	(0.039)	(0.035)	(0.039)	(0.035)
TV View: 4th Quartile	-0.096*	-0.090**	-0.093*	-0.086*
	(0.038)	(0.034)	(0.038)	(0.034)
Newspaper reading	0.026***	0.030***	0.028***	0.032***
	(0.006)	(0.005)	(0.006)	(0.005)
Age	0.008***	0.008***	0.008***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)
Gender, female	0.058***	0.049***	0.058***	0.048***
	(0.011)	(0.010)	(0.011)	(0.010)
Education, years of schooling	0.056***	0.054***	0.056***	0.055***
	(0.002)	(0.002)	(0.002)	(0.002)
Income, household	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Religious attendance, 1/month+	0.067***	0.069***	0.067***	0.068***
	(0.011)	(0.010)	(0.011)	(0.010)
Race, black	-0.359***	-0.340***	-0.358***	-0.339***
	(0.016)	(0.015)	(0.016)	(0.015)
Race, other	-0.168***	-0.127***	-0.165***	-0.125***
	(0.023)	(0.020)	(0.023)	(0.020)
Size location	-0.003	-0.003	-0.003	-0.003
	(0.003)	(0.002)	(0.003)	(0.002)
Incumbency	-0.022*	-0.019	-0.018	-0.013
	(0.011)	(0.010)	(0.011)	(0.010)
Year (Mean Centered)			-0.012***	-0.013***
			(0.001)	(0.001)
Year Squared			0.000***	0.000***
			(0.000)	(0.000)
Year Cubed			0.000***	0.000***
			(0.000)	(0.000)
Constant	19.483***	19.439***	0.688***	0.643***
	(1.108)	(1.020)	(0.051)	(0.044)
Observations	32929	38958	32929	38958

General Social Survey, 1972-2014. Entries are regression coefficients, followed by standard errors in parentheses. Models I and II, OLS; Models III and IV, year as cubic spline to approximate fixed effects. Note that in the FIML specification, the "weak" level of partisan strength becomes significant, but consistent with the findings of Table 3, the highest level of partisan strength ("strong") has a lower level of generalized trust in comparison to those who have weaker partisan strength. * $p < .05$, ** $p < .01$, *** $p < .001$

References

Allison, Paul D. 2001. Missing data. *Thousand Oaks, CA*: Sage.

Allison, Paul D. 2009. Missing Data. In *The Sage Handbook of Quantitative Methods in Psychology*, edited by R. E. Millsap and A. Maydeu-Olivares. Thousand Oaks, CA: Sage.

Enders, Craig K. 2001. A Primer on Maximum Likelihood Algorithms Available for Use with Missing Data. *Structural Equation Modeling* 8 (1):128-141.

Enders, Craig K. 2010. *Applied missing data analysis*. New York: Guilford Press.

Smith, Tom W, Peter Marsden, and Michael Hout. 2016. *General Social Surveys, 1972-2014: Cumulative Codebook*. Chicago: National Opinion Research Center.