

The Role of Student Researchers in Attenuating the Replication Crisis While Performing Original Research: Replication and Original Research on Political Polarization

Kenneth Collins

B.A., State University Of New York at Stony Brook, 2018

Abstract:

This thesis presents results from the replication of a paper from the Political Polarization literature as a first step in conducting original research. Also discussed is the importance of replication in context of the replication crisis and the role of student researchers in performing replication while building valuable research skills.

Table of Contents

Abstract:	1
1. Literature review	3
2. Replication	4
2.1 Replication of Baldassarri and Gelman 2008	4
2.2 The Replication Models	5
Table One	6
Table Two	6
2.3 Why Scientists Value Replication and the “Replication Crisis”	7
3. Original Research	10
3.1 Why Political Polarization Matters	10
3.2 Statistical Conclusion Validity of the Model	11
3.3 Hypothesis Exploration	12
Table 3	13
3.4 External Validity	13
3.5 Discussion of Results	14
References	15

1. Literature review

There has been an ongoing discussion in the literature for several years about whether polarization exists. Much of this is due to different ways of defining and operationalizing polarization. (Lelkes 2016) There was an attempt to prove polarization is or is not happening instead of a nuanced discussion of the ways the United States is and is not becoming polarized. (e.g. Abramowitz and Saunders 1998, Abramowitz and Saunders 2005, Fiorina et al. 2006, Abramowitz and Saunders 2008, Fiorina and Abrams 2008, Fiorina et al. 2008, Webster and Abramowitz 2017)

Diermeier and Li 2017 proposed a formal mathematical model. This model attempted to take into account the effects of partisan affect and elite polarization.

Ideology can be measured in multiple ways. Two popular ways are symbolic ideology and operational ideology. Symbolic ideology is a self-reported score, often on a seven-point scale. Operational ideology involves constructing a scale from the respondents position on multiple public policy questions. There is evidence that symbolic ideology does not always correspond to operational ideology. (Ellis and Stimson 2012)

Another way to define Political Polarization is as a tendency for correlations between an issue position and Party Identification (i.e. Party ID), Symbolic Ideology, or other issue positions to increase over time. Baldassarri and Gelman 2008 found that this type of polarization is occurring with Party ID as well as Symbolic Ideology. However the authors did not find this happening between issue positions.

One possible explanation is that the change in the correlations between issue positions had not shown up in the data by 2004 but would show up when a longer time period was considered. It could be that it was a weaker effect or there was a lag.

A second explanation is that the increases in the correlation between issue positions were “canceled out” by decreases, while that did not happen to the same degree with party identification and ideology.

A third explanation is that the wrong variables were modeled by Baldassarri and Gelman.

2. Replication

2.1 Replication of Baldassarri and Gelman 2008

The first step to replicate the paper was to contact the authors looking for any replication materials they may have. The authors provided an R-script. The second step was to find a copy of the dataset the authors used. The ANES no longer lists this datafile on their website. Luckily, Professor Reuben Kline was able to find a copy at the Association of Religious Data Archives or ARDA website. The data came as an xlsx file. The xlsx file was converted to a csv file with Microsoft Excel.

The next step was to attempt to run the R-script. The R-script did not include the process of cleaning the data and recoding the variables. As a result, the most feasible way to replicate this paper was to write a script to do clean the data, recode the variables, and perform the data analysis.

The R-script first had to read the data set into R. After reading the dataset into R, the data was cleaned. After the data was clean, the codebook was consulted to determine which variables from the data set were used in the analysis by Baldassarri and Gelman. There is no way to tell which variables were used, but there were clues in the authors paper and their R-script.

After determining which variables were to be used, variables were recoded in the same direction as Baldassarri and Gelman, liberal to conservative. There were comments in the code provided about how the variables were coded. Also, the dataset contained various types of missing values, which were recoded to "NA"s.

Next, a subsetting of the data was necessary. The datafile contained years not included in the authors model. After subsetting the data, the next step was to produce the correct correlation coefficients. After that, a variable was created that indicated which issue was correlated with party identification or ideology. Then, a variable that indicated the year of the correlation coefficient was created.

The first problem with replication of the paper was that the authors did not provide a data set. The ANES did not have a copy of that data set available. Having a copy of the dataset as part of the replication file would have made the process of replication simpler.

Also, the script did not have the code to clean the data and recode the variables. This made it unclear which variables were selected by the authors and how the data was cleaned. This ambiguity was dealt with by attempting to guess from the codebook, paper, and variable names in the R-script. It may not have been the same variables. Note that finding similar results despite challenges in replication would aid in confirming the statistical conclusion validity but not necessarily the construct validity of the study. These will be addressed in more detail below.

2.2 The Replication Models

The models were varying slope, varying intercept models. Four of the models from the paper Baldassarri and Gelman, 2008 were replicated. Table One contains two of the replicated models as well as their counterparts from the paper. The Party ID model from Table One was a varying slope varying intercept model. The dependent variable was a set of correlation coefficients. These coefficients were the correlations between the respondents Party ID and their responses to policy questions.

Only the complete pairs of variables were used in computing correlation coefficients. To put this another way, issue i and Party ID/Ideology were correlated if there was a response for issue i and Party ID/Ideology. An alternative way would be to only use data from respondents who had a response for all issues and the Party ID/Ideology question. The model formula is,

$$y_{it} = \alpha_i + \beta_i t + \varepsilon_{it}$$

There are 47 issues in the model, thus $i = \{1, 2, \dots, 46, 47\}$.

Multilevel Models are used when there are groups or some other multilevel structure. For example, if there were national standardized test data for students grouped within a class, within a school, within a school district, within a county, within a state for all 50 states, a Multilevel Model would be appropriate. This would be an example of a nested multilevel model.

Multilevel Models can be thought of as a middle group between not including the group in a regression model and estimating separate regressions for each group. (Gelman and Hill, 2007) The former is also called "complete pooling" while the later is also called "no pooling". Multilevel models allow for "partial pooling".

The replicated models are shown in Tables 1 and 2. The models from Baldassarri and Gelman 2008 are also in these tables. In these tables, the numbers in parentheses are standard errors.

Wald Tests were performed on the fit models. It is worth noting that issues with Wald Tests have been noted in the multilevel modeling literature. (eg Hox 1998, Berkhof and Snijders 2001)

The Wald Tests showed that the time component was significant in all four models. The p-values are in square brackets in the tables. This indicates a significant change in the correlation between Party Id / Ideology with issue position over time. The Wald test for the Party ID model from table 2 shows type significant at the .05 level (0.004855). The type variable in the Ideology model from table 2 does not have significance at the .05 level, with a p-value of about 0.5 (0.4943). This is potentially due to the issues associated with this test in Multilevel Modeling Literature. Baldassari and Gelman reported the type issue as a significant factor in Political Polarization for both Party ID and Ideology. Also, there were no interaction terms significant at the .05 level (0.225702, 0.1678). Hypothesis testing is discussed in more detail below.

Table One	Party ID {table 1}	Party ID {replication}	Ideology {table 1}	Ideology {replication}
Intercept	.17 (0.01)	0.152634 (0.014790)	.22 (0.01)	0.207114 (0.015325)
Time (decades)	.05 (0.01)	0.046443 (0.006554) $\left[\frac{1.383}{10^{12}} \right]$.04 (0.01)	0.032110 (0.007282) $\left[\frac{1.036}{10^5} \right]$
Residual SD : Intercept	.08	0.09752	.08	0.10097
Residual SD : Trends	.03	0.03387	.03	0.03910
Residual SD : Data	.04	0.04362	.04	0.04492

Table Two	Party ID (table 2)	Party ID (replication)	Ideology (table 2)	Ideology (replication)
Intercept	.24 (.02)	0.21383 (0.02405)	.24 (.02)	0.219687 (0.027568)
Time (decades)	.04 (.01)	0.03916 (0.01182) $\left[\frac{1.302}{10^{12}} \right]$.03 (.01)	0.029068 (0.013018) $\left[\frac{8.712}{10^6} \right]$
Economic issues	baseline	baseline	baseline	baseline
Civil rights issues	-.06 (.02)	-0.05891 (0.03280)	-.02 (.03)	-0.016691 (0.037617)
Moral issues	-.13 (.03)	-0.10706 (0.03756)	-.01 (.03)	0.005590 (0.043022)
Foreign policy	-.14 (.03)	-0.14001	.08 (.04)	-0.066146

issues		(0.04400)		(0.050427)
Time x economic	baseline	baseline	baseline	baseline
Time x civil rights	.02 (.01)	0.00371 (0.01587)	.01 (.01)	-0.008993 (0.017505)
Time x moral	.04 (.02)	0.03296 (0.01874)	.04 (.02)	0.034010 (0.020608)
Time x foreign policy	.00 (.02)	-0.00871 (0.02195)	.00 (.02)	-0.005771 (0.024146)
Residual SD: Intercepts	.06	0.08684	.08	0.10014
Residual SD: Trends	.03	0.03341	.03	0.03804
Residual SD: Data	.04	0.04366	.04	0.04498

Note: The models specified in the authors R-script had correlated slopes and intercepts.
Similar results were obtained with uncorrelated slopes and intercepts.

2.3 Why Scientists Value Replication and the “Replication Crisis”

In science, being able to reproduce researchers findings is vital. “Reproducibility is a core principle of scientific progress (1–6). Scientific claims should not gain credence because of the status or authority of their originator but by the replicability of their supporting evidence.” (Open Science Collaboration) This is a quote from a paper that attempted to replicate 100 psychology papers. “Ninety-seven percent of original studies had statistically significant results. Thirty-six percent of replications had statistically significant results; ...” The failure to replicate studies undermines the legitimacy of scientific research, and this is not just contained to the social sciences. Button et al. 2013 explored the issue of low statistical power in neuroscience research. The authors assert, “Three main problems contribute to producing unreliable findings in studies with low power, even when all other research practices are ideal. They are: the low probability of finding true effects; the low positive predictive value (PPV; see BOX 1 for definitions of key statistical terms) when an effect is claimed; and an exaggerated estimate of the magnitude of the effect when a true effect is discovered.” This is even more troubling given that best practices are not always followed. See 3.2 for the related discussion of threats to statistical conclusion validity.

The authors also discuss another issue which they call “Excess significance.” The authors define this as “... the phenomenon whereby the published literature has an excess of statistically significant results that are due to biases in reporting.” These biases stem from multiple sources.

One source of concern is publication bias. There are multiple types of publication bias. Berinsky et al addressed two distinct types of biases. File drawer bias is research that finds nonsignificant results is more likely to remain unpublished than research that finds significant results.. The second type of bias is gotcha bias, which increases the likelihood of results being published that contradict prior research.

There is evidence “editors and reviewers of psychology journals overwhelmingly recommend against the publication of replications (Neuliep & Crandall, 1990, 1993)” (Koole and Lakens 2012). There are other issues with bias in the peer review process. The process is intended “to vet the quality and feasibility of submitted work.” (Lee et al. 2013) However, this can break down in many systematic ways, introducing bias into the scientific literature.

Bias in research can be construed as a threat to the credibility of scientific evidence, despite its’ “Self-Correcting” framework. (Ioannidis, 2012) There is evidence that there can be negative impacts resulting from improper research, even when that research is later refuted by multiple studies (eg. Taylor et. al 1999, Dales et al 2001, Mäkelä et al 2002), as in the case of the measles-mumps-rubella vaccine where, “ the uptake of measles-mumps-rubella vaccine in Britain declined, resulting in a 25 fold increases in measles in the 10 year period following the original publication.” (Smith and Noble, 2014)

Instruction about research misconduct starting at the undergraduate level has been proposed as a way to combat perverse incentives in scientific research. (Edwards and Roy 2017). “Rather than pretending that the problem of research misconduct does not exist, science and engineering students should receive instruction on these subjects at both the undergraduate and graduate levels.” (Edwards and Roy 2017)

It is this authors’ opinion that this has a natural home in methodology classes. Instruction about how to effectively and ethically conduct research as well as discussing the value of replication should occur while learning about the methods to conduct this research.

Earp and Everett 2015 suggest graduate students be required to conduct replications as a condition for receiving their PhDs. While it makes sense to have graduate students replicating experiments involving human subjects, undergraduate students can still play a role in replication.

Undergraduate students may not be trained in all the appropriate tools and techniques required to perform certain replications as principal investigators, such as replications that involve human subjects in experiments or certain tools involved in experiments in other disciplines. However, there are projects that undergraduate students can work on replicating. In the social sciences, there are studies conducted with survey data, which undergraduate could attempt to replicate.

Undergraduate students can also assist in the lab. Undergraduate research assistantship and other research mentoring has shown other benefits, including increases in: preparation for graduate school, organizational skills, statistical skills, writing skills, library usage skills , as well as benefiting faculty by assisting with time consuming tasks. (Landrum and Nelson, 2002). It has also been found that some groups of students have higher rates of graduation when they are in this type of role. (Gegerman et al, 1998)

In other disciplines as well as social sciences involving experimental data, undergraduate students could use data collected by more experienced researchers to attempt to replicate studies. This would also be beneficial for students, as they would gain valuable skills analyzing the data sets. Attempting to replicate Baldassarri and Gelman 2008 imparted valuable experience with

using R to analyze data as well as quite a bit of knowledge about both regression analysis and multilevel modeling.

Also, performing exact replications is difficult. While attempting to replicate Baldassari and Gelman 2008, there were many non trivial decisions to make with no clear “right” answer. It was necessary to decide how to recode the 50 variables in my analysis. There was also some ‘guess work’ involved in deciding which variables the authors used in their analysis. There were also non-trivial decisions to be made as to how to create the variables that were not contained in the dataset as well as what form to store these variables in R. At the end of this, it is not clear why the replicated models differed from the original authors.

Authors can address the difficulty of replication by keeping track of all the decisions a researcher makes. Write out a document at the beginning of any study, and create an updated version everytime a new decision needs to be made. This could be because it was not foreseen before the study, or because the situation had evolved throughout performing the study that needs to be addressed. By the end, researchers will have a document to provide to anyone who wishes to replicate their study as well as to remember how they performed the study for when the researchers write a paper to submit to journals about the study. Documenting research is important even without taking into account replication!

The DART (Data Access and Research Transparency) initiative is an attempt to collectively push for more accountability and accessibility in empirical research. This initiative began with a committee of the American Political Science Association, and has grown since then. According to “The Journal Editors' Transparency Statement (JETS)”, it’s goals include the availability of all data used, with restrictions for sensitive data; for authors to clearly communicate “the analytic procedures” employed, as well as “access to all relevant analytic materials ... to the greatest extent possible”; ensuring that authors, “identify a dataset’s author(s), title, date, version, and a persistent identifier.”.

This goals of the DART initiative are relevant to the problems encountered in this replication. The exact datafile was no longer available via the ANES; it had to come from another source. It is presumed to be the same data; however, it could not be verified that precisely the same dataset was found. Also, the script used to clean the data and recode the variables was not part of the replication files. This may explain some of the differences between the replicated models and the authors’ models.

3. Original Research

There were several objectives for the original research. These are broken down into the following sections: an examination of why the question of Political Polarization matters, an examination of the Statistical Conclusion Validity, the exploration of a Hypothesis about a relationship seen in the replicated paper, an examination of the External Validity as well as a discussion of these results. These are related concepts and at times the explorations overlap.

3.1 Why Political Polarization Matters

The founding fathers thought about the ways in which problems occur in a free society. While much has changed over the centuries, their concerns about ‘factions’ remain relevant.

In Federalist 10, Madison explores “the violence of faction.” Madison explains a faction is the citizenry “actuated by some common impulse of passion, or of interest, adverse to the rights of other citizens, or to the permanent and aggregate interests of the community.” In other words, this is a group of citizens who collectively attempt to disregard the rights of other people, or to disregard the long term well-being of the community as a whole.

Madison proposed that a large republic would decrease the likelihood of a faction having the necessary support to overwhelm other factions. The most damaging potential issue of factions would come about when one group of voters were able to elect leaders that would have both the power and the will to pursue an “improper or wicked project”.

This concern about factions is relevant in the case of Political Polarization. If the Political Parties become so antagonistic that they could utilize the legal process to pursue an “improper or wicked project”. So long as the people tend to subdivide into different groups this is always a concern for the United States

It does not seem that this tendency will rectify itself anytime soon. It has been suggested that “differential intergroup behavior” can be induced by merely sorting individuals into groups. (eg. Tajfel et al 1971)

Research was conducted in the 1960’s on ‘cleavages’ in many different Political Systems. What the authors called cleavages are conceptually similar to factions. The current “cleavages” among the electorate depends upon the historical cleavages in the electorate. Long-term, “cross-cutting cleavages” facilitate political stability. (Lipset and Rokkan 1967, Karvonen and Kuhnle 2001) Social instability arises from the alignment of these cleavages into non-overlapping groups. James Madison called cleavages factions.

Increasing partisan constraint as well as the effects that come along with that seems to be here to stay. Since it seems to have been clear since at least the Eighteenth Century that people tend to sort themselves into groups, perhaps a more relevant question than “Is Political Polarization occurring in the United States?” is the question “What can be done to mitigate the negative consequences associated with people segmenting into ‘Factions’?”

This seems to be the question the Federalist Papers were discussing instead of whether factions would form in a free society. Segmenting into these groups is human nature and humans

will form these groups whenever they are free to do so. That does not seem likely to change any time soon. Whether this tendency is good or bad depends not just on how we define good and bad. It also depends on how members of the group choose to behave. Further Polarization research could look at group behavior in an effort to understand more about the conditions under which “good and bad” behavior are elicited.

3.2 Statistical Conclusion Validity of the Model

Statistical Conclusion Validity involves testing whether or not covariation between the variables exists as well as estimating the strength of this covariation. (Cook et al. 2002) Using frequentists statistical methods, these are commonly assessed via a hypothesis test and reporting a point estimate. Issues surrounding the reporting of point estimates in the Null Hypothesis Significance Testing (NHST) framework has been discussed for quite a few years.

“David Bakan said back in 1966 that his claim that "a great deal of mischief has been associated" with the test of significance "is hardly original," that it is "what 'everybody knows,'" and that "to say it 'out loud' is . . . to assume the role of the child who pointed out that the emperor was really outfitted in his underwear" (p. 423). If it was hardly original in 1966, it can hardly be original now” (Cohen, 1994)

One proposed way to address these issues is reporting confidence intervals instead of reporting point estimates. (Krantz 1999, Cook et al. 2002) This would give clearer information to other researchers, including those attempting replications. This would also shift the focus away from arbitrary p-value cutoffs. (Gardner and Altman, 1986)

The focus would be shifted away from testing a hypothesis and towards a statistical summary of the result of the empirical inquiry. Reporting confidence intervals could also shift some of the focus of replication towards looking for a reasonable range of effect sizes of a phenomenon, instead of rejecting the null hypothesis a second time.

There are also various threats to Statistical Conclusion Validity. Baldassari and Gelman used frequentist Multilevel Modeling to model Political Polarization. A possible threat to statistical Conclusion Validity is violation of the models Assumptions. Loy 2013 is a detailed explanation of model diagnostics for Hierarchical Models. This is beyond the scope of this paper and is a potential area of future inquiry into the Statistical Conclusion Validity.

Other threats to statistical conclusion validity exist. Gelman and Loken 2013 discuss the issue of “researcher degrees of freedom”. This is when a researcher decides upon one model, but this model is determined based upon the researchers evaluation of which model is most likely to fit after looking at the data. This is as opposed to testing a hypothesis derived from the theory behind the empirical study. (Gelman and Loken, 2013; Gelman and Loken, 2016)

Simulations of researcher degrees of freedom have found that researchers can inflate the false error rate to over 60%. (Simmons et al. 2011) The authors posit six “Requirements for authors” and four “Guidelines for reviewers” aimed at addressing researcher degrees of freedom.

One source of the improper analyses be a misunderstanding of the difference between exploratory and confirmatory data analysis. While NHST is an example of the later, the failure to differentiate between the two could lead to the issues from Gelman and Loken. For a more

thorough discussion of the difference between exploratory and confirmatory data analysis see Tukey 1980.

For practical reasons, linear regressions were fit and the residuals were analysed. It is hoped that the Multilevel Models would be no worse than an OLS regression with regards to the issues that residual analysis would discern.

One assumption of both linear regression models and varying slope varying intercept models is normally distributed errors. Quantile-quantile plots show that the assumption of normally distributed errors is violated. Maas and Hox 2004 performed simulations to explore violations of this assumption in multilevel models.

Another assumption is the identically independently distributed errors. If the regression model is specified correctly, then plotting the time on the X axis and the marginal residuals on the Y axis is a useful diagnostic for potentially autocorrelated data. (Fraccaro et al 2000)

3.3 Hypothesis Exploration

Exploring this hypothesis, the latest ANES cumulative data file was used. This is the 1948-2012 cumulative data file. The replication models were fit to this larger data set and similar results were found. The R-script for the replication models was largely reused for this analysis with minor changes to read in the 1948-2012 datafile instead of the 1948-2004 datafile.

Baldassarri and Gelman fit additional models in their paper not replicated and displayed these in Table 3. The authors found no significant results for their models. These models were about the correlation of issues with other issues. The authors assert that this constraint would be a serious concern for Polarization. The models did not model the issues directly, instead focusing on types of issues or whether or not the issues were of the same type of issue.

Instead, this paper models the change over time of the correlation between issue $i \times$ issue j in year t for $i, j \in \{1, 2, \dots, 46, 47\}$, $t \in \{1972, 1974, \dots, 2002, 2004, 2008, 2012\}$, and $i \neq j$.

For each year t , a correlation matrix was created for all pairs of issues i and j . Since the ANES did not ask all questions in all years there were some coefficients that could not be computed. Since the correlation matrix is symmetric, ie $\text{corr}(\text{issue } j, \text{issue } i) = \text{corr}(\text{issue } i, \text{issue } j)$, only one half was retained for the analysis.

Also, this is related to the reason that i cannot equal j ; this would have a correlation of one even when there were no relationships between variables. The interpretation of correlation coefficients as the inner product of two vectors v, w divided by the vectors Euclidean Norm feels like a particularly intuitive way of explaining this. (Gniazdowski 2013)

$$\rho_{v,w} = \cos(v, w) = \frac{v \cdot w}{\|v\| \|w\|}$$

When the v and w are the same vectors, i.e. when $i = j$ the angle between the vectors, $\theta = 0$ and thus $\cos(\theta) = 1$. For more about the correlation coefficient see Lee Rodgers and Nicewander 1988 as well as Gniazdowski 2013.

The model, stated in a formal equation,

$$y_{ijt} = \alpha_{ij} + \beta_{ij}t + \varepsilon_{ijt}$$

Where y_{ijt} is the correlation between issue i and issue j in year t . This is a varying slope varying intercept model where the issues are the second-level units, allowing both the mean correlation and the time trend to vary by issues i and j . This model is similar to the model from Table One in Baldassarri and Gelman 2008.

A Wald Test was also performed on this model. The Wald test found the time trend significant. The χ^2 statistic was 10.097 with 1 degree of freedom which has a p-value of 0.001485. Again, caution should be used in interpreting this as issues have been found with the use of Wald Tests with Multilevel Models.

In order to rule out the possibility of different results being obtained by using a different dataset, the models were fit to the 1948-2004 datafile and similar results were found. Also, note that the recommended procedure of fitting regression models before fitting the multilevel model was performed. (Gelman and Hill, 2007) Both no pooling and complete pooling regressions were fit. The details of those models can be provided upon request.

Table 3	Issues $i \times j$ model output
Intercept	Estimate (Std. Error) {t value} 0.167956 (0.018124) {9.267}
Time (decades)	0.015196 (0.004782) {3.178}
Residual SD : Intercept issue j	0.10018
Residual SD :Trends issue j	0.02308
Residual SD : Intercept issue i	0.06538
Residual SD :Trends issue i	0.01356
Residual SD : Data	0.08210

3.4 External Validity

The External Validity of a study is whether the results can be generalized beyond the specific circumstances of the study. Replication of the results of Baldassarri and Gelman 2008 on the same datafile enhances the confidence in the results because it is evidence that the relationships can be reproduced by another researcher. Also, similar findings with the models fit to the 1948-2012 datafile enhance the confidence in the External Validity of the results. These results were

generalized to a different time frame as well as to a potentially different set of variables. The latter would depend upon whether the exact same variables were chosen from the ANES codebook.

The conclusions for the models of the correlation of issue x issue however are called into question by my results. It is possible that modeling the issues instead of “within and between issue domains” or the “types of issue domains” lead to different results. (Baldassarri and Gelman 2008)

3.5 Discussion of Results

The United States is becoming Polarized. Narratives seem to exist supporting two different interpretations, one for both sides. “Clickbait”, “Fake News” and other low quality reporting is seen confirming the suspicions of the ‘other’.

The Motivated Reasoning literature has discussed when and why people are subject to biases that lead to accepting information that is congruent to their prior beliefs and disconfirmation of information that is not congruent with their previously held beliefs. (e.g. Taber and Lodge 2006) This is important because when people believe the ‘other side’ is ‘bad’, unsubstantiated stories can spread without effective evaluation.

When the “other” is seen as dangerous to the functioning of our Democracy, it is possible that one side will pursue an “improper or wicked project” believing it is necessary for the survival of our Democracy or its’ ability to function properly. (Federalist 10)

However that is merely the existential threat to the Democratic process. It is possible that this decreases our ability to efficiently and effectively pursue long term plans. Investments such as infrastructure, education, social services, as well as economic and fiscal plans being changed frequently could affect the United States. Exploring better ways to accomplish our objectives can be useful; however, it is hard to argue that being indecisive leads to better outcomes.

When policy discussions are replaced by discussions about the evils of the “other”, it seems likely that ideas will be rejected or accepted without effective consideration because it comes from the “other side” or “our side”.

References

Abramowitz, Alan I., and Kyle L. Saunders. "Ideological realignment in the US electorate." *The Journal of Politics* 60, no. 3 (1998): 634-652.

Abramowitz, Alan, and Kyle Saunders. "Why can't we all just get along? The reality of a polarized America." In *The Forum*, vol. 3, no. 2. De Gruyter, 2005.

Abramowitz, Alan I., and Walter J. Stone. "The Bush effect: Polarization, turnout, and activism in the 2004 presidential election." *Presidential Studies Quarterly* 36, no. 2 (2006): 141-154.

Abramowitz, Alan I., and Kyle L. Saunders. "Is polarization a myth?." *The Journal of Politics* 70, no. 2 (2008): 542-555.

Abramowitz, Alan I., and Morris P. Fiorina. "Polarized or Sorted? Just What's Wrong with our Politics, Anyway." *The American Interest* (2013).

Bakker, Marjan, Annette van Dijk, and Jelte M. Wicherts. "The rules of the game called psychological science." *Perspectives on Psychological Science* 7, no. 6 (2012): 543-554.

Baldassarri, Delia, and Andrew Gelman. "Partisans without constraint: Political polarization and trends in American public opinion." *American Journal of Sociology* 114, no. 2 (2008): 408-446.

Berinsky, Adam, James N. Druckman, and Teppei Yamamoto. "Why Replications Do Not Fix the File Drawer Problem: A Model and Evidence from a Large-Scale Vignette Experiment." (2017).

Berkhof, Johannes, and Tom AB Snijders. "Variance component testing in multilevel models." *Journal of Educational and Behavioral Statistics* 26, no. 2 (2001): 133-152.

Button, Katherine S., John PA Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma SJ Robinson, and Marcus R. Munafò. "Power failure: why small sample size undermines the reliability of neuroscience." *Nature Reviews Neuroscience* 14, no. 5 (2013): 365.

Bones, Arina K. "We knew the future all along: Scientific hypothesizing is much more accurate than other forms of precognition—A satire in one part." *Perspectives on Psychological Science* 7, no. 3 (2012): 307-309.

Box, George EP. "Science and statistics." *Journal of the American Statistical Association* 71, no. 356 (1976): 791-799.

Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O'donoghue, and Matthew Rabin. "Regulation for Conservatives: Behavioral Economics and the Case for" Asymmetric Paternalism". *University of Pennsylvania law review* 151, no. 3 (2003): 1211-1254.

Cohen, Jacob. "The earth is round ($p < .05$)." *American psychologist* 49, no. 12 (1994): 997.

Cook, Thomas D., Donald Thomas Campbell, and William Shadish. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin, 2002.

Dales, Loring, Sandra Jo Hammer, and Natalie J. Smith. "Time trends in autism and in MMR immunization coverage in California." *Jama* 285, no. 9 (2001): 1183-1185.

Data Access and Research Transparency <https://www.dartstatement.org/>

De Jonge, Edwin, and Mark van der Loo. "An introduction to data cleaning with R." Heerlen: Statistics Netherlands (2013).

Diekmann, Andreas. "Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data." *Journal of Applied Statistics* 34, no. 3 (2007): 321-329.

Diermeier, Daniel, and Christopher Li. "Partisan Affect and Elite Polarization." (2017).

Edwards, Marc A., and Siddhartha Roy. "Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition." *Environmental Engineering Science* 34, no. 1 (2017): 51-61.

Earp, Brian D., and Jim AC Everett. "How to fix psychology's replication crisis." *The Chronicle of Higher Education* (2015).

Ellis, Christopher, and James A. Stimson. *Ideology in America*. Cambridge University Press, 2012.

Fiorina, Morris P., Samuel J. Abrams, and Jeremy Pope. *Culture war?: The myth of a polarized America*. Longman Publishing Group, 2006.

Fiorina, Morris P., and Samuel J. Abrams. "Political polarization in the American public." *Annu. Rev. Polit. Sci.* 11 (2008): 563-588.

Fiorina, Morris P., Samuel A. Abrams, and Jeremy C. Pope. "Polarization in the American public: Misconceptions and misreadings." *The Journal of Politics* 70, no. 2 (2008): 556-560.

Fisher, Ronald A., and Winifred A. Mackenzie. "Studies in crop variation. II. The manurial response of different potato varieties." *The Journal of Agricultural Science* 13, no. 3 (1923): 311-320.

Fisher, Ronald Aylmer. "Theory of statistical estimation." In *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 22, no. 5, pp. 700-725. Cambridge University Press, 1925.

Fisher, Ronald A. "The logic of inductive inference." *Journal of the Royal Statistical Society* 98, no. 1 (1935): 39-82.

Franklin, Charles H. "Measurement and the dynamics of party identification." *Political Behavior* 14, no. 3 (1992): 297-309.

Gardner, Martin J., and Douglas G. Altman. "Confidence intervals rather than P values: estimation rather than hypothesis testing." *Br Med J (Clin Res Ed)* 292, no. 6522 (1986): 746-750.

Gelman, Andrew, and Jennifer Hill. (2007) *Data analysis using regression and multilevel/hierarchical models*. Vol. 1. New York, NY, USA: Cambridge University Press, 2007.

Gelman, Andrew. "Induction and deduction in Bayesian data analysis." (2011).

Gelman, Andrew, and Eric Loken. "The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time." Department of Statistics, Columbia University (2013).

Gelman, Andrew, and Eric Loken. "The statistical crisis in science." *The Best Writing on Mathematics* 2015 (2016): 305.

Gniazdowski, Zenon. "Geometric interpretation of a correlation." *Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki* 9 (2013): 27-35.

Gregerman, Sandra R., Jennifer S. Lerner, William von Hippel, John Jonides, and Biren A. Nagda. "Undergraduate student-faculty research partnerships affect student retention." *The Review of Higher Education* 22, no. 1 (1998): 55-72.

Hox, Joop. "Multilevel modeling: When and why." In *Classification, data analysis, and data highways*, pp. 147-154. Springer, Berlin, Heidelberg, 1998.

Ioannidis, John PA. "Why science is not necessarily self-correcting." *Perspectives on Psychological Science* 7, no. 6 (2012): 645-654.

Karvonen, Lauri, and Stein Kuhnle, eds. *Party systems and voter alignments revisited*. Routledge, 2001.

Koole, Sander L., and Daniël Lakens. "Rewarding replications: A sure and simple way to improve psychological science." *Perspectives on Psychological Science* 7, no. 6 (2012): 608-614.

Krantz, David H. "The null hypothesis testing controversy in psychology." *Journal of the American Statistical Association* 94, no. 448 (1999): 1372-1381.

Landrum, R. Eric, and Lisa R. Nelsen. "The undergraduate research assistantship: An analysis of the benefits." *Teaching of Psychology* 29, no. 1 (2002): 15-19.

Lee, Carole J., Cassidy R. Sugimoto, Guo Zhang, and Blaise Cronin. "Bias in peer review." *Journal of the Association for Information Science and Technology* 64, no. 1 (2013): 2-17.

Lee Rodgers, Joseph, and W. Alan Nicewander. "Thirteen ways to look at the correlation coefficient." *The American Statistician* 42, no. 1 (1988): 59-66.

Lelkes, Yphtach. "Mass polarization: Manifestations and measurements." *Public Opinion Quarterly* 80, no. S1 (2016): 392-410.

Lipset, Seymour and Stein Rokkan. "Cleavage Structures, Party Systems, and Voter Alignments." *Party Systems and Voter Alignments: Cross-National Perspectives* (New York, The Free Press, pp. 1-64) (1967).

Mäkelä, Annamari, J. Pekka Nuorti, and Heikki Peltola. "Neurologic disorders after measles-mumps-rubella vaccination." *Pediatrics* 110, no. 5 (2002): 957-963.

Neuliep, James W. "Editorial bias against replication research." *Journal of Social Behavior and Personality* 5, no. 4 (1990): 85.

Neuliep, James W., and Rick Crandall. "Reviewer bias against replication research." *Journal of Social Behavior and Personality* 8, no. 6 (1993): 21.

Neuliep, James W., and Rick Crandall. "Everyone was wrong: There are lots of replications out there." *Journal of Social Behavior and Personality* 8, no. 6 (1993): 1.

Open Science Collaboration. "Estimating the reproducibility of psychological science." *Science* 349, no. 6251 (2015): aac4716.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant." *Psychological science* 22, no. 11 (2011): 1359-1366.

Smith, Joanna, and Helen Noble. "Bias in research." *Evidence-based nursing* (2014): ebnurs-2014.

Taber, Charles S., and Milton Lodge. "Motivated skepticism in the evaluation of political beliefs." *American Journal of Political Science* 50, no. 3 (2006): 755-769.

Tajfel, Henri, Michael G. Billig, Robert P. Bundy, and Claude Flament. "Social categorization and intergroup behaviour." *European journal of social psychology* 1, no. 2 (1971): 149-178.

Tajfel, Henri, and John C. Turner. "An integrative theory of intergroup conflict." *The social psychology of intergroup relations* 33, no. 47 (1979): 74.

Taylor, Brent, Elizabeth Miller, CPaddy Farrington, Maria-Christina Petropoulos, Isabelle Favot-Mayaud, Jun Li, and Pauline A. Waight. "Autism and measles, mumps, and rubella vaccine: no epidemiological evidence for a causal association." *The Lancet* 353, no. 9169 (1999): 2026-2029.

Tukey, John W. "We need both exploratory and confirmatory." *The American Statistician* 34, no. 1 (1980): 23-25.

Webster, Steven W. and Abramowitz, Alan I., (2017) *The Ideological Foundations of Affective Polarization in the U.S. Electorate*, *American Politics Research* 207 Vol. 45(4) 621-647
<https://ase.tufts.edu/gsc/gradresources/guidetomixedmodelsinr/mixed%20model%20guide.html>