Ideological Extremification in the American Media Diet*

Matthew E. Dardet† Ben TerMaat‡

11 December 2021

Abstract

Replicating Guess (2021) and Tyler, Grimmer, and Iyengar (2021), we find substantial overlap in the online media diets of Democrats and Republicans in 2015 and moderately less overlap in 2016, important findings given the contemporary concern over polarization driven by media echo chambers on the Internet. However, we also find that models of media consumption that account for partisanship alone without considering ideology exhibit signs of model dependence. Including ideology in these models reduces model dependence and explains greater variance in Americans' media diets. Futher, stratifying respondents by ideological type (i.e., if they are ideologues, near ideologues, moderate partisans, true non-partisan moderates, or have no issue content at all in their beliefs) reveals that more ideological respondents are exhibiting less overlap in their online media consumption and are moving closer to the echo chamber model of media consumption. Finally, we present evidence that different methods for classifying news articles from URLs and different survey samples cause differing results in the two papers.

Introduction

One of the most important agendas for modern political science research involves quantifying and elucidating the causes of apparent increases in common measures and varieties of political polarization. In recent years, the literature has benefited from numerous studies that examine the relationship between heightened political polarization and digital technologies. However, stark disagreement remains regarding the fundamental question of whether Internet media and news contribute to political polarization. While some scholars conclude that social media

*Replication files are available at https://github.com/matthewdardet/GOV2001ReplicationPaper (GitHub) or https://doi.org/10.7910/DVN/AGIAZ4 (Dataverse). We thank Soubhik Barari, Chris Kenny, Gary King, and members of Ryan Enos' research group for their useful insights.

†Department of Government, Harvard University, matthewdardet@fas.harvard.edu

‡Department of Government, Harvard University, btermaat@g.harvard.edu

has created polarizing echo chambers (Sunstein, 2001, 2017) or information silos (Pariser, 2011; Shmargad & Klar, 2020), others argue that such concerns are overblown (Boxell et al., 2017; Guess, 2021).

In this study, we replicate two papers on opposing sides of this online media-driven polarization debate. The first paper, Guess (2021) "(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets," finds that "most people across the political spectrum have relatively moderate media diets, about a quarter of which consist of mainstream news websites and portals," (p. 1007). Guess (2021) reaches this finding by fielding opt-in YouGov Pulse surveys of respondents in 2015 ($N = 1,392$) and 2016 ($N = 2,512$), asking these respondents about their demographic characteristics, party identification, and ideological leanings. These respondents then agree to download a tool called the Wakoopa browser bar, which subsequently tracks and records all of the URLs that they view online.

Guess (2021) employs a machine learning algorithm, logistic regression classification, to determine which URLs constitute news and which constitute non-news browsing behavior. Using a mapping of URLs to ideological slants developed by Bakshy et al. (2015), a mapping which relies on the self-reported ideologies of the Facebook users who visit these sites, Guess (2021) computes a simple mean ideological slant for the entire media diet (comprising all of the URLs visited) of each respondent. Ultimately, by calculating a measure of distributional overlap that we describe in the analysis below, Guess (2021) is able to determine the degree to which media diets of Republicans, Democrats, and Independents are overlapping (in other words, "similar" or "moderate").

The second paper, Tyler, Grimmer, & Iyengar (2021), follows a very similar analytical scheme but comes to the conclusion that Democrats and Republicans are more siloed in their media consumption habits than were those in Guess' (2021) findings. That is, Democrats and Republicans respectively consume left and right-leaning news articles on moderate and mainstream media sources.

Two key differences exist in the data acquisition and methodological strategies of each

study. The first difference is one of survey construction and sampling. Whereas Guess (2021) uses the same YouGov Pulse web browsing panel in 2015 and 2016, Tyler, Grimmer, & Iyengar (2021) fields a different YouGov panel survey in 2016 of $N = 7704$ respondents, out of which 1076 or 14% of respondents agreed to download the Wakoopa browsing history tracker.

The second difference exists in how each paper classifies URL visits as "news" or as "not news." Whereas Guess (2021) learns a logistic regression classifier on document-feature matrices from a training set of news articles pre-coded as news/not news, Tyler, Grimmer, & Iyengar (2021) uses URL section-name proxies to get at the same classification. The classification procedure in Tyler, Grimmer, & Iyengar (2021) involves subsetting URLs to their second-level domain names (e.g, wikipedia.org, foxnews.com, etc.) and removing all of the URLs that are not on a pre-specified list of popular news domains. Then, the group focuses on explicitly U.S.-election-related news content within these news sites using a pre-specified vector of terms related to the 2016 election (e.g., Trump, Clinton, politics, Congress).

In this replication paper, we seek to analyze these disparate URL classification strategies in Guess (2021) and Tyler, Grimmer, & Iyengar (2021) as well as the overall models used in each paper to predict Americans' media consumption. Along the way, we introduce a simple "bellwether" statistic for detecting reductions in the magnitude of model dependence in parametric linear regression models, the mean absolute uncertainty deviation (MAUD) and its standardized version (SMAUD).

Ultimately, we come to three main findings. First, models that account for ideology in addition to party identification appear to be less model-dependent (when ideology is being properly captured by the URL classification strategy) than models that consider party identification alone. This finding has implications for the study of polarization as manifested in media diet consumption patterns. Should political scientists care more about the overall diets of Democrats and Republicans, or should they care more about the diets of ideologues within society?

The answer may be both, and coincides with a literature in political science that affirms

the distinction between partisanship and ideology. Abramowitz & Saunders (2006) find that partisanship comprises elements of both traditional social identity (i.e., identifying with a party in order to satisfy an emotional attachment to group loyalty) and ideology. Green, Palmquist, and Shickler (2002) assert that party identification is influenced more by identification with social groups as opposed to a rational choice examination of party ideology or policies. They also contend that party loyalties themselves motivated citizens' issue positions, evaluations of political leaders, and vote choices rather than the other way around.

Second, we find that, by stratifying respondents by ideological types similar to those outlined in Converse's (1964) seminal work, "The Nature of Belief Systems in Mass Publics," (which first sorted the mass public into the ideologues, near ideologues, group interest, nature of the times, and no issue content categories), the media diets of more ideological citizens have been becoming much more extreme over time, whereas those of the "moderate majority" have remained largely stable over time.

Third, using simple correlations between ideology and party identification in each study as well as more advanced, machine-learning-based strategies that implement Lei et al.'s (2018) leave-one-covariate-out (LOCO) inference, we show that the URL section name proxy method seems to classify news/not news differently than the logistic regression classification method. This in combination with differential survey samples likely account for the difference in results exhibited by both papers.

Data and Research Design

Data for these analyses are national survey responses connected to URL browsing histories for each respondent generated by the Wakoopa web browsing tracker bar and come from Guess (2021) and Tyler, Grimmer, and Iyengar (2021), respectively. Data for this replication were limited by the fact that the URL data could contain private and sensitive information,

and we were only able to obtain a cleaned version of the 2015 URL browsing history data from Guess (2021).

As overviewed in the introduction, the surveys fielded by the two papers are distinct. The 2015 survey in Guess (2021) contained $N = 1,392$ respondents from the YouGov Pulse panel with $6,319,441$ URL observations collected between February 27 and March 19 of 2015, and the 2016 survey from the same paper was from another YouGov Pulse panel with $N = 2,512$ respondents and $16,984,989$ URL visits collected between October 7 and October 31 of 2016. Tyler, Grimmer, and Iyengar (2021) construct a 2016 YouGov panel (seemingly different from the company's Pulse panel offerings) using 7,704 respondents, out of which $N = 1,076$ or 14% of respondents agreed to download the Wakoopa browsing history tracker that recorded all of the URLs they visited on their computers. One key question this study seeks to shed light on is that of whether it is the different survey composition or the URL classification method itself that could be driving these disparities in results.

In the analysis that follows, we compute the average ideological slant of Americans' media diets as in Guess (2021). Then, we show that the linear models in Guess (2021) suffer from model dependence and that they can be improved by including respondent ideology as a parameter in the main model specifications. Lastly, using LOCO inference, we show that the URL section name proxy method seems to classify news/not news differently than the logistic regression classification method. This in combination with differential survey samples likely account for the difference in results exhibited by both papers.

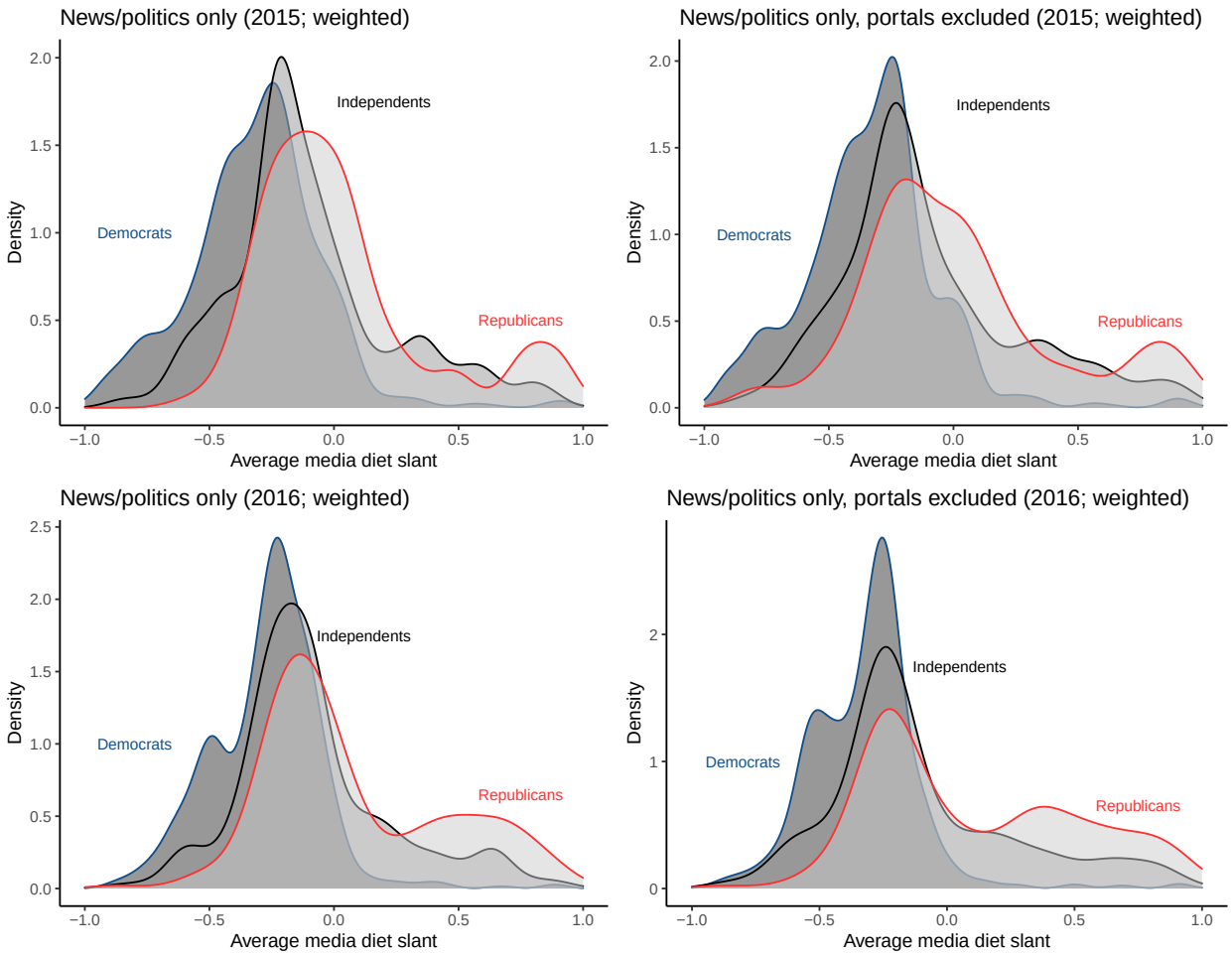
Quantifying the American Media Diet

Are most Americans consuming the same sorts of online media, or are U.S. citizens polarized in their consumption patterns? In **Figure 1**, we use survey and URL browsing history data from Guess (2021) to plot the distributions of Americans' average media diet slants by party identification. Average media diet slant is computed by classifying respondent URL browsing

histories as news/not news using a logistic regression classification algorithm as outlined above, then using the alignment scores from Bakshy et al. (2015) to classify the ideological leaning of a news article visit. Then, a simple mean is taken for each respondent. The figure displays distributions of average media diet slant in the years 2015 and 2016, modified by YouGov survey weights, of respondents using URLs that consist of news and politics articles only as well as those that explicitly remove portal sites like AOL or Yahoo.

Figure 1 shows that most Americans, regardless of party identification, are very much overlapping in the average ideological slant of their media diets. In other words, most Americans of differing party affiliations seem to be mostly moderate and similar in their news consumption, although less distributional overlap occurs in 2016.

Figure 1. Americans' Online Media Diets by Partisanship



Next, we compute the "overlapping coefficient" between the distributions of average Democrat, Republican, and Independent media diet slants. To do so, we use the same formula as Inman and Bradley Jr. (1989), Clemons and Bradley Jr. (2000), and Guess (2021):

$$1 - \frac{1}{2} \int_{-\infty}^{+\infty} \Big| f(x) - g(x) \Big| dx \ . \tag{1}$$

This simple statistic computes the shared area underneath two PDFs as a fraction of the total area. This is feasible due to the fact that any PDF integrated over its support yields a value of 1. In a sense, it quantifies the extent to which two PDFs "agree" with one another.

As shown in **Table 1**, in 2015, Democrats and Republican media diet slant distributions exhibited about 60-63% overlap, Democrats and Independents about 77-83%, and Republicans and Independents about 79-80%. In 2016, Democrat and Republican media diet slant distributions exhibited about 43-48% overlap, Democrats and Independents about 68-73%, and Republicans and Independents about 72-74%. Δ_P and Δ_{NP} capture the change in overlap coefficients between 2015 and 2016 for URLs that include and exclude portal sites, respectively. Notably, every measure of media diet slant change is negative, meaning that Americans of all party identifications became more extreme in their media diets. The extremification of media diets between 2015 and 2016 represents a finding that was relatively glossed over in Guess (2021).

Table 1: **Overlap Coefficients for Americans' Partisan Media Diets**

	2015	2015 (no portals)	2016	2016 (no portals)	Δ_P	Δ_{NP}
Dem-Rep	0.629	0.596	0.478	0.429	-0.150	-0.167
Dem-Ind	0.825	0.770	0.731	0.679	-0.094	-0.090
Rep-Ind	0.790	0.799	0.741	0.720	-0.048	-0.078

Models of Partisan Media Consumption in Guess (2021) Exhibit Model Dependence

Notably, Guess' (2021) main model of media diet consumption (replicated in Appendix A.1) includes respondents' party identification, but not ideology. We are interested in seeing whether these models exhibit symptoms of model dependence and whether the inclusion of respondent ideology could alleviate the model dependence.

To do so, we introduce two simple, descriptive "bellwether" statistics for glimpsing hints of model dependence, the mean absolute uncertainty deviation (MAUD) and the standardized mean absolute uncertainty deviation (SMAUD), defined as follows:

$$\text{MAUD} = \frac{1}{n}\sum_{i=1}^{n} |\sigma_{i_{robust}} - \sigma_{i_{vanilla}}| \ , \tag{2}$$

$$\text{SMAUD} = \frac{1}{n}\sum_{i=1}^{n} \frac{|\sigma_{i_{robust}} - \sigma_{i_{vanilla}}|}{\sigma_{i_{vanilla}}} \ . \tag{3}$$

Here, $\sigma_{i_{robust}}$ and $\sigma_{i_{vanilla}}$ are the coefficient standard errors for parametric linear regression models estimated using the heteroskedasticity-robust estimator and classical regression standard errors, respectively. These estimators take form from the assertion in King & Roberts (2015) that "robust and classical standard errors that differ need to be seen as bright red flags that signal compelling evidence of uncorrected model misspecification," (p. 159).[1]

We also employ the generalized information matrix (GIM) test proposed by King & Roberts (2015), which uses a parametric boostrap-based information matrix test to diagnose model dependence. As Aronow (2016) notes, King and Roberts (2015) is valid only when the researcher specifies a fully parametric regression model. Guess (2021) does exactly that, specifying the linear regression model with finite-dimensional parameterization for media diet

[1]We have not yet attempted to rigorously prove or disprove any properties of these statistics as they pertain to model dependence, so they should be treated with extreme caution and skepticism.

slant, Y_i, as follows:

$$Y_i = \alpha + \beta_1 \text{Democrat}_i + \beta_2 \text{Republican}_i + \beta_3 \text{Independent}_i + \gamma \boldsymbol{X}_i + \epsilon_i \,, \tag{4}$$

where $\boldsymbol{X}_i$ is a "vector of demographic characteristics that includes age, race, gender, family income, and educational attainment," (p. 1016).

When MAUD, SMAUD, and GIM test statistics are computed in **Table 2**, we find clear evidence of model dependence in Guess (2021). In the next section, we attempt to address and alleviate this model dependence using a measure of respondent ideology.

Table 2: **Evidence for Model Dependence in Main Consumption Determinants Models**

	MAUD	SMAUD	(GIM) Rule of Thumb	Test Statistic	p-value
2015	0.008	0.167	1.865	1479.103	0.032
2016	0.018	0.349	1.865	1495.154	0.032

Including Respondent Ideology Reduces Model Dependence in Guess (2021)

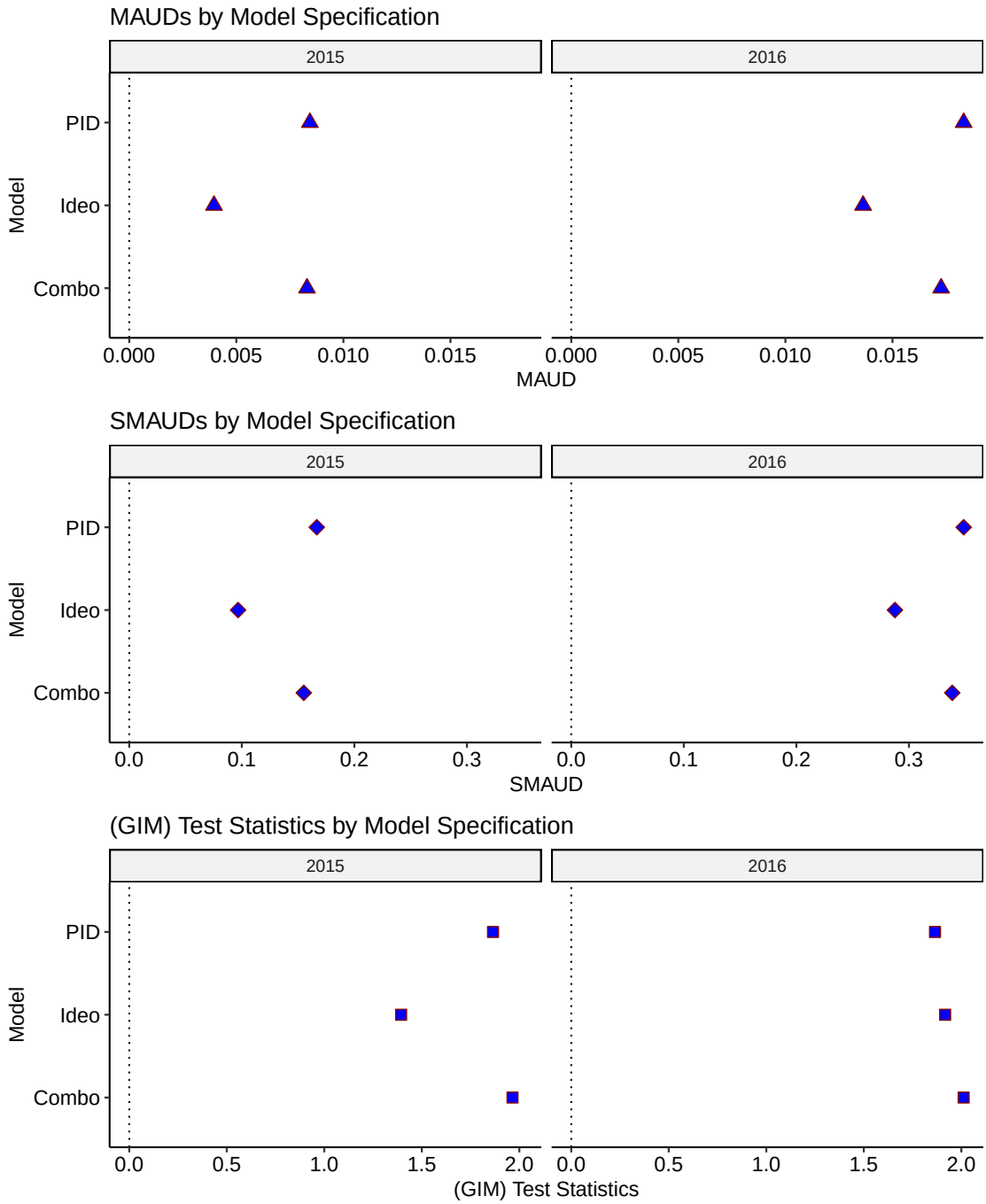
We extend the Guess (2021) parametric linear model of media diet slant (4) by including respondent ideology as follows:

$$\begin{aligned} Y_i = \alpha &+ \beta_1 \text{Democrat}_i + \beta_2 \text{Republican}_i + \beta_3 \text{Independent}_i + \mu_1 \text{Conservative}_i \\ &+ \mu_2 \text{Liberal}_i + \mu_3 \text{Moderate}_i + \mu_4 \text{Very Conservative}_i \\ &+ \mu_5 \text{Very Liberal}_i + \gamma \boldsymbol{X}_i + \epsilon_i \,. \end{aligned} \tag{5}$$

The results of this extended model can be found in Appendix A.2. When ideology is included in the model, the coefficient significance attributed to party identification variables disappears in favor of the ideology variables.

But what of model dependence? **Figure 2** shows that including ideology in models of media diet slant decrease model dependence across all three of the previously outlined statistics for detecting model dependence.[2]

Figure 2. Inclusion of Ideology Reduces Model Dependence



[2]One exception is the GIM Test statistic in 2016. Perhaps the 2016 election made some partisans more ideological in orientation?

Evidence of Ideological Extremification in the American Media Diet

Following the results providing initial evidence for the influence of ideology instead of or in addition to party identification on Americans' media consumption, we investigate how the interpretation of Guess' (2021) findings change if respondents are stratified by ideology rather than partisanship.

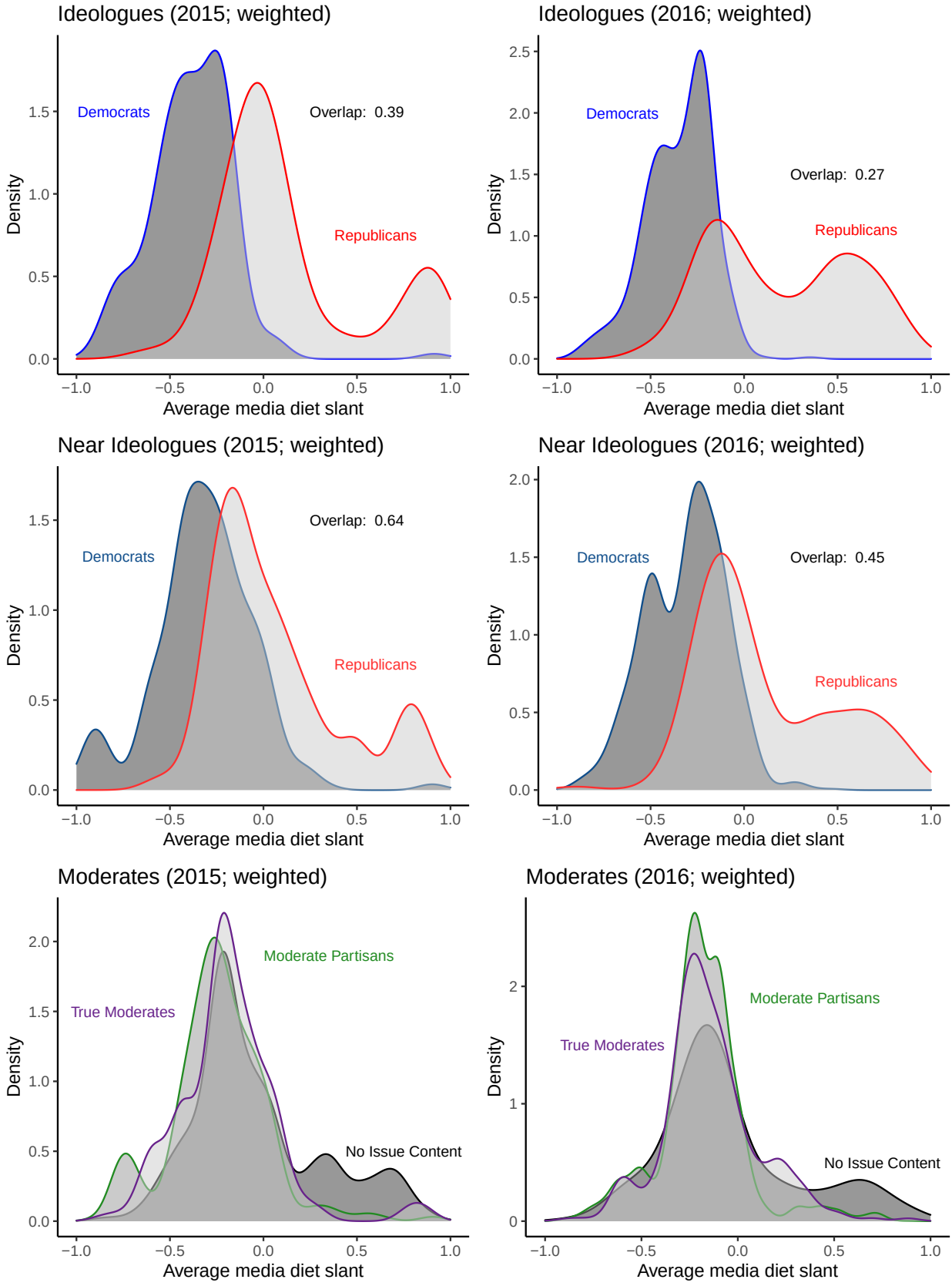
To do so, we follow ideological strata outlined by Converse (1964) in his seminal work, "The Nature of Belief Systems in Mass Publics."[3] We incorporate the spirit of Converse (1964) in dividing respondents into "ideologues" (Republicans who identify as "very conservative" and Democrats who identify as "very liberal"), "near ideologues" (Republicans and Democrats who are respectively "conservative" and "liberal"), "moderate partisans" (Republicans and Democrats who are ideologically "moderate"), "true moderates" (political Independents who are "moderate"), and "no issue content" (people who answered "Don't know" or who refused to respond to questions). When tabulated (Appendix A.3), true moderates, moderate partisans, and no issue content voters vastly outweigh near ideologues and ideologues.

When plotted in **Figure 3**, we notice that respondents who are more ideological in nature consume much more extreme online media along the left-right ideological continuum, with each subclass becoming more extreme in 2016 than 2015.[4] Guess (2021) may have generated results suggesting moderate media diet slants due to the outsize presence of moderates in the survey sample. Indeed, we find that ideology and partisanship have correlations of 0.48 in 2015 and 0.57 in 2016 using the Guess (2021) data.

[3]Converse found that individuals could be categorized as ideologues (those who use ideology in their evaluations of policies and politicians), near ideologues (those who mention items that fall along the liberal-conservative dimension, but place less emphasis on them), group interest voters (those who vote with parties as social identity), nature of the times voters (those who made decisions based on whatever social or economic conditions like the Korean War were ongoing at the time), and no issue content voters (those whose evaluations had no shred of coherence or policy relevance at all).

[4]Ideologues exhibit overlap coefficients between 0.27-0.39 and near ideologues between 0.45-0.64.

Figure 3. Americans' Media Diets by Ideological Type

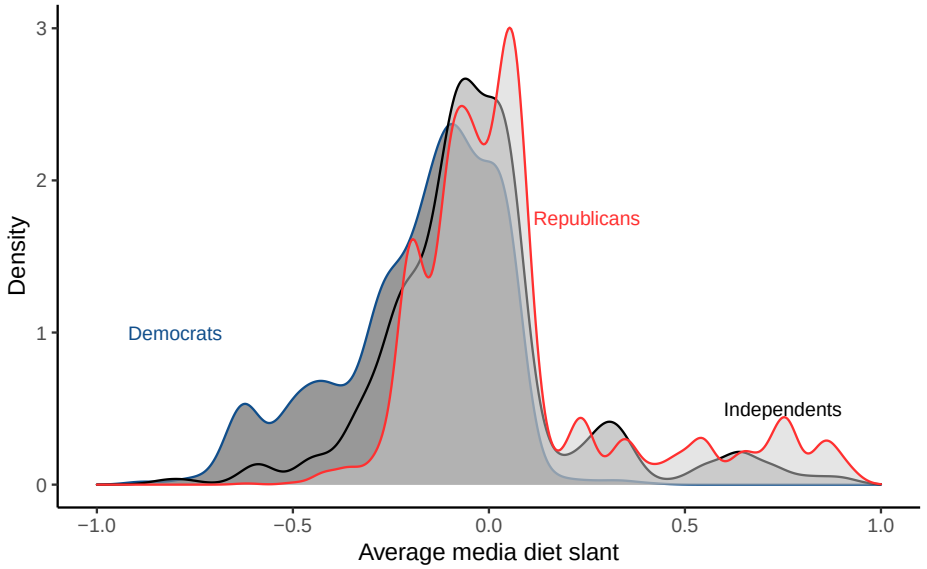


Using URL Section Name Proxies Produces Different Results Than Logistic Classification for News/Not News

Another study on American media diets, Tyler, Grimmer, & Iyengar (2021), differs from Guess (2021) in finding that Republicans and Democrats are not consuming the same types of online media and that, rather, they are consuming mostly media from right and left-leaning news sources, respectively.[5] However, Tyler, Grimmer, & Iyengar (2021) employ a different URL news/not news classification method that Guess (2021) based on URL section name proxies. We outline this procedure in the introduction.

Figure 4 shows the results of applying the URL section name proxy classification method used in Tyler, Grimmer, & Iyengar (2021) to the 2015 web browsing data from Guess (2021) in order to test whether the divergent classification strategies employed in each paper produce comparable results. Here, we see that most Americans have even more overlap in their media consumption patterns and the probability densities appear to be much noisier. In the next section, we investigate whether different classification methods affect variable importance in models of American's media diet slants.

Figure 4. Americans' Online Media Diets (URL Proxy Method)



[5]We replicate the main findings and test model dependence in Tyler, Grimmer, & Iyengar (2021) in Appendix A.4.

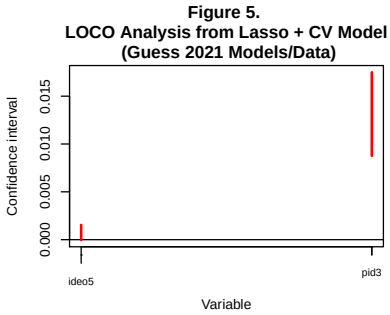
Let's Get LOCO: What's Driving These Disparities in Results?

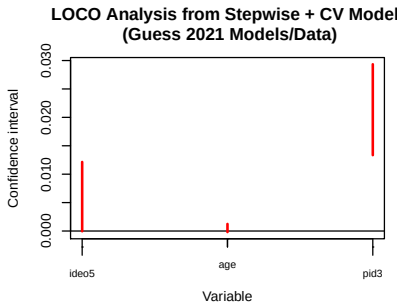
Are the differences between Guess (2021) and Tyler, Grimmer, & Iyengar (2021) due to differential URL news/not news classification methods, underlying differences in survey samples, or something else? Firstly, the correlation between respondents' party identification and ideology in the 2015 data is 0.48 and in 2016 is 0.57 in Guess (2021), whereas the correlation in Tyler, Grimmer, & Iyengar (2021) is much higher at 0.83, suggesting that differences in each paper's survey samples are partially driving disparate results and conclusions in the papers.

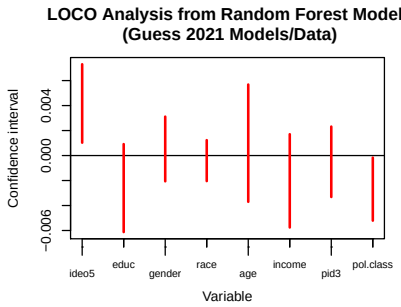
Next, we apply three machine learning algorithms, LASSO, stepwise regression, and random forests within Lei et al.'s (2018) leave-one-covariate-out (LOCO) inference framework[6] to three datasets and models of American media consumption coming from (1) the Guess (2021) data with the logistic regression classification of URLs as news/not news [**Figure 5**], (2) the Tyler, Grimmer, & Iyengar (2021) data with the URL section name proxy classification method [**Figure 6**], and (3) the Guess (2021) data with the URL section name proxy classification method [**Figure 7**]. The red lines in each figure represent a 100% marginal confidence interval on the prediction error that occurs in each model if the specific covariate on the x-axis is removed from the model. In a sense, this is a unique way to get at the importance of each variable in a predictive framework.

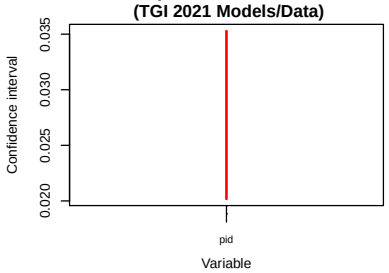
Figures 5-7 display that, in the first dataset, ideology and partisan identification appear to be the most important variables in the prediction model. Partisanship and age are the most important variables in the second prediction model. And ideology, education, gender, and partisanship are the most important variables in the third dataset and model. Ultimately, these results support our intuition that both the different URL classification methods and underlying asymmetries in the survey sample in Guess (2021) and Tyler, Grimmer, & Iyengar (2021) are driving the disparate results in each paper.

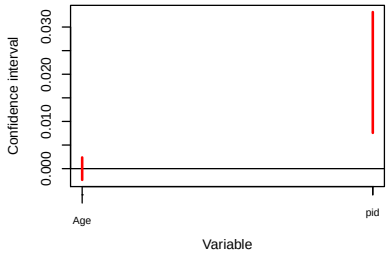
[6]We explain these methodologies in greater detail in Appendix A.6.

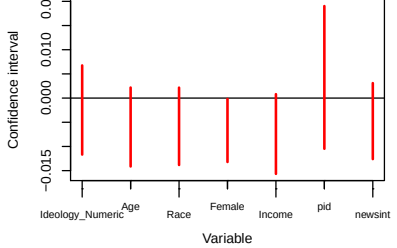


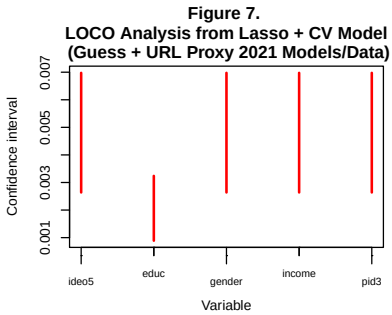


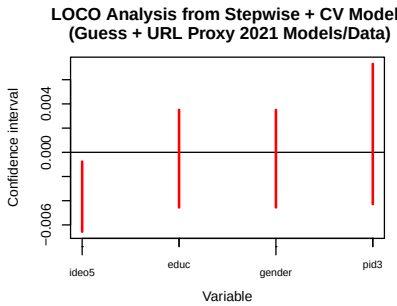


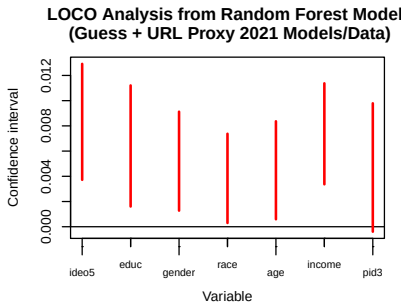












Discussion and Conclusions

Overall, our replication study found that, when stratified by party, most Americans seem to overlap in their media consumption patterns regardless of their affiliation. However, when stratified by ideology, we find that more ideological partisan respondents exhibit much less overlap in their media consumption. The "moderate majority" or moderate partisans, non-ideological individuals, or people with no issue content whatsoever in their beliefs may be driving findings that suggest lack of polarization in the electorate. Finally, we found that two papers making conclusions about American media diets, Guess (2021) and Tyler, Grimmer, & Iyengar (2021), came to different conclusions as a result of *both* different URL classification

methods *and* differences in the underlying survey sample.

There were several limitations to our replication. The first is that we were unable to obtain the 2016 data in Guess (2021) due to privacy restrictions. The second is that all of the data in each year for the two papers come from different surveys with nebulous and potentially differing question wordings and sampling designs. We also did not have a full suite of data between the years 2015-2020 that we could use to determine whether long-term trends in media diet slants point towards ideological extremification in the American mass public.

Another question that this work leaves open is, who are the people who allow their web browsing data to be tracked? Guess (2021) states that the respondents tended to be younger and more Democratic than a representative sample of Americans. Tyler, Grimmer, and Iyengar (2021) note that online panel participants tended to be more active Internet users than the general population. But, just as previous telephone surveys find that their samples tend to have differential rates of older women as respondents, and may be correlated along an underlying psychological dimension of agreeableness, a concern may be that certain panels are including very ideologically motivated partisans that are not representative of the average American partisan. Future studies should investigate whether poststratification weights can really offset differential nonresponse and psychological-attributional groupings among the willing survey sample.

We conclude with several recommendations for researchers who wish to study how Americans consume media online using similar research designs. The first is that study samples may need to be much larger in order to capture the full heterogeneity of Americans' web browsing behavior. The second is that more background characteristics should be collected in these surveys. Party identification and ideology should be measured by a 7-point scale wherein independents and moderates are pressed to provide their party and ideological leanings so as to facilitate greater specificity and precision in model estimation. Further, surveys should include questions asking about political interest and about how many days

during the average week do respondents engage with TV/magazine/internet news in order to triangulate stated versus realized observational patterns. Third, researchers should pay more attention to the composition of the samples in the panel or other survey from which they obtain browsing data; it is unclear whether heterogeneities in browsing data can be rectified simply by applying posstratification weights to unrepresentative data plagued by differential nonresponse. Finally, researchers studying American media diets should ask, which variable is really more important and influential: party, ideology, or their interaction?

Bibliography

Abramowitz, Alan I., and Kyle L. Saunders. 2006. "Exploring the Bases of Partisanship in the American Electorate: Social Identity Vs. Ideology." *Political Research Quarterly* 59 (2): 175–87. https://doi.org/10.1177/106591290605900201.

Aronow, Peter M. 2016. "A Note on "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It"," September. https://arxiv.org/abs/1609.01774v1.

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348 (6239): 1130–32. https://doi.org/10.1126/science.aaa1160.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. 2017. "Greater Internet Use Is Not Associated with Faster Growth in Political Polarization Among US Demographic Groups." *Proceedings of the National Academy of Sciences of the United States of America* 114 (40): 10612–17. http://www.jstor.org/stable/26488105.

Budak, Ceren, Sharad Goel, and Justin M. Rao. 2016. "Fair and Balanced? Quantifying Media Bias Through Crowdsourced Content Analysis." *Public Opinion Quarterly* 80 (S1): 250–71. https://doi.org/10.1093/poq/nfw007.

Clemons, Traci E, and Edwin L Bradley. 2000. "A Nonparametric Measure of the Overlapping Coefficient." *Computational Statistics & Data Analysis* 34 (1): 51–61. https://doi.org/10.1016/S0167-9473(99)00074-2.

Converse, Philip E. 2006. "The Nature of Belief Systems in Mass Publics (1964)." *Critical Review* 18 (1-3): 1–74. https://doi.org/10.1080/08913810608443650.

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence From U.S. Daily Newspapers." *Econometrica* 78 (1): 35–71. https://doi.org/10.3982/ECTA7195.

Green, Donald, Bradley Palmquist, and Eric Schickler. 2002. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. New Haven: Yale University Press.

Groseclose, Tim, and Jeffrey Milyo. 2005. "A Measure of Media Bias." *The Quarterly Journal of Economics* 120 (4): 1191–1237. http://www.jstor.org/stable/25098770.

Guess, Andrew M. 2021. "(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets." *American Journal of Political Science* 65 (4): 1007–22. https://doi.org/10.1111/ajps.12589.

Inman, Henry F., and Edwin L. Bradley. 1989. "The Overlapping Coefficient as a Measure of Agreement Between Probability Distributions and Point Estimation of the Overlap of Two Normal Densities." *Communications in Statistics - Theory and Methods* 18 (10): 3851–74. https://doi.org/10.1080/03610928908830127.

King, Gary, and Margaret E. Roberts. 2015. "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It." *Political Analysis* 23 (2): 159–79. http://www.jstor.org/stable/24572966.

Lei, Jing, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. 2018. "Distribution-Free Predictive Inference for Regression." *Journal of the American Statistical Association* 113 (523): 1094–1111. https://doi.org/10.1080/01621459.2017.1307116.

Lewis, Verlan. 2021. "The Problem of Donald Trump and the Static Spectrum Fallacy." *Party Politics* 27 (4): 605–18. https://doi.org/10.1177/1354068819871673.

Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You.* UK: Penguin.

Prior, Markus. 2009. "Improving Media Effects Research Through Better Measurement of News Exposure." *The Journal of Politics* 71 (3): 893–908. https://doi.org/10.1017/S0022381609090781.

Sunstein, Cass. 2017. *#Republic.* Princeton, NJ: Princeton University Press.

Tyler, Matthew, Justin Grimmer, and Shanto Iyengar. 2021. "Partisan Enclaves and Information Bazaars: Mapping Selective Exposure to Online News." *The Journal of Politics*, August. https://doi.org/10.1086/716950.

Yotam Shmargad & Samara Klar. 2020. "Sorting the News: How Ranking by Popularity Polarizes Our Politics." *Political Communication* 37 (3): 423–46. https://doi.org/10.1080/

10584609.2020.1713267.

Appendix

(A.1) Replication of Main Guess (2021) Model

We replicate the original Guess (2021) media consumption prediction model below.

Table 1: **Determinants of Media Diet Slant (Guess 2021)**

	Average media diet slant (news/politics only)	
	2015	*2016*
	(1)	(2)
Age: 30-44	0.136**	0.062
	(0.034)	(0.035)
Age: 45-59	0.205**	0.123**
	(0.037)	(0.029)
Age: 60+	0.242**	0.203**
	(0.051)	(0.028)
Race: Black	0.010	−0.047
	(0.047)	(0.046)
Race: Hispanic	−0.017	0.110
	(0.054)	(0.066)
Race: White	−0.038	0.021
	(0.037)	(0.040)
Female	−0.022	−0.024
	(0.030)	(0.021)
Income level	−0.002	0.005
	(0.005)	(0.003)
High school	0.019	0.079
	(0.071)	(0.073)
Some college	−0.048	0.055
	(0.062)	(0.071)
College graduate	−0.026	0.017
	(0.068)	(0.071)
Postgraduate	−0.117	−0.006
	(0.076)	(0.072)
Democrat	−0.258**	−0.189**
	(0.060)	(0.044)
Independent	−0.074	−0.017
	(0.062)	(0.046)
Republican	0.040	0.127**
	(0.067)	(0.048)
Constant	−0.093	−0.283**
	(0.075)	(0.081)
N	861	1,903
Adjusted R^2	0.192	0.238

*p < .05; **p < .01; ***p < [.***]
OLS regressions with HC2 robust standard errors in parentheses; YouGov survey data with weights applied.

(A.2) Full Results of Ideology-Extended Models of Media Consumption

We present a full table of the extension of the Guess (2021) media consumption prediction model by varying the inclusion of measures of respondent ideology.

Table 3: **Determinants of Media Diet Slant (Including Ideology)**

	Average media diet slant (news/politics only)					
	2015 (PID)	*2015* (Ideo)	*2015* (Combined)	*2016* (PID)	*2016* (Ideo)	*2016* (Combined)
	(1)	(2)	(3)	(4)	(5)	(6)
Age: 30-44	0.136**	0.080*	0.088**	0.062	0.007	0.020
	(0.034)	(0.033)	(0.033)	(0.035)	(0.033)	(0.034)
Age: 45-59	0.205**	0.126**	0.136**	0.123**	0.068**	0.079**
	(0.037)	(0.035)	(0.035)	(0.029)	(0.026)	(0.027)
Age: 60+	0.242**	0.149**	0.159**	0.203**	0.120**	0.139**
	(0.051)	(0.044)	(0.044)	(0.028)	(0.027)	(0.028)
Race: Black	0.010	−0.044	−0.020	−0.047	−0.056	−0.031
	(0.047)	(0.043)	(0.044)	(0.046)	(0.052)	(0.050)
Race: Hispanic	−0.017	−0.063	−0.061	0.110	0.121*	0.114
	(0.054)	(0.046)	(0.048)	(0.066)	(0.061)	(0.062)
Race: White	−0.038	−0.091**	−0.085*	0.021	0.047	0.033
	(0.037)	(0.032)	(0.034)	(0.040)	(0.039)	(0.040)
Female	−0.022	−0.009	−0.006	−0.024	−0.006	−0.004
	(0.030)	(0.027)	(0.026)	(0.021)	(0.021)	(0.021)
Income level	−0.002	−0.002	−0.002	0.005	0.005	0.004
	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
High school	0.019	0.038	0.035	0.079	0.006	0.034
	(0.071)	(0.067)	(0.070)	(0.073)	(0.061)	(0.066)
Some college	−0.048	−0.031	−0.030	0.055	−0.013	0.011
	(0.062)	(0.063)	(0.067)	(0.071)	(0.059)	(0.064)
College graduate	−0.026	−0.005	−0.002	0.017	−0.020	−0.007
	(0.068)	(0.071)	(0.072)	(0.071)	(0.061)	(0.065)
Postgraduate	−0.117	−0.061	−0.057	−0.006	−0.050	−0.025
	(0.076)	(0.075)	(0.077)	(0.072)	(0.062)	(0.066)
Democrat	−0.258**		−0.106	−0.189**		−0.124*
	(0.060)		(0.059)	(0.044)		(0.050)
Independent	−0.074		−0.020	−0.017		−0.024
	(0.062)		(0.060)	(0.046)		(0.050)
Republican	0.040		−0.013	0.127**		0.027
	(0.067)		(0.070)	(0.048)		(0.054)
Conservative		0.213**	0.214**		0.213**	0.183**
		(0.055)	(0.068)		(0.044)	(0.055)
Liberal		−0.110*	−0.053		−0.127**	−0.061
		(0.047)	(0.060)		(0.043)	(0.050)
Moderate		−0.043	−0.019		0.008	0.027
		(0.042)	(0.055)		(0.038)	(0.046)
Very conservative		0.303**	0.297**		0.308**	0.266**
		(0.057)	(0.063)		(0.059)	(0.065)
Very liberal		−0.208**	−0.144*		−0.165**	−0.099*
		(0.048)	(0.060)		(0.040)	(0.048)
Constant	−0.093	−0.144*	−0.136*	−0.283**	−0.288**	−0.276**
	(0.075)	(0.059)	(0.066)	(0.081)	(0.074)	(0.077)
Party ID	✓		✓	✓		✓
Ideology		✓	✓		✓	✓
N	861	861	861	1,903	1,903	1,903
Adjusted R^2	0.192	0.291	0.299	0.238	0.281	0.302

*p < .05; **p < .01; ***p < [.***]
OLS regressions with HC2 robust standard errors in parentheses; YouGov survey data with weights applied.

(A.3) Tabulation of Respondent Ideological Types Using Guess (2021) Data

Table 4: **Proportion of Respondents in Guess (2021) by Ideological Type**

	Rep Ideologue	Dem Ideologue	Rep Near Ideologue	Dem Near Ideologue	True Moderate	Moderate Partisan	No issue content
2015	7.112	8.980	9.124	11.638	13.649	18.032	31.466
2016	9.116	11.704	11.943	12.898	14.968	16.799	22.572

(A.4) Replication of Tyler, Grimmer, & Iyengar (2021) Results

We replicate the main media diet distributions by partisanship figure using the data from Tyler, Grimmer, & Iyengar (2021) as well as their URL section name proxy method for classifying whether URLs are news or not news in **Figure 8**. **Table 5** displays five different model specifications that include party identification, ideology, both, both and a political interest variable, and both with a political interest and party identification interaction. Finally, **Figure 9** shows measures of model dependence in the Tyler, Grimmer, & Iyengar (2021) models and data. The model containing only party identification and the model that contains party identification, ideology, and the measure of political interest appear to have the lowest model dependence, findings that diverge significantly from the data and models in Guess (2021)—that, respectively, are improved dramatically by the model that includes only ideology—and were proven to be a likely artefact of the survey sampling process and differential classification methodology.

Figure 8. Americans' Online Media Diets (URL Proxy)

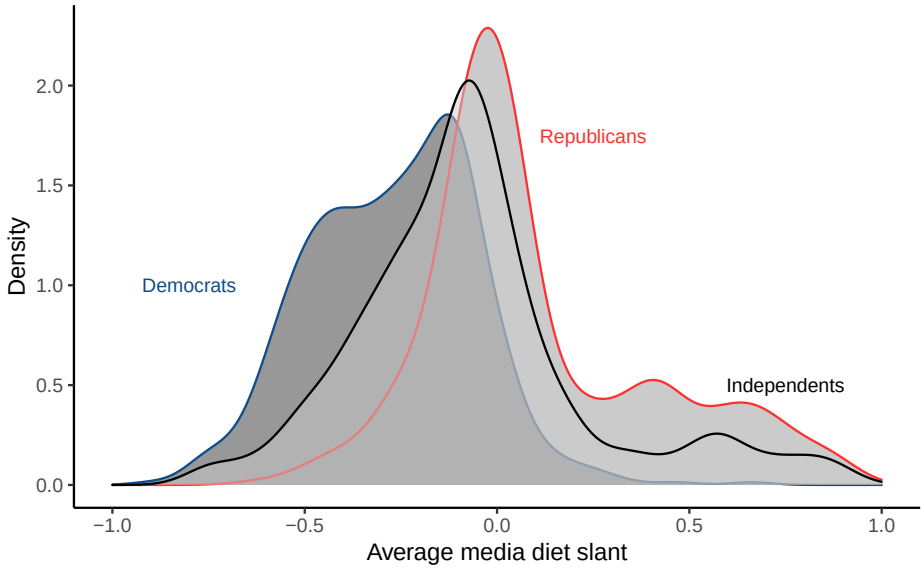
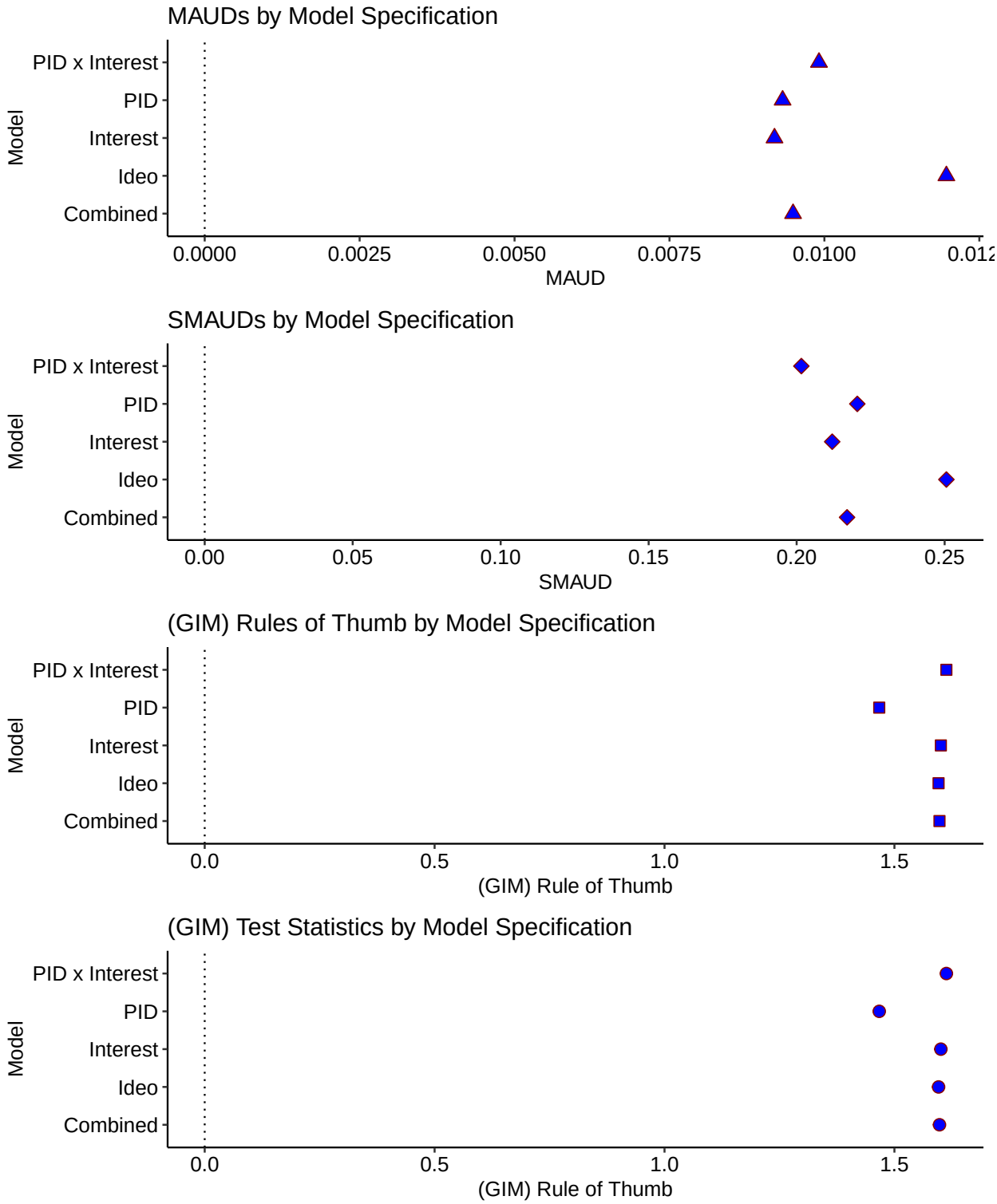


Table 5: **Determinants of Media Diet Slant (Using URL Section Proxies)**

	Average media diet slant (news/politics only)				
	(PID)	(Ideo)	(Combined)	(Interest)	(Party x Interest)
	(1)	(2)	(3)	(4)	(5)
Age: 30-44	0.101**	0.083*	0.090*	0.090*	0.068**
	(0.035)	(0.037)	(0.035)	(0.035)	(0.025)
Age: 45-59	0.122**	0.082*	0.103**	0.104**	0.086**
	(0.034)	(0.036)	(0.035)	(0.036)	(0.025)
Age: 60+	0.177**	0.149**	0.161**	0.162**	0.144**
	(0.035)	(0.037)	(0.036)	(0.037)	(0.025)
Race: White	0.011	0.038	0.017	0.016	0.010
	(0.038)	(0.038)	(0.038)	(0.038)	(0.032)
Race: Black	0.095*	0.093*	0.105*	0.105*	0.104**
	(0.040)	(0.045)	(0.041)	(0.041)	(0.038)
Race: Hispanic	0.053	0.059	0.057	0.056	0.047
	(0.054)	(0.059)	(0.054)	(0.054)	(0.043)
Female	−0.043*	−0.053*	−0.043*	−0.043*	−0.040*
	(0.021)	(0.022)	(0.021)	(0.022)	(0.016)
Income level	0.002	0.002	0.002	0.002	0.002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
High school	−0.129*	−0.156*	−0.142*	−0.142*	−0.130**
	(0.065)	(0.075)	(0.069)	(0.069)	(0.049)
Some college	−0.149*	−0.163*	−0.159*	−0.159*	−0.148**
	(0.062)	(0.070)	(0.065)	(0.065)	(0.048)
College graduate	−0.149*	−0.167*	−0.153*	−0.153*	−0.142**
	(0.064)	(0.074)	(0.067)	(0.067)	(0.050)
Postgraduate	−0.188**	−0.198**	−0.184**	−0.184**	−0.181**
	(0.066)	(0.075)	(0.069)	(0.070)	(0.052)
Democrat	−0.138**		−0.101**	−0.101**	−0.199**
	(0.035)		(0.038)	(0.038)	(0.032)
Republican	0.116**		0.076	0.076	0.055
	(0.035)		(0.042)	(0.042)	(0.033)
Conservative		0.173**	0.098	0.098	0.059
		(0.041)	(0.051)	(0.051)	(0.052)
Liberal		−0.078	−0.020	−0.020	−0.035
		(0.040)	(0.043)	(0.043)	(0.054)
Moderate		0.047	0.047	0.047	0.025
		(0.039)	(0.040)	(0.041)	(0.050)
Very conservative		0.262**	0.163**	0.164**	0.117*
		(0.049)	(0.058)	(0.060)	(0.060)
Very liberal		−0.120*	−0.060	−0.059	−0.061
		(0.047)	(0.050)	(0.051)	(0.059)
Political Interest				−0.002	0.140**
				(0.022)	(0.044)
Democrat x Interest					0.240**
					(0.049)
Republican x Interest					0.069
					(0.050)
Constant	−0.060	−0.124	−0.099	−0.098	−0.150
	(0.079)	(0.084)	(0.083)	(0.084)	(0.078)
Party ID	✓		✓	✓	✓
Ideology		✓	✓	✓	✓
Political Interest				✓	✓
N	920	924	920	920	920
Adjusted R^2	0.245	0.220	0.265	0.264	0.290

*p < .05; **p < .01; ***p < [.***]
OLS regressions with HC2 robust standard errors in parentheses; YouGov survey data with weights applied.

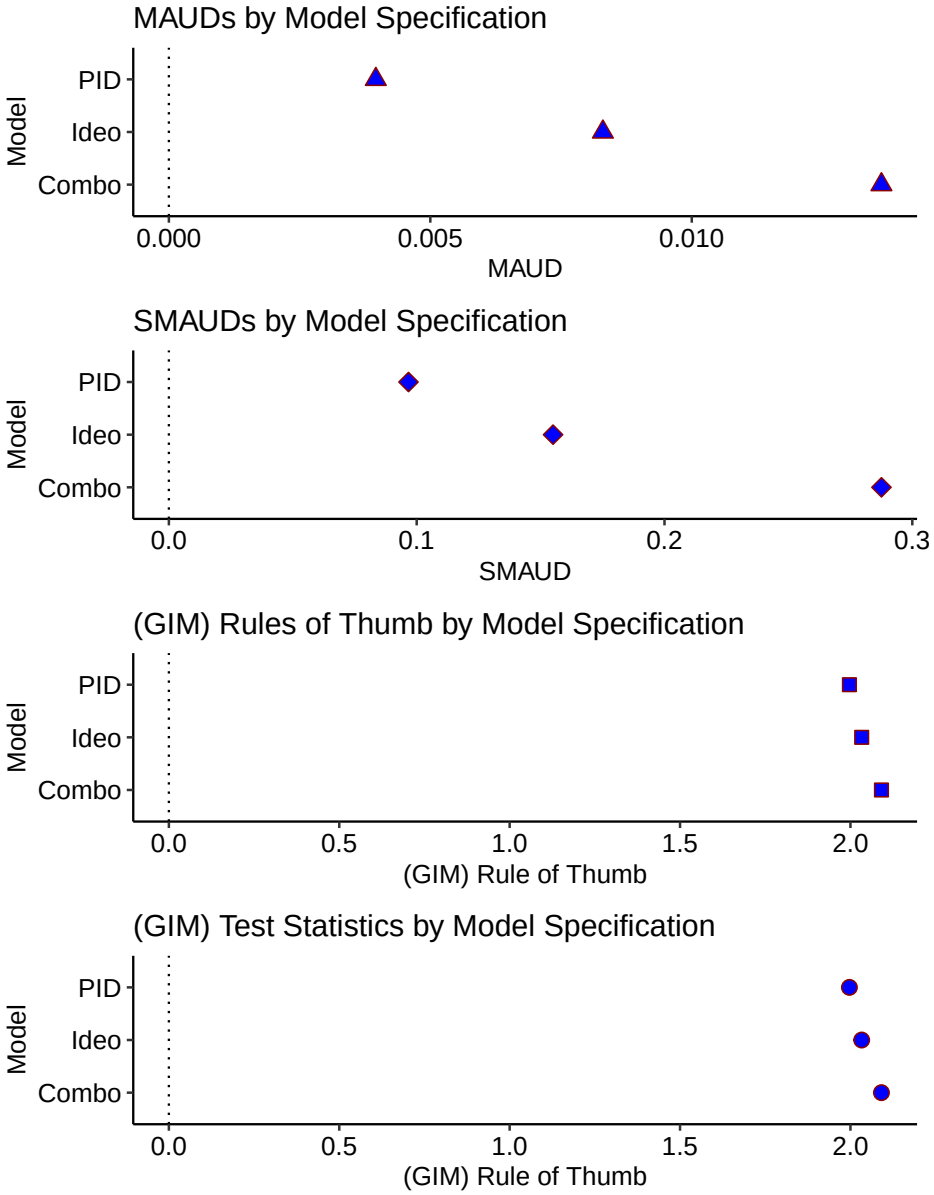
Figure 9. Model Dependence by Specification (URL Proxies)



(A.5) Model Dependence in Guess (2021) Data + URL Section Name Proxy Method

We find model dependence results in **Figure 10** that differ yet again using data from Guess (2021) alongside the URL section name proxy classification method from Tyler, Grimmer, & Iyengar (2021).

Figure 10. Model Dependence in URL Proxy Models



(A.6) Leave-One-Covariate-Out (LOCO) Inference

Leave-one-covariate-out (LOCO) inference was first proposed in Lei et al. (2018) and generates assessments of variable importance via conformal prediction bands.

The first step in conducting LOCO inference is to select some machine learning algorithm, f, to use for predicting the outcome variable. Then, split the sample in half into D_1 and D_2 such that $D_1 \cup D_2 = 1, \ldots, n$. The algorithm will then estimate $\hat{f}_{n_1}$ on the first sample, D_1. Next, we select variable j and estimate $\hat{f}_{n_1}^{-j}$, the prediction of the outcome variable on a model

Finally, D_2 is used to construct a finite-sample, distribution-free confidence interval for

$$\theta_j(D_1) = \text{med}\left(|Y - \hat{f}_{n_1}^{-j}(X)| - |Y - \hat{f}_{n_1}(X)| \middle| D_1 \right) .$$

The beauty of this methodology is twofold. First, it can generate a highly interpretable and intuitive measure of relative importance of your explanatory variables in predicting the outcome variable. Second, you can employ any machine learning algorithm for estimating the conditional expectation function, $f(x) = \mathbb{E}(Y|X = x)$.

We use three machine learning algorithms with cross-validation in our study: LASSO (least absolute shrinking and selection operator), stepwise regression, and random forests. LASSO works similarly to the classical ordinary least squares estimator, but with a constraint imposed as a sublevel set of the ℓ_1 norm so as to "shrink" certain covariates that are less important towards or to 0. Stepwise regression generates 2^j potential models with j covariates and uses either forward selection or backward elimination to fit the optimal model. Stepwise regression is a somewhat antiquated method, and many critiques of it have been generated, particularly as model selection is path-dependent and selecting on p-values may not be an optimal objective function. Random forests are advantageous in accounting for potential nonlinearities and work by bagging an ensemble of random trees (CART).

For more details on these specific machine learning algorithms, please see Hastie, Tibshirani,

& Friedman (2008).