Measuring Elite Personality Using Speech

Adam J. Ramey, Jonathan D. Klingler and Gary E. Hollibaugh, Jr

Political Science Research and Methods / *FirstView* Article / April 2016, pp 1 - 22
DOI: 10.1017/psrm.2016.12, Published online: 18 March 2016

Link to this article: http://journals.cambridge.org/abstract_S2049847016000121

How to cite this article:
Adam J. Ramey, Jonathan D. Klingler and Gary E. Hollibaugh, Jr Measuring Elite Personality Using Speech. Political Science Research and Methods, Available on CJO 2016 doi:10.1017/psrm.2016.12

 doi:10.1017/psrm.2016.12

Measuring Elite Personality Using Speech*

ADAM J. RAMEY, JONATHAN D. KLINGLER AND GARY E. HOLLIBAUGH, JR.

We apply recent advances in machine learning to measure Congressmember personality traits using floor speeches from 1996 to 2014. We also demonstrate the superiority of text-based measurement over survey-based measurement by showing that personality traits are correlated with survey response rates for members of Congress. Finally, we provide one empirical application showcasing the importance of personality on congressional behavior.

Recently, political scientists have taken notice of the role of personality in political behavior. The five-factor, or "Big Five," model has been applied widely in the behavior literature to the study of participation, ideology, and vote choice (e.g., Caprara et al. 2006; Gerber et al. 2010; Mondak et al. 2010; Gerber et al. 2011a). This model decomposes individual personality into five traits—Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability. Though growing rapidly, this literature has eschewed a focus on elite behavior. This is unfortunate, as the study of personality among the public has yielded many findings linking personality to ideology (Jost, Nosek and Gosling 2008; Gerber et al. 2010), political information (Gerber et al. 2011b), civic engagement (Mondak et al. 2010), and many other dynamics important to the political sphere. Indeed, while personality is related to ideology and other factors that predict behavior, it offers predictive power above and beyond these more traditional measures.

A key reason for the lack of focus on elites is that measuring elite personality is more problematic than it may seem at first, as most political science applications involve surveys (Caprara, Barbaranelli and Zimbardo 2002; Gerber et al. 2011a). While useful, these studies tell us little about the traits of elites. One solution might be to survey legislators, but this approach would prevent us from being able to study the deceased and would be difficult to perform for retired legislators. Moreover, even if we restrict ourselves to the contemporary legislative period, most legislators would be unwilling to participate, and any responses would likely be subject to selection bias and strategic decisions. Indeed, to our knowledge, only one study has attempted to apply survey-based inventories with legislators (Dietrich et al. 2012).[1]

* Adam J. Ramey is an Assistant Professor of Politics, New York University Abu Dhabi, PO Box 129188, Abu Dhabi (adam.ramey@nyu.edu). Jonathan D. Klingler is a Postdoctoral Fellow in the Institute for Advanced Study in Toulouse, Université Toulouse 1 Capitole, 21 alleé de Brienne, 31000 Toulouse (jonathan.klingler@iast.fr). Gary E. Hollibaugh, Jr. is an Assistant Professor in the Department of Political Science, University of Notre Dame, Notre Dame, IN 46556 (gholliba@nd.edu). All authors contributed equally to the paper. Support through ANR-Labex IAST is gratefully acknowledged. The authors thank Ken Benoit, Matt Blackwell, Richard Bonneau, Drew Dimmery, Conor Dowling, Michael Gill, Andy Harris, Pablo Hernandez-Lagos, John Jost, Slava Mihkaylov, Jeff Mondak, Jonathan Nagler, David Nickerson, Elena Panova, John Patty, Michael Peress, Dave Primo, Molly Roberts, Larry Rothenberg, Maya Sen, Jo Silvester, Arthur Spirling, Karine van der Straeten, and participants at the 5th Annual Text as Data Conference and the Rooney Center for the Study of American Democracy for comments and feedback. All remaining errors are their own. To view supplementary material for this article, please visit http://dx.doi.org/10.1017/psrm.2016.12

[1] Though see Silvester, Wyatt and Randall (2014).

Response rates were low—ranging from 17 to 26 percent. These rates, coupled with the inability to get estimates for past time periods, have prevented scholars of behavior in institutions from incorporating personality into their analyses.

In this paper, we present a method for measuring elite personality without surveys. Drawing on recent advances in machine learning, we measure legislator personality using floor speeches, estimating Big Five personality traits for all incumbent members of the US House from 1996 to 2014. We show that these estimates parallel findings in the behavior literature linking personality with ideology (Jost, Nosek and Gosling 2008; Gerber et al. 2010) and that they are robust to authorship and speechwriting effects. We then use our estimates to shed light on the propensity of legislators to respond to surveys. While perhaps uninteresting on its face, our results suggest the propensity to respond is a function of personality, and that attempting to measure elite personality using surveys will result in selection bias. Finally, we provide one example of the empirical utility of these estimates and examine the relationship between personality and the types of bills legislators propose.[2]

MEASURING PERSONALITY: FROM SPEECHES TO SCORES

Given the aforementioned issues with survey-based measurement of personality, studying the role of personality in legislator behavior—or, more generally, elite behavior—requires a method different from direct surveys. We therefore draw on a recent literature in machine learning that uses traditional psychometric personality inventories in conjunction with written texts and auditory transcriptions to train predictive models for personality (Mairesse et al. 2007; Mairesse and Walker 2008; Schuller et al. 2013). In a foundational piece in this literature Mairesse et al. (2007) develop a widely applicable method for generating personality estimates from speech and text. Using Pennebaker and King's (1999) corpus of nearly 1.9 million words from laboratory experiments, and Mehl, Gosling and Pennebaker's (2006) corpus of ~100,000 words from recorded conversations, Mairesse et al. (2007) train several machine learning models to predict personality traits. Machine learning methods are a class of models that seek to predict an observed output with optimal combinations of features. The models are "trained" on a subset of data and the estimates are used to predict the rest of the data using only right-hand-side variables. An example would be to perform a linear regression of some variable $\boldsymbol{y}$ on a matrix $\boldsymbol{X}$ of covariates for a fraction of a sample of data. Then, the coefficients β estimated using the "training" set would be validated by predicting $\boldsymbol{y}$ using only the $\boldsymbol{X}$ for the unused part of the sample.[3]

As mentioned, Mairesse et al. (2007) use written and spoken language to predict personality traits, which are measured by independent observers and self-placement.[4] Crucial for our purposes, words are categorized according to Pennebaker, Francis and Booth (2001) *Linguistic Inventory and Word Count* (LIWC) dictionary (2001 edition), as well as Coltheart's (1981) *MRC Psycholinguistic Database* (MRCPD).[5] Doing so allows scholars to generalize to domains beyond the confines of the laboratory. Both the LIWC and MRCPD search for linguistic features in texts, such the number of second person pronouns, punctuation, six letter words, and more. After processing the data using these dictionaries and reducing the spoken and written text to a collection of linguistic features, Mairesse et al. (2007) train several machine

[2] For a more thorough examination of the substantive relationships between personality and congressional behavior, see Ramey, Klingler and Hollibaugh (2015).

[3] A fuller description of the method is found in the Online Appendix.

[4] We use the models trained on independent observer data.

[5] The categories are found in the Online Appendix to this chapter.

learning algorithms on the data.[6] They find Support Vector Machines for Regression (Shevade et al. 1999; Smola and Schölkopf 2004; Bishop 2006; Hastie, Tibshirani and Friedman 2009)—hereafter referred to as SMOreg—best recovers personality measures in written trials.[7] Thus, we use SMOreg in the analyses that follow.[8]

Using Mairesse et al.'s (2007) pre-trained SMOreg model, our goal is to predict the personality traits of members of Congress. However, applying this approach to Congress requires texts to feed to the models. To that end, the most systematic data available are speeches made on the floors of the House and Senate, which can be found in the *Congressional Record*. Though there exist concerns with using these data, as legislators might use floor speeches to strategically convey their policy preferences, constituency preferences, or (remote, but possibly) personality or leadership profiles, we show below that our measures of personality, while correlating with Common Space ideological scores (and generally in the expected directions), explain only a small fraction of the variance, suggesting personality is not equivalent to ideology and our estimates have meaning independent of it. As for the third concern, this is only a problem if legislators try to convey "fake" profiles. Relative to one-time self-administered personality surveys, the Mairesse et al. (2007) method should result in more accurate profiles, as it will be harder to maintain a "fake" profile over the course of a career than over the span of a short survey. Nevertheless, this effect would simply attenuate our results. Furthermore, if some legislators were speaking sincerely and others strategically, further attenuation would occur, and our estimates would have little predictive power.

Last, we might be concerned that the subject matter reflected in the Pennebaker essay corpus might not map well to legislator floor speeches. Specifically, the kinds of words used by laboratory participants are almost certainly different from those employed by legislators in their floor speeches. To assess this concern, Figure 1 plots the mean LIWC category usage from the Pennebaker data against the mean usage of members of the 114th House of Representatives.[9] Each point represents a LIWC category; a version of this plot with category labels in found in the Online Appendix. As we see, there is a close correspondence between the two; the correlation on the untransformed scale is ~0.99. Additionally, the τ_B rank correlation coefficient for the rank orderings across the two is 0.72 ($p < 0.001$). Thus, while the subject matter of the texts are different, legislators' usage of the various categories does not differ substantively from the laboratory essay writers.[10]

[6] JAVA code for Mairesse et al.'s (2007) Personality Recognizer program is found at http://people.csail.mit.edu/francois/research/personality/recognizer.html. The program performs both the LIWC and MRCPD processing as well as the fitting of the machine learning models. While each of these steps can be performed separately using other programs, the JAVA program provides a convenient wrapper for performing all necessary steps.

[7] A description of the SMOreg model is in the Online Appendix. Intuitively, SMOreg is simply a more complicated version of linear regression, with the dependent variables in our case being the personality trait measures, and the independent variables the LIWC scores. However, in contrast to linear regression that simply minimizes the sum of squared errors, SMOreg minimizes the sum of squared coefficients, subject to a prespecified precision ε. This necessitates that the problem be viewed as one of constrained optimization (Karush 1939; Kuhn and Tucker 1951)—as opposed to the typical unconstrained optimization of OLS or (usually) maximum likelihood. With these exceptions, however, the underlying intuition is similar. Implementation of the SMOreg model was performed in Weka (Hall et al. 2009; Hornik, Buchta and Zeileis 2009).

[8] However, for our application, the choice of model used—in particular, SMOreg versus M5 model trees with linear models (Quinlan 1992)—seems to make little substantive difference, presumably due to the size of our data.

[9] The hyperbolic arcsine transformation used for the axes behaves similarly to a log transform, except that it "stretches" small values more, "compresses" larger values more, and is not undefined for zeroes. This transformation is necessary for presentation, as some LIWC categories have much larger observed values than others.

[10] To be sure, we are not claiming that the vocabularies or distributions of actual words used are the same—or even comparable—between laboratory essay writers and members of Congress. Rather, the distributions of LIWC categories (which simply characterizes the types of words used by structure, tone, and topic) are similar across the two types. Moreover, we also refrain from claiming that the personality types exhibited by members of

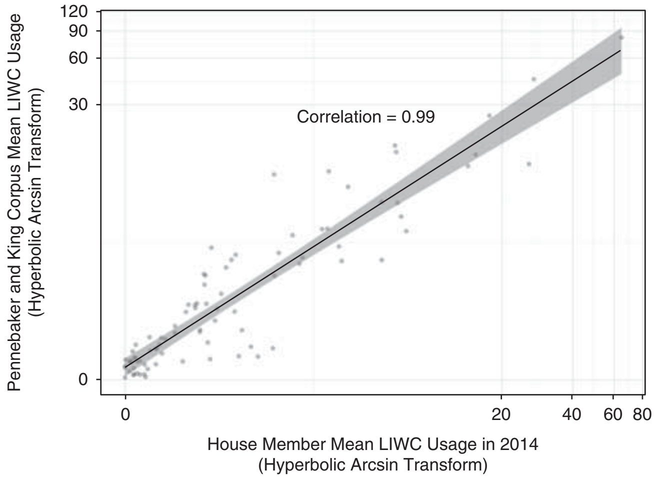

Fig. 1. Comparing Linguistic Inventory and Word Count (LIWC) (2001) usage between the Pennebaker corpus and floor speeches

As none of these concerns seem to be particularly damning, we apply Mairesse et al.'s (2007) SMOreg model to the corpus of every legislative speech by every sitting member of the House of Representatives in the 104th–113th Congresses (1996–2014). The procedure is as follows:

1. For each set of speeches by a given legislator in a fixed time span (in our case, by year, though Congress-level estimates can be similarly estimated), process the raw text through LIWC and MRCPD to get counts of word usage across all LIWC and MRCPD categories.[11]
2. Once all speeches have been processed for the given span, feed the directory of texts through the Mairesse et al. (2007) models in corpus mode. Specifically, all LIWC and MRCPD features are standardized and then fed in to the standardized SMOReg models. Though we showed above that the category usage of the training (essay) sample is substantively similar

(F'note continued)
Congress are distributionally similar to those in the general population (or even the laboratory sample). Indeed, there is almost certainly a great deal of self-selection occurring; for example, the average member of Congress is almost certainly more Extraverted than the average member of the general population. However, this latter concern is not necessarily problematic for us, as the personality scores are estimated relative to the corpus used. That is, the personality scores we estimate are normalized relative to the *Congressional Record*. While this means that the scores presented here are not comparable with those from the general population, and we cannot directly compare our estimates with those recovered from common personality inventories like the Ten Item Personality Inventory ("TIPI") (Gosling, Rentfrow and Swann 2003)—or even other estimates recovered from the method presented here that are based on different corpora—it does mean that they are comparable with each other, as they are based on the same source material (the *Congressional Record*). However, they are comparable only in a relative sense, and a score of 4 on our scale is not necessarily comparable with a score of 4 recovered from other personality inventories.

[11] Our choice to do the estimation by year is not related to the debate over whether personality is static or dynamic. Rather, as language changes from Congress to Congress, it is important to ensure language is measured relative to the time period in which it is delivered. Additionally, there are practical computational considerations at play here, as estimating Congress-level scores taxed our available computing power to its limit. Increasing the amount of data used at any one time by an order of magnitude (as we estimated scores for the 104th–113th Congresses) would require us to describe the time necessary to estimate the scores and bootstrapped standard errors in terms of months as opposed to weeks.

to the legislator sample, using the standardized models offers additional leverage on estimating the personality traits within the legislative domain.

3. The output from Step 2 (Big Five estimates for each legislator) is saved and the process is repeated for as many years as there are available.
4. (Optional correction) A legislator i's personality on dimension d in year $t = 1, 2, \ldots, T_i$ (T_i is the number of years in which legislator i served) is rescaled as the jackknifed average of his/her personality estimates on that dimension for all years except t That is, $\tilde{\theta}^d_{it} = \frac{1}{T_i - 1} \sum_{t' \neq t} \hat{\theta}^d_{it'}$.[12]

   This ensures that personality trait estimates in session t are not corrupted by language associated with particular actions in that session (e.g., legislators that cosponsor more might be perceived as more Agreeable, not because they cosponsor more legislation *per se*, but simply because the act of cosponsorship inherently entails using Agreeable language).[13] In short, this correction alleviates concerns of endogeneity.[14]
5. To generate measures of uncertainty, we follow Lowe and Benoit (2011) in using a sentence-level bootstrap. Specifically, suppose legislator i utters N_{it} sentences in time period t. For all i, we resample from the set of their sentences N_{it} times with replacement (Efron and Tibshirani 1994). At the year level, we conduct 100 bootstrap replications/member and compute the empirical 95 percent confidence interval.[15] To measure uncertainty in the jackknifed estimates, we take the legislator's estimates by year across each of the 100 bootstraps, calculate the jackknife per Step 4 (if desired), and then compute the 95 percent confidence interval.[16]

Importantly, we are agnostic as to whether our personality scores are measures of "sincere" legislator personality. Like with ideal point estimates based on roll calls, we simply consider these estimates as *revealed* and potentially strategic preferences. Estimates of the traits and confidence intervals are presented for key House members in Figure 2. The scale for each trait ranges from 1 to 7. As we see, the estimates are stable and precise, despite the fact that only single Congresses—as opposed to the entire corpus—were used at any one time during the estimation process.

Additionally, Figure 3 presents the relationship between word count and confidence interval width. The median member utters around 11,000 words/year. Once the word count exceeds about 5000 words, the confidence interval width is about 0.3 or less, meaning the 95 percent interval is the point estimate ± 0.15 on the seven-point scale. Thus, the estimates are relatively precise for the vast majority of legislators. Moreover, the estimates make intuitive sense. Former Representative Ron Paul (R-TX) is significantly more Open and less Emotionally Stable than the rest of the legislators in this subsample. Representative Paul's out-of-the-box libertarian ideology and his frequent diatribes against the status quo politics embodied by the two major parties are in line with common characterizations of both of these traits (McCrae and John 1992; Jost, Nosek and Gosling 2008). That said, our main goal in Figure 2 is not necessarily to make any claims about the personalities (either relative or absolute) of any particular members of Congress.

[12] Absent strategic concerns, the raw scores or simple arithmetic means across years may be used.

[13] Another way to estimate the jackknifed scores would be to pool all speeches excluding the current year, perform the estimation, and repeat for each year. We avoid this for two reasons. First, the agenda changes substantially from year-to-year and feature usage patterns can change over time. Second, it is computationally expensive to do so.

[14] Congress-level jackknife estimates can be estimated in a similar way. Let $c_i = 1, 2, \ldots, C_i$ be the number of Congresses in which member i served. As each Congress is a two-year period, denote the years associated with Congress c_i by $t(c_i)$. Member i's jackknifed score on dimension d in Congress c is thus $\tilde{\theta}^d_{ic} = \frac{1}{T_i - 2} \sum_{t' \notin t(c_i)} \hat{\theta}^d_{it'}$.

[15] The number of bootstrap replications has little effect on the estimates.

[16] Upon publication, the scores will be released.

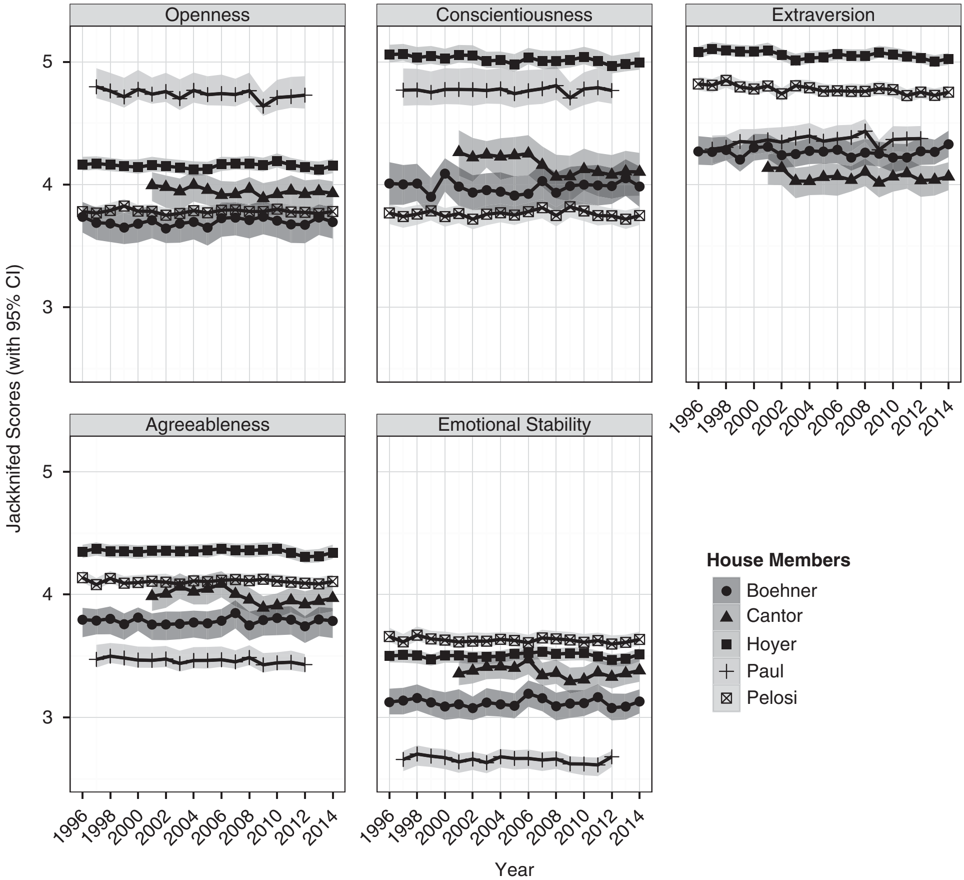

Fig. 2. House scores over time (selected members)
Note: CI = confidence interval.

Rather, we simply want to show that the scores are stable over time and not subject to great deals of uncertainty. Indeed, at the present time, more stringent forms of validation are not yet possible; as we have mentioned earlier, previous efforts to ascertain the personalities of legislators via more traditional personality inventories have resulted in low response rates and high levels of clustering at the "more desirable" ends of each scale (Dietrich et al. 2012). Whether this clustering reflects the desire of legislators to make themselves look more appealing to constituents or the natural self-selection of certain personality types into elected office has not yet been determined.[17] Additionally, as we show later, personality traits themselves help explain the rate of response to less intrusive surveys that ask about policy preferences; the relationship between personality and the rates of response to personality

[17] It might also be worth considering that response rates might also be correlated with the prominence of the office—with those in more prominent offices having more to lose if they admit to "undesirable" personality profiles, and also having more staff members acting as "gatekeepers" to the officeholder in question—and that surveys of local officeholders might elicit higher response rates. However, even in this case, validation using our method would be useful to examine the survey results that we do receive for desirability bias. This is a question left for future research.

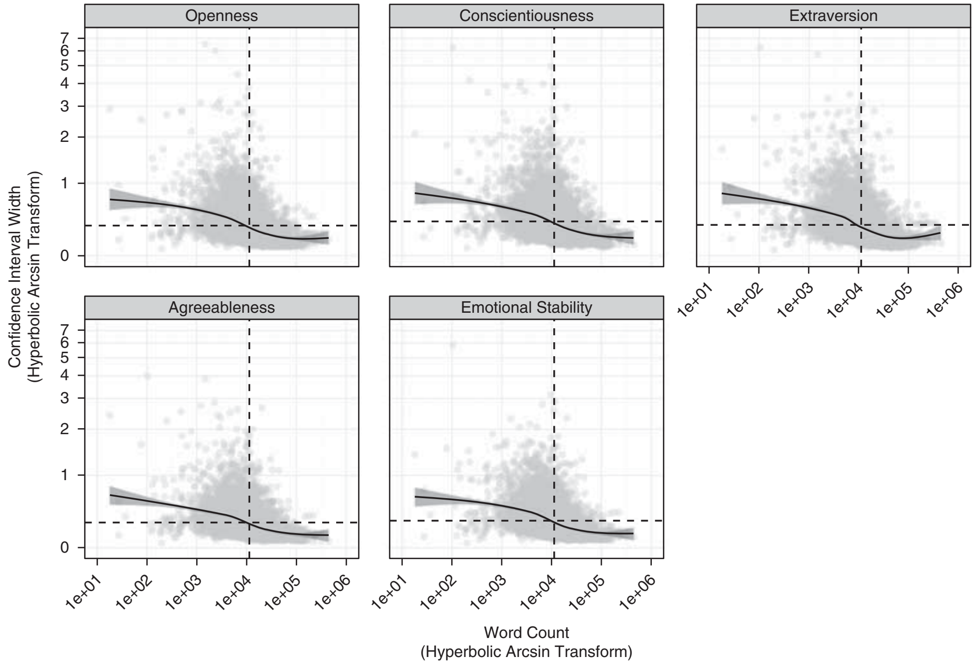

Fig. 3. Word count and precision
Note: Points are individual members. The smoothed line is a loess-smoothed trend. The vertical dashed lines are the median member's word count (~11,000 words/year). Horizontal dashed lines are the average confidence interval width (~0.35). Thus, most members' personality estimates are their point estimates ±0.175.

inventories would almost assuredly be stronger, especially given the more intrusive nature of the latter. All that aside, we can perform some more indirect tests of validity, which we discuss in the following section.

VALIDITY CHECKS

Strategic Concerns

We acknowledge legislators may have incentives to misrepresent their "true" personality traits and adopt false personas. In this case, their behavior and language would be affected as well. While we acknowledge this concern as a serious issue, based on the evidence currently available, we do not believe it is a serious impediment to the measurement of legislators' beliefs and/or preferences, as exhibited through the Big Five personality traits.

Indeed, a common objection to the standard NEO-PI-R personality inventory (Costa and McCrae 1992) is that the "desired" trait or answer is transparent. Critics of the inventory argue the average respondent would certainly prefer to think of him or herself as extraverted rather than introverted, emotionally stable rather than neurotic, agreeable rather than selfish, conscientious rather than lazy, and open to experience rather than insular. This concern appears to gain support from personality inventories administered to state legislators (Dietrich et al. 2012), as the responding legislators' responses are substantially skewed toward these "favorable" personality trait values, suggesting personality traits serve as valence characteristics.

Indeed, legislators may consciously adopt speech patterns in order to signal these valence characteristics. There are likely many verbal cues that both indicate personality types to legislators and voters, and are known to all and can be voluntarily suppressed or performed insincerely. However, we also assume there exists some subset of these cues tied to personality, which are unknown to legislators, but may be captured by the personality recognizer described above.

We suppose legislators also have unconscious verbal cues that signal personality and that the algorithm can read them. In that case, it is in the interests of legislators to misrepresent their personalities by appearing to hold personality trait values as favorable as possible throughout their public speech. While all known cues should be utilized to develop valence trait values, the unknown subset of cues should exhibit variation as legislators are simply unaware of their link with personality. Combining these known and unknown subsets of speech elements should result in measures that appear skewed toward the valence characteristics, but with substantial variation reflecting the true variation in the underlying trait values.

Face Validity

Theoretical concerns aside, it is important to ensure our estimates have some degree of face validity. Traditionally, personality psychologists have relied on surveys to develop measures on the Big Five dimensions. However, as we note above, surveying current and former federal legislators is not likely to yield adequate responses given the difficulty other scholars have faced in reaching currently serving state legislators. Therefore, we will instead rely on more indirect validity checks based on associations repeatedly uncovered in the literature. Our analysis here will focus on the relationship between personality and ideology.

As mentioned, much recent research has focused on the relationship between personality and partisan attitudes (Gerber et al. 2010; Mondak 2010; Gerber et al. 2011a). Strong and consistent relationships have been uncovered between ideology and Openness, Conscientiousness, and Emotional Stability, with the former being associated with liberalism, and the latter two associated with conservatism. Effects for the other two traits are more mixed; Mondak (2010) finds no relationship between Extraversion and ideology, and only a weak—and inconsistent—relationship between Agreeableness and liberalism. Gerber et al. (2011a) find the same relationships between ideology and Openness, Conscientiousness, and Emotional Stability. Moreover, they find no relationship between Agreeableness and self-reported ideology, but uncover divergent relationships between Agreeableness and economic attitudes, and Agreeableness and social attitudes; more Agreeable people are more economically liberal and socially conservative. Finally, they find a positive relationship between Extraversion and conservatism, though the relationship is weaker than those between ideology and the other four traits. Similarly, Gerber et al. (2010) find consistent relationships between ideology and Openness, Conscientiousness, and Emotional Stability in the same directions, but the other relationships are more nuanced.

Nonetheless, we can use these findings to perform an indirect test of the validity of our estimates. Using Common Space (Poole and Rosenthal 1997) scores, we can examine the direction of the relationships between the Big Five traits and ideology. From these findings, we should expect a negative relationship between Openness and the Common Space score, and positive relationships between Conscientiousness, Emotional Stability, and the Common Space Score. Additionally, as Common Space scores tap into the underlying economic conflict between the two dominant parties (roughly approximate to the contemporary liberal-conservative dimension), we should find a negative relationship between Agreeableness and the

TABLE 1 *Ordinary Least Squares Models of Personality and Common Space Scores (1996–2014)*

	Model 1	Model 2	Model 3	Model 4
Openness	–0.26***	–0.26***	–0.22***	–0.22***
	(0.04)	(0.04)	(0.04)	(0.04)
Conscientiousness	0.26***	0.24***	0.27***	0.25***
	(0.04)	(0.03)	(0.03)	(0.03)
Extraversion	–0.15***	–0.13***	–0.16***	–0.14***
	(0.03)	(0.03)	(0.03)	(0.02)
Agreeableness	–0.17**	–0.15**	–0.18***	–0.16**
	(0.07)	(0.07)	(0.07)	(0.07)
Emotional Stability	0.20***	0.17***	0.22***	0.19***
	(0.04)	(0.04)	(0.04)	(0.04)
Female	–	–0.21***	–	–0.19***
		(0.04)		(0.04)
Birth Year	–	–	0.01***	0.01***
			(0.00)	(0.00)
Constant	0.64***	0.75***	−15.21***	−14.31***
	(0.19)	(0.19)	(2.53)	(2.50)
R^2	0.10	0.13	0.14	0.16
Adjusted R^2	0.09	0.12	0.13	0.16
Number of observations	844	844	844	844

Note: Standard errors in parentheses.
Two-tailed tests: ***p < 0.01, **p < 0.05, *p < 0.1.

Common Space score (Poole and Rosenthal 1997; Crespin and Rohde 2010).[18] Finally, the literature is much more mixed on the relationship between ideology and Extraversion; as such, we remain agnostic regarding the expected sign (or significance). Table 1 presents several ordinary least squares (OLS) models of ideology as a product of personality (as measured by the lifetime means of each trait) and demographic traits.[19]

All four traits with expected relationships have statistically significant coefficients in the expected directions in each model. Additionally, in line with previous literature, the coefficients on Extraversion are of smaller magnitudes than those for the other traits, and this holds for all four models (though the relationship between ideology and Extraversion is still a point of contention). Moreover, these results allay one natural concern with using potentially ideologically-tinged legislative speeches to estimate personality, in that our personality estimates may be simply summaries of legislator ideology. However, our model R^2s in Table 1 are not large, even when controlling for the additional demographic variables, suggesting personality traits alone do not account for large proportions of the variance in ideology.[20] All together, the results are in line with expectations from the literature (Jost, Nosek and Gosling 2008;

[18] No other trait has these divergent effects.

[19] We use *Birth Year* instead of age because Common Space scores are static, and we therefore use lifetime means for the personality traits. This ensures that each House member in our analysis is included only once.

[20] Though ideology is the dependent variable in this analysis, we refrain from commenting on the debate over whether personality traits are causally prior to ideology (e.g., Mondak and Halperin 2008; Gerber et al. 2010; Mondak 2010; Kandler, Bleidorn and Riemann 2012) or whether personality traits and ideology are caused by the same—usually genetic—underlying factor (e.g., Eaves and Eysenck 1974; Verhulst, Hatemi and Martin 2010; Verhulst, Eaves and Hatemi 2012). All we assume is that ideology and personality are somehow related, which an assumption congruent with both sides of the aforementioned debate and also consistent with the statistical significance reached in our results, yet measure different concepts (Alford and Hibbing 2007), which is consistent with the relatively low R^2s.

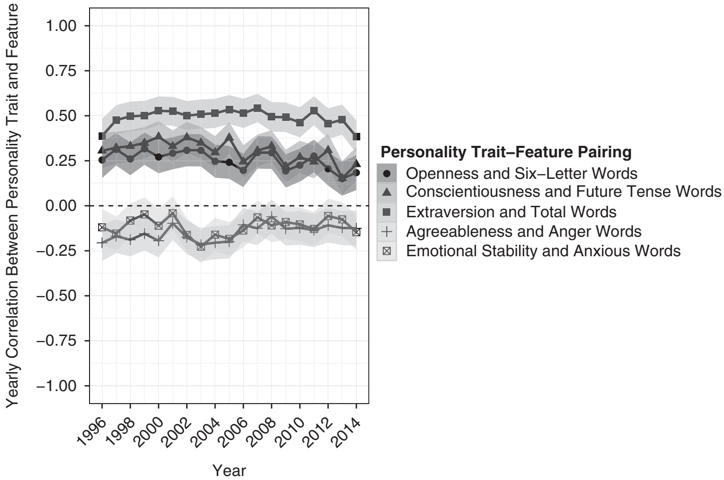

Fig. 4. Correlations between personality traits and selected Linguistic Inventory and Word Count features

Gerber et al. 2010). Specifically, Jost, Nosek and Gosling (2008) find that while there are correlations between the Big Five and ideology, the magnitudes of these are not large. Our model R^2 is very much consistent with this finding. This suggests whatever theoretical concerns one might have about the dependence of the Big Five on ideology or the ideological content of the legislative record are, in practice, not a problem.

Personality Traits and Linguistic Features

We can also examine the relationships between LIWC/MRCPD features and the estimated personality traits to see if they have face validity. While a full feature exploration is beyond the scope of this paper due to reasons of length (as there are over 80 features), we can examine the stability of selected relationships. To wit, Figure 4 presents the yearly correlations between the Big Five and one feature per trait.[21] The correlations are remarkably stable over time and generally are distinct from 0 at the 95 percent level.

However, as the LIWC/MRCPD traits are used in the estimation of the personality traits, it is not surprising that they are relatively stable over time. What is more important is that these relationships make sense. As Openness is associated with intellectual pursuits (Borgatta 1964; Tupes and Christal 1992)—indeed, the trait is sometimes referred to as Intellect rather than Openness (Saucier and Goldberg 1996)—we might expect more Open members to express themselves in more conventionally “intellectual” ways; one rough proxy for intellect is the length of words used (Vetterli and Furedy 1997).[22] As such, it should be unsurprising that more Open members use longer words at higher rates. Additionally, more Conscientious individuals tend to be more goal-oriented and have greater abilities to engage in delayed gratification (Duckworth, Tsukayama and Kirby 2013), and should therefore focus on the future and long-term goals more than those who are less Conscientious (Barrick, Mount and Strauss 1993; Ozer and Benet-Martinez 2006). As Figure 4 shows, this is also supported, as more

[21] All yearly correlations for all trait–feature pairs are in the Online Appendix.

[22] Pennebaker and King (1999) uncovered the same relationship between word length and Openness. Additionally, Küfner et al. (2010) uncovered positive correlations between Openness and “sophisticated writing.”

Conscientious members use future tense words at higher rates. Similarly, more Extraverted members should be more sociable and talkative, and they do speak more in our data.[23] Additionally, as Agreeableness is almost tautologically at odds with the propensity to anger (Gosling, Rentfrow and Swann 2003)—indeed, the TIPI asks respondents if they are "quarrelsome" and the responses are used to code for Agreeableness or the lack thereof—legislators should use angry, off-putting language at lower rates.[24] This relationship, though weaker than those described above, is also found in our data, and in the expected direction. Finally, as Emotional Stability is "believed to reflect the general predisposition to develop psychological symptoms such as anxiety" (Muris et al. 2005, 1106), more Emotionally Stable members should exhibit less outward anxiety; again, this relationship is found in our data, though it is somewhat weaker than those for Openness, Conscientiousness, and Extraversion. Nonetheless, the relationships between the features and the estimated traits are not only stable, but are generally in the expected directions, thus providing additional face validity.

Authorship and Speechwriter Effects

An additional important concern to address is the potential for speechwriting effects. Specifically, as at least some of members' speeches are written by their staff members, our estimates might be reflective of the writers' personalities and not those of the Congressmembers. This turns out to be a nonissue for a number of reasons. First, few members have professional speechwriting staff. A cursory look at the House disbursement records demonstrates that, outside of the Speaker and a few key members, speechwriters are simply not hired by most members.[25] This does not mean that other staffers do not write speeches. However, as staff turnover is notoriously high (Congressional Management Foundation and Society for Human Resource Management 2013), it is reasonable to assume that *many* staffers would contribute to many speeches and, if so, our estimates should be unstable or clustered for all members. As Figure 2 shows, for some of the highest ranking and/or most prominent members of Congress during our time period, there is marked stability in the jackknifed personality estimates. Perhaps most assuring, neither Pelosi nor Boehner were Speaker at the start of the data, yet the estimates remain consistent. This suggests their speechwriters maintained the same linguistic patterns over time.

Indeed, the Congressional Research Service admonishes those individuals charged with speechwriting for Congressmembers to make every effort to mimic the language and diction of the member for whom they are writing (Neale 1998):

> Congressional speechwriters should make every effort to become familiar with the speaking style of the Member for whom they are writing, and adjust their drafts accordingly.

These results somewhat contradict those of Sigelman (2002), who argues that President Reagan's revealed persona was different in ghostwritten and off-the-cuff remarks. However, he notes that while differences exist between Reagan with speechwriters and Reagan without

[23] This is consistent with the very strong correlations ($r > 0.6$) between Extraversion and total word count found in previous work (Mehl, Gosling and Pennebaker 2006).

[24] Barlett and Anderson (2012) showed that Agreeableness was not only negatively related to physical aggression and violent behavior, but it was also with aggressive emotions and attitudes more generally. Additionally, Jovanović et al. (2011) found that more Agreeable drivers had lower levels of driving-related anger and drove less aggressively. These results provide further evidence that more Agreeable individuals are not only slower to anger, but less likely to express that particular emotion.

[25] See, for example, http://disbursements.house.gov/2013q4/2013q4_singlevolume.pdf

speechwriters, they are simply variations on a theme. As presidents have arguably the most incentive and ability to shape their public personalities, and the difference is still minimal, there should be even less of an effect for members of Congress. This is precisely what we observe.

SPEECHES VERSUS SURVEYS

Despite the advantages proffered by the speech-based approach, there may be circumstances in which a survey-based approach might be preferable, at least for contemporaries. Indeed, surveys might be preferable for these types of legislators if they were perfectly responsive. However, missingness would not necessarily be a problem, as personality traits could conceivably be imputed using existing algorithms (e.g., Rubin 1987; Gelman, King and Liu 1998). For example, this is a potentially appealing route, as this would allow us to recover estimates for those who did not respond. However, this approach has limitations. Most important for our purposes here is that the missingness must obey the so-called missing at random (MAR) assumption, which is satisfied if missingness can be modeled as a function of observed data.[26] A canonical example of this sort of missingness is the case of high wage earners who fail to report their income in surveys. While the income may be unobserved, several known correlates (e.g., education) are not missing. By conditioning on these observed values, we may model missingness using existing algorithms. However, if the missing observations cannot be predicted from observed covariates, MAR is not satisfied and multiple imputation is no longer an option (Weisberg 2009).

This is most problematic if the missingness is in part due to personality. In these cases, missingness is a function of the underlying latent traits the surveys seek to recover, and several characteristics that help predict missingness are themselves missing. Therefore, imputation cannot be used with any degree of certainty. Moreover, the distribution of responsive legislators is subject to selection bias, and any inferences drawn from these data will themselves be problematic. Using predicted values from Heckman (1976) selection models will not suffice either, as important covariates that predict selection into the sample (i.e., personality traits) will be missing.[27]

To check whether this is potentially a problem, we examine the relationship between our personality scores and response rates to the National Political Awareness Test (NPAT), a survey developed by Project Vote Smart that asks candidates for congressional elections to provide answers to a series of political issue questions.[28] This survey has received a great deal of attention by political scientists and has been used to estimate the ideology of legislators and candidates (e.g., Ansolabehere, Snyder and Stewart 2001; Shor and McCarty 2011; Battista, Peress and Richman 2012). Response rates to the NPAT are higher than those from previous (anonymous) studies of personality, presumably because the NPAT questions ask about policy positions, which are more in line with the roles played by legislators, as opposed to the more personally intrusive personality surveys. However, both types of surveys have potential consequences for giving the “wrong” response. As Dietrich et al. (2012) rightly note about their

[26] Following King et al. (2001), let D_{obs} denote observed data, D_{mis} denote missing data, D denote the total data, and M represent missingness. Data are MAR if $\Pr(M|D_{\text{obs}}, D_{\text{mis}}) = \Pr(M|D_{\text{obs}})$.

[27] The lack of personality traits for those not selecting into the sample is also why we cannot take the approach of Klingler, Hollibaugh and Ramey (2015) and directly model the underlying decision-making process.

[28] Project Vote Smart was discussed only twice in the *Congressional Record* during the years under analysis, and only by Reps. Henry Hyde and Alcee Hastings. Therefore, the likelihood of discussion of the survey or the organization affecting the estimates in any meaningful sense is effectively zero.

surveys, "respondents may have been concerned about social desirability, leaving them unwilling to admit to possessing certain traits" (201). Similarly, Project Vote Smart notes that "[m]ost candidates, fearing their opponents might use their positions in attack ads, refuse or only respond to a few questions that their consultants stamp as safe."[29] Indeed, this attitude is institutionalized within the major party leadership; in 2006, a Democratic leader in the Florida House of Representatives told a *Wall Street Journal* reporter that "We tell our candidates not to do it. It sets them up for a hit piece" (Grant 2006).

We therefore estimate a series of logistic regression models where the dependent variable equals 1 if the member in question ever responded to the NPAT, and 0 otherwise. This is arguably a very low bar, and therefore biases against finding significant results, as we minimize the amount of variance in the dependent variable. Thus, significant results should provide strong evidence of a relationship between personality and nonresponse. Independent variables include all personality traits, as well as host of other demographic information on legislators. These include dummy variables for gender (*Female*), region (*South*), whether or not legislators are *Leaders*, and whether or not they are in the *Majority Party*. Additionally, we control for district partisanship (*Democratic Normal Vote*), *Seniority*, electoral vulnerability (*Legislator Voteshare* in the last election), and *Ideological Extremism* (absolute value of their DW-NOMINATE score). Because we are estimating the likelihood of *ever* responding to the NPAT, each member has but one observation in our data; as such, the independent variables are actually set to the Congressmember-level means.[30]

As Table 2 shows, the propensity of nonresponse is a function of personality.[31] Perhaps most notably, Emotional Stability is positive and significant at conventional levels at all models, regardless of which controls are included. This finding provides strong evidence that the propensity to respond to surveys on which social desirability may be a factor is at least in part a function of personality traits.[32] Given that Dietrich et al. (2012) note that social desirability likely plays a role in the decision to respond (and, perhaps more worryingly, the revealed traits conditional on response), we have every reason to believe that data gathered through survey methods alone are inappropriate for investigating the relationships between personality traits and elite behavior.

Figure 5 makes these effects even clearer. In this figure, we plot the predicted probability of NPAT response varying Emotional Stability while holding other variables at their means (using the results from Model 4 in Table 2). As we see, increasing Emotional Stability from 2 SD below its mean to 2 SD above increases the predicted probability of response from about 30 to

[29] See http://votesmart.org/about/political-courage-test

[30] Project Vote Smart only reports a legislator's most recent survey. Thus, at best, we can only tell if a legislator in our data has *ever* responded to the survey. This necessitates our pooling strategy.

[31] Results are substantively similar if we include fixed effects for states in which the delegation size is larger than 5.

[32] While our goal in the present analysis is to demonstrate that personality traits predict survey nonresponse, we should also note that the pattern uncovered here also makes substantive sense. To wit, Klingler, Hollibaugh and Ramey (2015) show that higher levels of Emotional Stability are associated with lower levels of survey nonresponse among voters. Additionally, they demonstrate that this pattern is consistent with a theory wherein Emotional Stability captures the weight individuals place on potential negative outcomes for their actions; also see Ramey, Klingler and Hollibaugh (2015). Therefore, as responding to surveys about policy positions might "[set] them up for a hit piece," to borrow the words of the aforementioned Democratic leader in the Florida House, less Emotionally Stable legislators might be more likely to place undue weight on this possibility of the fallout that might ensue and therefore decline to respond to the survey. We can only assume that "hit pieces" would also be possible, and might even be more likely, if legislators were to admit to being disagreeable, unconscientious, or emotionally unstable.

TABLE 2 *Predicting Response to the National Political Awareness Test At Least Once*

	Model 1	Model 2	Model 3	Model 4
Openness	−0.15	−0.08	−0.14	−0.08
	(0.23)	(0.25)	(0.25)	(0.25)
Conscientiousness	0.17	0.26	0.38*	0.29
	(0.20)	(0.21)	(0.22)	(0.22)
Extraversion	−0.09	−0.10	−0.17	−0.20
	(0.14)	(0.15)	(0.16)	(0.16)
Agreeableness	−0.39	−0.53	−0.61	−0.57
	(0.39)	(0.40)	(0.41)	(0.41)
Emotional Stability	0.47*	0.49*	0.62**	0.58**
	(0.25)	(0.27)	(0.28)	(0.28)
Democratic Normal Vote	–	0.02**	0.02*	−0.02
		(0.01)	(0.01)	(0.01)
South	–	−0.17	−0.14	−0.13
		(0.19)	(0.19)	(0.20)
Female	–	0.22	0.19	0.10
		(0.26)	(0.26)	(0.27)
Seniority	–	−0.05*	−0.06**	−0.07**
		(0.03)	(0.03)	(0.03)
Legislator Voteshare$_{t-1}$	–	0.01	0.01	0.02*
		(0.01)	(0.01)	(0.01)
Leader	–	−0.49	−0.36	−0.60
		(0.91)	(0.91)	(0.90)
Ideological Extremism	–	–	−0.04*	1.08***
			(0.02)	(0.27)
Majority Party	–	–	–	−11.38***
				(2.76)
Constant	−0.12	−1.89	−1.72	4.79**
	(1.09)	(1.28)	(1.29)	(2.04)
BIC	881.23	869.07	871.99	860.51
Log likelihood	−421.32	−396.17	−394.44	−385.50
Number of observations	622	598	598	598

Note: Standard errors in parentheses. Observations are pooled by Congressmember.
BIC = Bayesian information criterion.
Two-tailed tests: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

about 53 percent. This increase is enormous and suggests estimates of personality from surveys of elites are highly susceptible to personality-induced selection bias.[33,34] The effect would almost certainly be stronger if we were modeling the rates of nonresponse to personality inventories, given that they are more personally intrusive than policy-specific surveys like the

[33] That said, surveys of legislators do have potential applications in the measurement of legislator—and, more generally, elite—personality traits. For starters, Mairesse et al.'s (2007) Personality Recognizer was trained on nonpolitical speech. One concern might be that, despite the similarity in word frequencies and distributions noted earlier, political and nonpolitical speech are fundamentally different from one another. In this case, elite surveys—coupled with spoken word transcripts—could be used to train elite- and/or domain-specific recognizers. Additionally, it might be worthwhile to use the personality scores to examine the relationships between legislator personality traits and the likelihood of response (or, alternatively, the types of responses) on personality surveys.

[34] While surveys *of* legislators may prove problematic for the study of elite personality, surveys of experts may prove to be a fruitful avenue for improving our machine learning approach. For example, Rubenzer and Faschingbauer (2004) administer a Big Five inventory to experts in order to measure personality traits for US presidents. One could use the expert ratings to retrain the machine learning algorithm discussed above, using these ratings in conjunction with presidential speeches. This would undoubtedly help to hone in on language pertinent to the political domain.

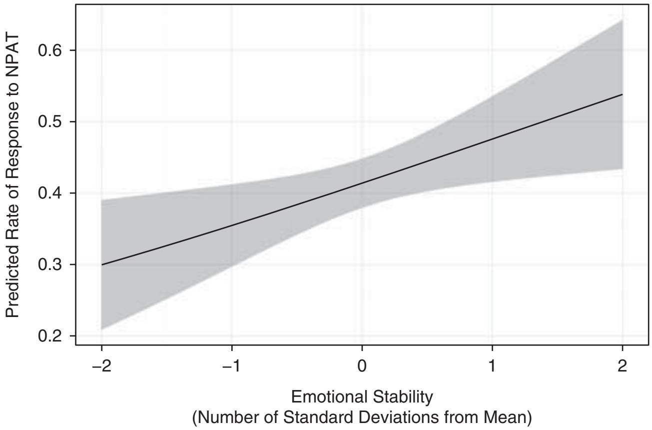

Fig. 5. Probability of responding to the National Political Awareness Test (NPAT)

NPAT and do not offer legislators the ability to take policy positions on important issues of the day. Therefore, more indirect measures of recovering legislators' personality traits are necessary. At its heart, that is the *raison d'être* for the text-based method presented here.

AN EMPIRICAL APPLICATION: PERSONALITY AND SYMBOLIC BILL PROPOSALS

Finally, we present a substantive application of these scores to the legislative domain—determining the types of bills legislators propose. Do they propose symbolic bills intended to rename post offices after local figures, or substantive bills geared toward effecting major policy changes? Different types of bills should reflect different underlying goals and should also require varying degrees of effort to successfully shepherd through the legislative process. However, different types of bills also face different hurdles, and more substantive bills are at greater risk of being prevented from reaching the floor due to partisan concerns (e.g., Aldrich 1995; Cox and McCubbins 2005). Given these hurdles, we expect more Conscientious legislators to propose fewer ceremonial/symbolic bills, as Conscientious individuals are described as having tendencies toward hard work, responsibility, and planning (VandenBos 2007), and they tend to be more driven, goal-oriented, uptight, organized, and have more willpower (Ozer and Benet-Martinez 2006).[35] They also tend to exhibit higher levels of performance on work-related and academic tasks (e.g., Barrick, Mount and Strauss 1993; Poropat 2009), are much more likely to set goals (Barrick, Mount and Strauss 1993), and are also more likely to have high levels of Grit, which is associated with perseverance and the ability to stick to long-term goals (Duckworth et al. 2007).

Lacking substantive importance, ceremonial/symbolic bills should be less controversial and therefore easier to pass. This suggests more Conscientious members should propose fewer ceremonial/symbolic bills as a proportion of total bills proposed. However, these effects should be conditional on the abilities of members of Congress to get their bills enacted into law; those in the majority should find it easier to have their preferred policies passed by the chamber. Thus, we should expect the effects of personality to interact with majority party status.

[35] Also see Ramey, Klingler and Hollibaugh (2015), who model Conscientiousness as a discount factor.

Members of the minority will have to work harder and expend more effort shaping policy-relevant bills to the liking of the majority; unless they are extremely forward-looking and have immense willpower—that is, unless they are highly Conscientious—they should propose more ceremonial/symbolic bills. More Conscientious members of the minority, however, should be less deterred by the partisan obstacles in their paths—as they place more weight on future payoffs—and should be similar to highly Conscientious members of the majority.

To estimate the importance of Conscientiousness, we analyzed all bills proposed in the House during the 104th–112th Congresses. Each bill was categorized using Volden and Wiseman's (2014) criteria as being ceremonial/symbolic or of substantive importance.[36] We then estimated a series of binomial logistic regressions; for these, the proportion of proposed bills that are ceremonial/symbolic in nature is the dependent variable.[37] The independent variables include our jackknifed personality scores, as well as several other independent variables that might affect the decisions. *Ideology* is parameterized as the member's first-dimension DW-NOMINATE score (Poole and Rosenthal 1997), and *Ideological Extremism* is the squared first-dimension DW-NOMINATE score. *Majority Party* is an indicator variable equaling 1 if the member and Speaker are of the same party, and 0 otherwise. *Seniority* denotes the number of terms a member has served. *Legislator Voteshare*$_{t-1}$ denotes the percentage of the vote (on a 0–100 scale) the member received in the election to the current Congress. *Committee Chair* and *Subcommittee Chair* are indicator variables equaling 1 if the member served in the relevant role in the Congress under analysis, and 0 otherwise. *Power Committee* is an indicator variable that equals 1 if the member in question sat on at least one of the three most powerful committees in the House—Appropriations, Rules, and Ways and Means—and 0 otherwise. Finally, we account for the importance of party leadership by including indicator variables for *Speaker*, *Majority Leadership*, and *Minority Leadership*, which are indicator variables equaling 1 if the member served in the relevant role in the Congress under analysis, and 0 otherwise. We account for previously established correlations between personality and personal demographics by including variables for *Age* and whether or not the member is *Female*; the former denotes the member's age at the end of the first session of each Congress, and the latter is an indicator variable equaling 1 if the member identifies as female and 0 otherwise. We also include interactions between personality and *Majority Party*.

Results are in Table 3, and Figure 6 presents the predicted proportions of ceremonial bill proposals, conditional on Conscientiousness and majority party status, along with 95 percent confidence intervals.[38] More Conscientious members propose fewer ceremonial/symbolic bills, instead dedicating their energies to more substantive arenas. Moreover, the effect is mitigated

[36] Volden and Wiseman classify bills as ceremonial/symbolic if any of the following occur in the title: "commemoration, commemorate, for the private relief of, for the relief of, medal, mint coins, posthumous, public holiday, to designate, to encourage, to express the sense of Congress, to provide for correction of, to name, to redesignate, to remove any doubt, to rename, and retention of the name" (2014, 21).

[37] We drop all observations where zero bills were proposed. This is because the binomial regression model takes as its dependent variable a matrix of trials and success, thus allowing for implicit weighting of proportions. However, this also means that rows with zero trials (and therefore zero successes) must be dropped from the analysis. This is not problematic, as standard count models (e.g., Poisson and negative binomial)—with logged offsets of one plus the total number of bills proposed per member-Congress dyad—recover substantively similar estimates. Nonetheless, as the rate of bill proposals overall are related to personality traits—a dynamic discussed by Ramey, Klingler and Hollibaugh (2015)—estimating the number of ceremonial bills proposed, as opposed to the rate, is prone to selection bias. It is for this reason (as well as the easier interpretation of the rates as opposed to raw counts) that we use binomial regression models instead of count models.

[38] For Figure 6, we use the results from Model 4 from Table 3. All continuous variables are set their means, and all categorical variables are set to their modes.

TABLE 3 *Personality and Ceremonial Bill Proposals*

	Model 1	Model 2	Model 3	Model 4	Model 5
Openness	0.21***	0.13*	0.18***	0.20*	0.17
	(0.06)	(0.07)	(0.07)	(0.10)	(0.10)
Conscientiousness	−0.52***	−0.43***	−0.39***	−0.53***	−0.58***
	(0.05)	(0.05)	(0.06)	(0.08)	(0.08)
Extraversion	0.10***	0.03	0.02	0.19***	0.22***
	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)
Agreeableness	0.24**	0.18*	0.13	0.20	0.28*
	(0.10)	(0.10)	(0.11)	(0.16)	(0.15)
Emotional Stability	0.11*	0.19***	0.19***	0.24**	0.21**
	(0.07)	(0.07)	(0.07)	(0.10)	(0.10)
Majority Party	–	−0.24***	−0.18***	1.35**	1.40**
		(0.05)	(0.06)	(0.59)	(0.60)
Majority Party × Openness	–	–	–	−0.07	−0.07
				(0.14)	(0.14)
Majority Party × Conscientiousness	–	–	–	0.27**	0.29***
				(0.11)	(0.11)
Majority Party × Extraversion	–	–	–	−0.33***	−0.35***
				(0.08)	(0.08)
Majority Party × Agreeableness	–	–	–	−0.18	−0.23
				(0.21)	(0.21)
Majority Party × Emotional Stability	–	–	–	−0.09	−0.05
				(0.14)	(0.14)
Ideology	–	−0.27***	−0.24***	−0.27***	−0.28***
		(0.05)	(0.06)	(0.06)	(0.06)
Ideological Extremism	–	0.02	−0.14	−0.12	−0.00
		(0.13)	(0.14)	(0.14)	(0.14)
Age	–	–	0.01**	0.01**	0.01***
			(0.00)	(0.00)	(0.00)
Female	–	–	−0.22***	−0.23***	−0.25***
			(0.07)	(0.07)	(0.07)
Seniority	–	–	−0.03***	−0.03***	−0.03***
			(0.01)	(0.01)	(0.01)
Legislator Voteshare$_{t-1}$	–	–	0.01***	0.01***	0.00***
			(0.00)	(0.00)	(0.00)
Committee Chair	–	–	−0.29**	−0.31**	−0.32**
			(0.12)	(0.12)	(0.12)
Subcommittee Chair	–	–	−0.15**	−0.16**	−0.16**
			(0.07)	(0.07)	(0.07)
Power Committee	–	–	−0.07	−0.08	−0.09
			(0.06)	(0.06)	(0.06)
Speaker	–	–	0.50	0.62	0.53
			(0.62)	(0.62)	(0.62)
Majority Leadership	–	–	0.11	0.13	0.16
			(0.16)	(0.16)	(0.16)
Minority Leadership	–	–	0.32**	0.30**	0.33**
			(0.15)	(0.15)	(0.15)
Constant	−3.67***	−3.34***	−4.18***	−4.83***	−5.08***
	(0.28)	(0.30)	(0.35)	(0.45)	(0.46)
Congress FE?	No	No	No	No	Yes
Wald test	158.82***	132.28***	104.90***	121.81***	143.22***
BIC	7953.04	7888.17	7784.17	7799.81	7770.90
Log likelihood	−3951.81	−3907.03	−3814.09	−3801.38	−3754.08
Number of observations	3736	3732	3648	3648	3648

Note: Standard errors in parentheses. Observations are at the Congressmember level. Null hypotheses for the Wald tests are that all coefficients related to the personality traits are 0.
BIC = Bayesian information criterion.
Two-tailed tests: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

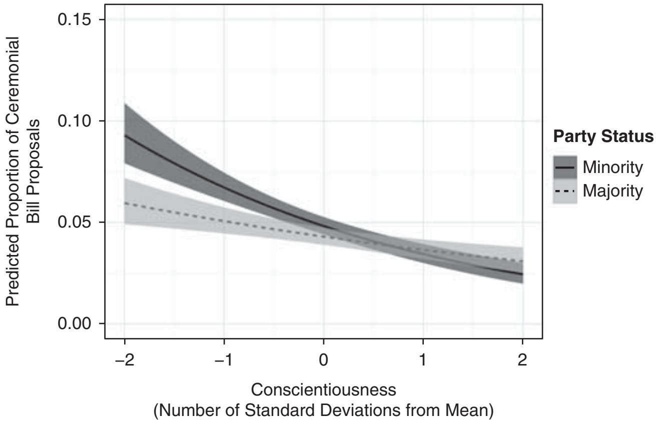

Fig. 6. Conscientiousness and symbolic bill proposals

by structural factors. For example, Conscientiousness seems to least affect those in the majority—that is, those most likely to see bills they agree with pass without their own influence. These members will be able to free ride on the efforts of the majority party leadership with some assurance that substantive bills with which they agree will come to the floor. Conversely, those in the minority will be more inclined to propose substantive bills—as there will be fewer fellow members of the House on whose efforts they can free ride—so long as they are sufficiently Conscientious.

CONCLUSION

This paper demonstrates measurement of elite personality is possible and that estimates are consistent and robust over time. We have also demonstrated that survey methods for recovering measures of elite personality are problematic at best and that the Big Five do indeed play a role in legislative decisionmaking. The next step, incorporating these measures into theoretically motivated studies of elite behavior, requires careful consideration of how personality fits into standard frameworks of elite decisionmaking. These efforts are underway (e.g., Hall 2015; Hollibaugh, Ramey and Klingler 2015; Ramey, Klingler and Hollibaugh 2015).

There are several ways in which the methods discussed in this paper may be extended in future work. First of all, while we believe text-based measures of elite personality are more appropriate for the study of legislative behavior than survey-based measures, there are a number of ways in which the strengths of survey-based methods may be used to enrich the project, as discussed above. Additionally, while the *Congressional Record* corpus is very similar to the Pennebaker corpus on the relevant dimensions, comparable sources are not available for many political elites of interest, including challengers in congressional races. Training a model on alternate text sources (e.g., social media corpora, press releases, etc.) would allow for the estimation of personality for many other actors. Additionally, digitizing the pre-1996 Congressional Record will allow for the estimation of personality traits for a wide variety of historical figures of interest.[39]

[39] This, of course, is conditional on the distribution of linguistic features being reasonably similar to that in the training set. This is yet to be determined.

This ability to generate measures of personality from speech offers scholars a powerful tool for the investigation of the Big Five personality traits among elites. For the first time, personality scores within the five-factor model are available for virtually all members of the US Congress going back to 1996. With these tools, established findings from personality psychology and the growing literature on personality in political behavior may be applied to develop new theories of legislative behavior. Furthermore, there is an opening for a modeling framework that would allow the Big Five to be incorporated into models of political institutions. The personality scores introduced in this paper now allow the implications of such models to be examined empirically.

In sum, text-based measures of legislator personality make significant improvements over existing measures derived from expert ratings or personality inventories administered through surveys. They represent a significant innovation and serve to fill a critical gap in the burgeoning literature on elite behavior in institutions.

REFERENCES

Aldrich, John H. 1995. *Why Parties?: The Origin and Transformation of Political Parties in America*. Chicago, IL: University of Chicago Press.

Alford, John R., and John R. Hibbing. 2007. 'Personal, Interpersonal, and Political Temperaments'. *The Annals of the American Academy of Political and Social Science* 614(1):196–212.

Ansolabehere, Stephen, James M. Snyder Jr., and Charles Stewart III. 2001. 'The Effects of Party and Preferences on Congressional Roll-Call Voting'. *Legislative Studies Quarterly* 26(4):533–72.

Barlett, Christopher P., and Craig A. Anderson. 2012. 'Direct and Indirect Relations Between the Big 5 Personality Traits and Aggressive and Violent Behavior'. *Personality and Individual Differences* 52(8):870–75.

Barrick, Murray R., Michael K. Mount, and Judy P. Strauss. 1993. 'Conscientiousness and Performance of Sales Representatives: Test of the Mediating Effects of Goal Setting'. *Journal of Applied Psychology* 78(5):715–22.

Battista, James Coleman, Michael Peress, and Jesse Richman. 2012. 'Common-Space Ideal Points, Committee Assignments, and Financial Interests in the State Legislatures'. *State Politics & Policy Quarterly* 13(1):70–87.

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. New York: Springer.

Borgatta, Edgar F. 1964. 'The Structure of Personality Characteristics'. *Behavioral Science* 9(1):8–17.

Caprara, Gian Vittorio, Claudio Barbaranelli, and Philip G. Zimbardo. 2002. 'When Parsimony Subdues Distinctiveness: Simplified Public Perceptions of Politicians' Personality'. *Political Psychology* 23(1):77–95.

Caprara, Gian Vittorio, Shalom Schwartz, Cristina Capanna, Michele Vecchione, and Claudio Barbaranelli. 2006. 'Personality and Politics: Values, Traits, and Political Choice'. *Political Psychology* 27(1): 1–28.

Coltheart, Max. 1981. 'The MRC Psycholinguistic Database'. *The Quarterly Journal of Experimental Psychology: Section A (Human Experimental Psychology)* 33(4):497–505.

Congressional Management Foundation and Society for Human Resource Management. 2013. 'Life in Congress: Job Satisfaction and Engagement of House and Senate Staff'. Available at http://www.congressfoundation.org/publications/life-in-congress, accessed 7 July 2015.

Costa, Paul T. Jr., and Robert R. McCrae. 1992. 'Normal Personality Assessment in Clinical Practice: The NEO Personality Inventory'. *Psychological Assessment* 4(1):5–13.

Cox, Gary W., and Mathew D. McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the U.S. House of Representatives*. New York: Cambridge University Press.

Crespin, Michael H., and David W. Rohde. 2010. 'Dimensions, Issues, and Bills: Appropriations Voting on the House Floor'. *Journal of Politics* 72(4):976–89.

Dietrich, Bryce J., Scott Lasley, Jeffery J. Mondak, Megan L. Remmel, and Joel Turner. 2012. 'Personality and Legislative Politics: The Big Five Trait Dimensions Among U.S. State Legislators'. *Political Psychology* 33(2):195–210.

Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly. 2007. 'Grit: Perseverance and Passion for Long-Term Goals'. *Journal of Personality and Social Psychology* 92(6):1087–101.

Duckworth, Angela L., Eli Tsukayama, and Teri A. Kirby. 2013. 'Is it Really Self-Control? Examining the Predictive Power of the Delay of Gratification Task'. *Personality and Social Psychology Bulletin* 39(7):843–55.

Eaves, Lindon J., and J. Eysenck Hans. 1974. 'Genetics and the Development of Social Attitudes'. *Nature* 249(454):288–89.

Efron, Bradley, and Robert J. Tibshirani. 1994. *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press.

Gelman, Andrew, Gary King, and Chuanhai Liu. 1998. 'Not Asked and Not Answered: Multiple Imputation for Multiple Surveys'. *Journal of the American Statistical Association* 93(443):846–57.

Gerber, Alan S., Gregory A. Huber, David Doherty, and Conor M. Dowling. 2011a. 'The Big Five Personality Traits in the Political Arena'. *Annual Review of Political Science* 14(1):265–87.

Gerber, Alan S., Gregory A. Huber, David Doherty, and Conor M. Dowling. 2011b. 'Personality Traits and the Consumption of Political Information'. *American Politics Research* 39(1):32–84.

Gerber, Alan S., Gregory A. Huber, David Doherty, Conor M. Dowling, and Shang E. Ha. 2010. 'Personality and Political Attitudes: Relationships Across Issue Domains and Political Contexts'. *American Political Science Review* 104(1):111–33.

Gosling, Samuel D., Peter J. Rentfrow, and William B. Swann Jr. 2003. 'A Very Brief Measure of the Big-Five Personality Domains'. *Journal of Research in Personality* 37(6):504–28.

Grant, Peter. 2006. 'Politicians Grow Wary of Survey as Internet Spreads Attack Ads'. *The Wall Street Journal*, 25 October, p. B1.

Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. 'The WEKA Data Mining Software: An Update'. *ACM SIGKDD Explorations Newsletter* 11(1):10–18.

Hall, Matthew E. K. 2015. 'Judging with Personality: The Justices' Personality Traits and Decision Making on the U.S. Supreme Court'. Manuscript.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* 2nd ed. New York: Springer.

Heckman, James J. 1976. 'The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models'. *Annals of Economic and Social Measurement* 5(4):475–92.

Hollibaugh, Gary E., Jr., Adam J. Ramey, and Jonathan D. Klingler. 2015. 'Welcome to the Machine: A Model of Legislator Personality and Communications Technology Adoption'. Available at http://dx.doi.org/10.2139/ssrn.2629657, accessed 7 July 2015.

Hornik, Kurt, Christian Buchta, and Achim Zeileis. 2009. 'Open-Source Machine Learning: R Meets Weka'. *Computational Statistics* 24(2):225–32.

Jost, John T., Brian A. Nosek, and Samuel D. Gosling. 2008. 'Ideology: Its Resurgence in Social, Personality, and Political Psychology'. *Perspectives on Psychological Science* 3(2):126–36.

Jovanović, Dragan, Krsto Lipovac, Predrag Stanojević, and Dragana Stanojević. 2011. 'The Effects of Personality Traits on Driving-Related Anger and Aggressive Behaviour in Traffic Among Serbian Drivers'. *Transportation Research Part F: Traffic Psychology and Behaviour* 14(1):43–53.

Kandler, Christian, Wiebke Bleidorn, and Rainer Riemann. 2012. 'Left or Right? Sources of Political Orientation: The Roles of Genetic Factors, Cultural Transmission, Assortative Mating, and Personality'. *Journal of Personality and Social Psychology* 102(3):633–45.

Karush, William. 1939. 'Minima of Functions of Several Variables with Inequalities as Side Constraints'. MSc Thesis, University of Chicago, Chicago, IL.

King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 2001. 'Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation'. *American Political Science Review* 95(1):49–69.

Klingler, Jonathan D., Gary E. Hollibaugh, Jr., and Adam J. Ramey. 2015. 'Don't Know What You Got: A Bayesian Hierarchical Model of Neuroticism and Nonresponse'. Available at http://dx.doi.org/10.2139/ssrn.2608719, accessed 7 July 2015.

Küfner, Albrecht C. P., Mitja D. Back, Steffen Nestler, and Boris Egloff. 2010. 'Tell Me a Story and I Will Tell You Who You Are! Lens Model Analyses of Personality and Creative Writing'. *Journal of Research in Personality* 44(4):427–35.

Kuhn, Harold W., and Albert W. Tucker. 1951. 'Nonlinear Programming'. In Jerzy Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 481–92. Berkeley, CA: University of California Press.

Lowe, Will, and Kenneth Benoit. 2011. 'Estimating Uncertainty in Quantitative Text Analysis'. Available at http://www.kenbenoit.net/pdfs/Midwest_2011_Lowe_Benoit.pdf, accessed 7 July 2015.

Mairesse, Francois, and Marilyn A. Walker. 2008. 'Trainable Generation of Big-Five Personality Styles Through Data-Driven Parameter Estimation'. In Johanna D. Moore, Simone Teufel, James Allan, and Sadaoki Furui (eds.), *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 165–73. Madison, WI: Omnipress.

Mairesse, Francois, Marilyn A. Walker, Matthias R. Mehl, and Roger K. Moore. 2007. 'Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text'. *Journal of Artificial Intelligence Research* 30(1):457–500.

McCrae, Robert R., and Oliver P. John. 1992. 'An Introduction to the Five-Factor Model and its Applications'. *Journal of Personality* 60(2):175–215.

Mehl, Matthias R., Samuel D. Gosling, and James W. Pennebaker. 2006. 'Personality in its Natural Habitat: Manifestations and Implicit Folk Theories of Personality in Daily Life'. *Journal of Personality and Social Psychology* 90(5):862–77.

Mondak, Jeffery J. 2010. *Personality and the Foundations of Political Behavior*. New York: Cambridge University Press.

Mondak, Jeffery J., and Karen D. Halperin. 2008. 'A Framework for the Study of Personality and Political Behaviour'. *British Journal of Political Science* 38(2):335–62.

Mondak, Jeffery J., Matthew V. Hibbing, Damarys Canache, Mitchell A. Seligson, and Mary R. Anderson. 2010. 'Personality and Civic Engagement: An Integrative Framework for the Study of Trait Effects on Political Behavior'. *American Political Science Review* 104(1):85–110.

Muris, Peter, Jeffrey Roelofs, Eric Rassin, Ingmar Franken, and Birgit Mayer. 2005. 'Mediating Effects of Rumination and Worry on the Links Between Neuroticism, Anxiety and Depression'. *Personality and Individual Differences* 39(6):1105–111.

Neale, Thomas H. 1998. 'Speechwriting in Perspective: A Brief Guide to Effective and Persuasive Communication'. *CRS Report for Congress*, 98–170 GOV. Available at http://digital.library.unt.edu/ark:/67531/metacrs581/m1/1/high_res_d/98-170gov_1998Feb25.pdf.

Ozer, Daniel J., and Veronica Benet-Martinez. 2006. 'Personality and the Prediction of Consequential Outcomes'. *Annual Review of Psychology* 57(1):401–21.

Pennebaker, James W., and Laura A. King. 1999. 'Linguistic Styles: Language Use as an Individual Difference'. *Journal of Personality and Social Psychology* 77(6):1296–312.

Pennebaker, James W., Martha E. Francis, and Roger J. Booth. 2001. *Linguistic Inquiry and Word Count: LIWC 2001*. Mahwah, NJ: Lawrence Erlbaum Associates.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.

Poropat, Arthur E. 2009. 'A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance'. *Psychological Bulletin* 135(2):322–38.

Quinlan, John R. 1992. 'Learning with Continuous Classes'. In Anthony Adams and Leon Sterling (ed.), *AI '92: Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, 343–48. Singapore: World Scientific.

Ramey, Adam J., Jonathan D. Klingler, and Gary E. Hollibaugh Jr. 2015. 'More Than a Feeling: Personality, Polarization, and the Transformation of the U.S. Congress'. Manuscript.

Rubenzer, Steven J., and Thomas R. Faschingbauer. 2004. *Personality, Character, and Leadership in the White House: Psychologists Assess the Presidents*. Sterling, VA: Potomac Books Inc.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Saucier, Gerard, and Lewis R. Goldberg. 1996. 'The Language of Personality: Lexical Perspectives on the Five-Factor Model'. In Jerry S. Wiggins (ed.), *The Five-Factor Model of Personality: Theoretical Perspectives*, 363–84. New York: Guilford Press.

Schuller, Björn, Stefan Steidl, Anton Batliner, Felix Burkhardt, Laurence Devillers, Christian Muller, and Shrikanth Narayanan. 2013. 'Paralinguistics in Speech and Language—State-of-the-Art and the Challenge'. *Computer Speech & Language* 27(1):4–39.

Shevade, Shirish K., S. Sathiya Keerthi, Chiranjib Bhattacharyya, and K. R. Krishna Murthy. 1999. 'Improvements to the SMO Algorithm for SVM Regression'. *IEEE Transactions on Neural Networks* 11(5):1188–193.

Shor, Boris, and Nolan McCarty. 2011. 'The Ideological Mapping of American Legislatures'. *American Political Science Review* 105(3):530–51.

Sigelman, Lee. 2002. 'Two Reagans? Genre Imperatives, Ghostwriters, and Presidential Personality Profiles'. *Political Psychology* 23(4):839–51.

Silvester, Jo, Madeleine Wyatt, and Ray Randall. 2014. 'Politician Personality, Machiavellianism, and Political Skill as Predictors of Performance Ratings in Political Roles'. *Journal of Occupational and Organizational Psychology* 87(2):258–79.

Smola, Alex J., and Bernhard Schölkopf. 2004. 'A Tutorial on Support Vector Regression'. *Statistics and Computing* 14(3):199–222.

Tupes, Ernest C., and Raymond E. Christal. 1992. 'Recurrent Personality Factors Based on Trait Ratings'. *Journal of Personality* 60(2):225–51.

VandenBos, Gary R. 2007. *APA Dictionary of Psychology*. Washington, DC: American Psychological Association.

Verhulst, Brad, Lindon J. Eaves, and Peter K. Hatemi. 2012. 'Correlation Not Causation: The Relationship Between Personality Traits and Political Ideologies'. *American Journal of Political Science* 56(1):34–51.

Verhulst, Brad, Peter K. Hatemi, and Nicholas G. Martin. 2010. 'The Nature of the Relationship Between Personality Traits and Political Attitudes'. *Personality and Individual Differences* 49(4):306–16.

Vetterli, Charles F., and John J. Furedy. 1997. 'Correlates of Intelligence in Computer Measured Aspects of Prose Vocabulary: Word Length, Diversity, and Rarity'. *Personality and Individual Differences* 22(6):933–35.

Volden, Craig, and Alan E. Wiseman. 2014. *Legislative Effectiveness in the United States Congress: The Lawmakers*. New York: Cambridge University Press.

Weisberg, Herbert F. 2009. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago, IL: University of Chicago Press.