Text as Behavior

Omar Wasow*

December 20, 2023

Abstract

Text analysis typically focuses on content—such as sentiment or topic—but writing is also a form of effortful action. Building on this insight, I propose using simple features of open-ended writing tasks to study *text as behavior*. This approach treats writing as cognitively and emotionally "costly" for subjects but inexpensive for researchers. I show basic statistics like the number of characters can approximate effort and significantly improve models estimating quantities of interest, from the probability of turning out to vote, to psychological states about which a subject may not be fully aware. Further, these methods can convert nonresponse into informative data; validate survey instruments; serve as mechanism checks; be hard for a subject to "game"; work across different languages and analogize well to real-world situations. In sum, text as behavior can help address a range of issues related to quantifying attitudes and actions.

Keywords: Text analysis, text as data, political behavior, survey methods

Word Count: 9,953

*Department of Political Science, UC Berkeley, owasow@berekely.edu. This is a draft, constructive feedback welcome. I thank Risa Gelles-Watnick and Christina Im for excellent research assistance. I thank Brandon McGhee for sharing data, and Andrew Little, Elena Llaudet and Alexander Agadjanian for helpful feedback.

Writing is hard. Linguist Peter Hugoe Matthews (2003) said of humanity, "No one would describe us as the 'writing' species." At the same time, writing is a pervasive part of contemporary life. The rising ubiquity of writing follows long-term educational and technological trends. In 1900, only one in five people in the world was literate. By the end of the twentieth century, only one in five people was *illiterate* (van Zanden et al. 2014). Further, as Thompson (2010) notes, "This generation doesn't make phone calls, because everyone is in constant, lightweight contact in so many other ways: texting, chatting, and social-network messaging." In contrast to Matthews' statement two decades ago, now, for many modern, connected and educated people, "Homo scribens" might actually be an appropriate description.

Quantifying sincere attitudes and preferences is also hard and a central challenge for social science (Campbell and Stanley 1963). Researchers often attempt to gauge affect and behavior through games or by attaching costs or rewards to an action in an effort to encourage subjects to reveal genuine tendencies and tastes (McDermott 2002). These approaches, though, have limitations. Paying subjects additional money to reveal preferences can be expensive. Economic games, such as the "Dictator Game," may not translate well to real-world situations in which subjects are not participants in a study (Winking and Mizer 2013). In addition, many measures suffer from concerns about external validity (Findley, Kikuta, and Denly 2021). Other kinds of behavioral measures, such as asking subjects to "submit their email address to 'sign' a petition," may be insufficiently costly or a poor approximation of real-world behavior. Another concern is that a strategic subject might attempt to infer the goals of the study which could induce researcher demand effects (Mummolo and Peterson 2019).

Given the nontrivial cognitive and affective load of writing, I propose using simple metadata from open-ended writing tasks—like nonresponse and number of characters—as alternative measures of effortful action. The nearly universal difficulty of writing combined with its increasing prevalence allows for a measure that is "costly" for subjects but inexpensive for researchers. Open-ended prompts can also be evaluated in numerous ways that are unlikely to be deduced by subjects and, in many cases, these tasks correspond neatly to many real-world behaviors. I test this "text as behavior" method across a range of questions, data sets and languages with three simple features of open-ended writing tasks: the number of characters, nonresponse and time. In contrast to methods like sentiment analysis that attempt to extract meaning from terms, I use metadata, like the number of characters, to determine typically unobserved qualities like effort, intensity of feeling or ambivalence. I find these simple features of text provide meaningful signals of subject attitudes and behaviors. All other factors being equal, subjects who write more generally reveal significantly more intense support, engagement and capacity for relevant action related to that particular topic as compared with subjects who write less. Likewise, all things being equal, subjects who fail to enter a single character in response to an open-ended text prompt—nonresponders—are much more likely to exhibit negative attitudes and conflicted behaviors as compared to subjects who write at least one character. Further, I present evidence that different sorts of prompts and/or experimental manipulations may be useful for eliciting or confirming different types of sincere attitudes and behaviors.

This approach offers multiple advantages over many current and more complicated methods of measuring attitudes and behavior. First, the use of open-ended prompts is already widespread and growing in social science (Li 2023). Second, text responses—whether collected in surveys or via services like Twitter—can offer researchers good equivalence to real-world situations, often called "mundane realism" (Aronson and Carlsmith 1968). In modern life, writing short messages is nearly ubiquitous on phones, in social media and via work platforms like Slack. Third, the non-trivial demands of writing allow measurement of otherwise hard to quantify traits such as the degree of commitment to voting. Fourth, writing can offer a window into psychological states about which a subject may not be fully aware (Wilson 2004). Fifth, nonresponse can be interpreted as informative rather than as missing data. Sixth, the open-ended nature of text response and the nearly infinite number of ways to measure text make it harder for a subject to "game" and therefore can serve as means to validate other measures for which there might be concerns about response bias, such as social desirability bias. Seventh, any measure of human attitudes or behavior suffers some risk of subjects inferring the study's purpose. This may result in insincere behavior or could induce researcher demand effects (Mummolo and Peterson 2019). With open-ended text responses, it is typically not obvious what is being measured or how a strategic subject might, in turn, adjust their behavior. Finally, I show using text as behavior can plausibly work across languages, transcription, modes and translation.

Related Work

The use of text as data is now widespread in social science (Grimmer, Roberts, and Stewart 2022; Li 2023). Researchers commonly use methods like topic models to rapidly categorize open-ended text responses in surveys (Roberts et al. 2014), or forms of sentiment analysis using tools like dictionaries of affect (c.f., Mossholder et al. 1995; Taboada et al. 2011; Bisgaard 2019). Using text as a measure of behavior or as a proxy for a person's state of mind, however, is less typical though has been done in a variety of fields and forms.

One common approach uses terms as a measure of emotions or mental states (Pennebaker, Francis, and Booth 2001; Kramer, Guillory, and Hancock 2014; Gillion 2016). Rude, Gortner, and Pennebaker (2004), for example, find the use of "I/me/my" terms and increased use of negatively valenced words is associated with greater depression in college students. Kuo, Malhotra, and Mo (2017) experimentally induced feelings of exclusion and, among other measures, asked subjects to write lists of items they liked and disliked about both the Democratic and Republican Parties. Of particular relevance, Kuo, Malhotra, and Mo (2017) note, "This task required a great deal of effort on the part of respondents and therefore can be interpreted as a behavioral manifestation of liking or aversion toward a political party" (27).

Another approach uses text recorded in transcripts, court cases, social media, short messaging services and Internet searches as possible indicators of bias that may circumvent efforts to offer socially acceptable answers. Data such as text logs or surveilled and transcribed speech can offer an unvarnished, non-survey alternative indicator of attitudes. Google search queries, for example, have been used to measure levels of racial and gender bias across regions (Stephens-Davidowitz 2014; Maloney 2021). Correspondence studies conducted via email or text message offer another method of detecting bias solely via written responses (Butler and Broockman 2011; Lowande and Proctor 2020; Yan and Bernhard 2023). Transcribed deliberative discussions, Supreme Court interruptions and police stops have been used to measure gender and racial bias (Karpowitz, Mendelberg, and Shaker 2012; Jacobi and Schweers 2017; Voigt et al. 2017).

Aggregated posts on social media have also been used to measure attitudes of the mass public and predict a range of future behaviors from voting in elections to stock market movements, movie attendance and disruptive events (c.f., Tumasjan et al. 2010; Bollen, Mao, and Zeng 2011; Alsaedi, Burnap, and Rana 2017). Social media sites have also been used to observe interactions in which exchanges are used to detect dynamics of status hierarchies and polite or conflictual conversations (Zhang et al. 2018; Danescu-Niculescu-Mizil et al. 2013; Panteli 2002). Similarly, social media interactions have also been used to conduct experiments in which both treatments and outcomes are short written exchanges (Munger 2017; Mosleh et al. 2021).

Writing tasks are also routinely used in psychology-related studies as a form of treatment to induce different states of mind, from reducing anxiety and trauma to increasing awareness of certain moral frames (Pennebaker 1997; Day et al. 2014). Other applications of text analysis on open-ended responses include validating and extending survey instruments (ten Kleij and Musters 2003), serving as a type of mechanism check (Collins 2021; Kuo, Malhotra, and Mo 2017) and detecting inattentive subjects (Ziegler 2022). Metadata like character length has also been used to assess potential differences in mode effects between surveys conducted by paper and online (Denscombe 2008).

Text as behavior

Building on prior work, I propose that much of this broad range of scholarship can usefully be understood under the category of "text as behavior." Text as behavior is a subset of text as data but with particular attention to cases in which writing or, in some cases, transcriptions of speech can be understood as a form of moderately costly action and, consequently, can help to reveal, validate or predict attitudes, preferences and behaviors. Two primary insights undergird the text as behavior approach: (1) writing is often cognitively and affectively demanding; (2) writing tasks provide a partial window into often inaccessible or difficult to articulate thoughts and feelings.

On the cognitive and affective demands of writing, Hayes (1996) details the complex range of capacities that must be coordinated and executed, from visual and motor skills to short-term memory and language abilities. Kellogg (1999) argues writing does "not simply unfold automatically and effortlessly in the manner of a well learned motor skill ... writing anything but the most routine and brief pieces is the mental equivalent of digging ditches" (17, quoted in Graham (2018)). Further, writing is not simply cognitively challenging but often emotionally hard, too. Consider, for example, sympathy cards with pre-written inscriptions to help solve the problem of conveying an emotion when one feels at a loss for words. Reviewing several decades of psychology research on the relationship between feelings and thoughts, Wright (2017) concludes there is a "fine entanglement of affect and cognition" (120). In short, writing tasks are often modestly taxing for subjects and, as a result, potentially useful for researchers as a behavioral measure that reduces the likelihood of "cheap talk."

Writing tasks may also provide insight to the numerous nonconscious perceptual systems used to make sense of and interpret the world (Wilson 2004). Haidt (2012) offers a useful metaphor, suggesting the mind is divided like "a rider on an elephant." The rider is that about which we are aware, our conscious reasoning, while the elephant is "the other 99 percent of mental processes—the ones that occur outside of awareness but that actually govern most of our behavior" (xxi). Writing techniques like journaling, free association and automatic writing have all been suggested as methods to surface nonconscious thoughts. Though researchers have developed many creative instruments and games to reveal otherwise subterranean thoughts and feelings, writing as a window into nonconscious thought processes remains underutilized in social science (Roberts et al. 2014). Table 1 presents an overview of six analyses across which I demonstrate a number of related applications of text as behavior.

Data and Methods

The analyses below rely on two large public data sets and data from two smaller studies. The public data sets are the 2016 American National Election Study, and 2016 Afrobarometer. The larger data sets are described below while the data used in the two other analyses are discussed in more detail within Studies 3a and 3b.

American National Election Study

Studies 1a, 1b and 2a rely on the 2016 American National Election Study (ANES) survey of a cross-section of eligible United States voters both before and after the 2016 election. The 2016 ANES investigated a broad range of questions including public opinion, voting behavior and media exposure. The survey was conducted with both a face-to-face sample (N = 2,238) and an Internet sample (N = 5,680). In addition to a large battery of multiple choice-style survey questions, some open-ended prompts were also administered. Of

Table 1: Overview of Study Types, Questions, Data, Measures and Models

Study	Type	Data	Text Measures	Models
1a.	Predict survey response	ANES (2016)	Character count and nonresponse	Vote Choice $\sim$ Text + $\mathbf{X}$
1b.	Predict behavior	ANES (2016)	Character count and nonresponse	Turnout $\sim$ Text + $\mathbf{X}$
2a.	Monolingual instrument validation	ANES (2016)	Nonresponse	Text $\sim$ Hostile Sexism + $\mathbf{X}$ Text $\sim$ Racial Resentment + $\mathbf{X}$
2b.	Multilingual instrument validation	Afrobarometer (2016)	Character count	Text $\sim$ Democracy View + $\mathbf{X}$
3a.	Manipulation Check	Kuo, Malhotra, and Mo (2017)	Character count and time	Time $\sim$ Exclusion Treatment
3b.	Manipulation Check	McGhee (2018)	Character count and nonresponse	Text $\sim$ Cross-pressure Treatment

specific interest were four questions that provided open-ended prompts to subjects with questions along the lines, “Is there anything in particular about Hillary Clinton that might make you want to vote against her?” The four questions each prompted for affect (like or dislike) and one of the two major party nominees (Democratic or Republican).

In addition, data from two more sources were used. First, the 2016 ANES included four open-ended prompts in which subjects were asked three times, “What do you think are the most important problems facing this country?” and, once, “Which among mentions is the most important problem?” Second, the 2016 ANES provided supplemental data that validated actual turnout using publicly available voter files (Enamorado and Imai 2017). For Study 1b, metadata about the most important problems is used to predict validated voter turnout along with the candidate-affect questions mentioned earlier

Afrobarometer

Study 2b tests whether text as behavior methods might work across multiple languages using data from the 2016 Afrobarometer study. The Afrobarometer describes itself as “a pan-African, non-partisan survey research network that conducts public attitude surveys on democracy, governance, the economy, and society.” The 2016 Afrobarometer surveyed subjects in 36 countries in Africa. A total of 53,921 subjects were interviewed in their native languages and results were then translated into English (N=34,838), French (N=14,116), or Portuguese (N=4,693). To test whether metadata about text could validate a survey instrument across languages, transcription and translation, I test whether responses to open-ended prompts asking, “What, if anything, does ‘democracy’ mean to you?” to validate a separate battery of survey questions about the importance of democracy.

Defining Measures

Across all studies I use one or more of three measures: the number of characters, nonresponse (i.e., zero characters) and/or time to completion of relevant writing tasks. Across most analyses, when multiple questions are related, I pool congruent responses and in some cases upweight informative nonresponse. These combined scales are explained in more detail within each study. To simplify notation in those scales, I define two basic functions below, one for the number of characters and another for nonresponse.

Let s be a text string. I define the function $nchar(s)$ as:

$$nchar(s) = \text{number of characters in } s$$

Further, I define a *nonresponse* function $non(s)$ that tests if $nchar(s)$ is equal to zero:

$$non(s) = \begin{cases} 1 & \text{if } nchar(s) = 0 \\ 0 & \text{otherwise} \end{cases}$$

I opt for the number of characters rather than number of words or stemmed terms because the basic underlying assumption of this approach is that writing is a kind of effortful work and, therefore, each keystroke can be thought of as the most granular measure of exertion expended by a subject. Further, other plausible measures—such as counting terms—likely discard information (e.g., `jobs" versus`unemployment''). Finally, as shown in Table 1, I deliberately use a variety of text metadata and combined scales, both as explanatory variables and outcomes, to try and demonstrate the versatility of this method.

Study 1a: Text as Predictor of Vote Choice

Can text metadata, independent of content, predict an outcome of interest like vote choice? Using the 2016 ANES, I begin with a simple test attempting to predict self-reported support for Hillary Clinton or Donald Trump using only the number of characters from two open-ended prompts asking is there something that would make the respondent vote for each of the main party presidential nominees. Figure 1 presents a diagram of how the candidate-affect writing prompts are hypothesized to influence subjects and the amount they write.

Figure 2 presents two plots showing how the number of characters written in response to one open-ended candidate-affect question predicts self-reported vote choice, by party identification, holding a large number of individual-level demographic and attitudinal variables constant. In Panels A and B, a subject writing nothing has about a 10 to 60% chance of self-reporting support for the relevant candidate, depending on party. As the

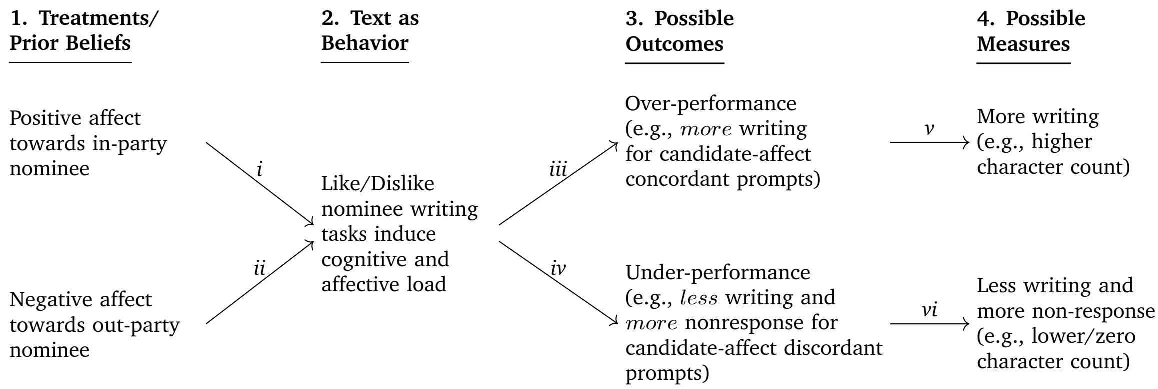


Figure 1: Hypothesized model of how (1) experimental treatments and/or prior beliefs might combine to (2) induce complicated cognitive and affective reactions following a writing prompt about what subjects like/dislike about their respective in- and out-party nominees, (3) which potentially produces two distinct outcomes — writing more or less/not-at-all — that can be detected by measuring, (4) total number of characters written in candidate-affect concordant questions and nonresponse.

number of characters rises, the predicted likelihood of support increases substantially and, at 250 characters, exceeds a predicted probability of 50% support for all subjects. At 500 characters, the predicted likelihood of support exceeds about 80% for all subjects. In addition, in Panels A and B, the differing slopes by party suggest the amount of writing is particularly informative for positive feelings among outpartisans (e.g., a Democrat who writes a lot about liking Trump). In Figure 1, this result can be understood as consistent with pathways *iii* and *v*, that is the relevant text prompts generate *more* writing in a way that is predictive of self-reported vote choice.

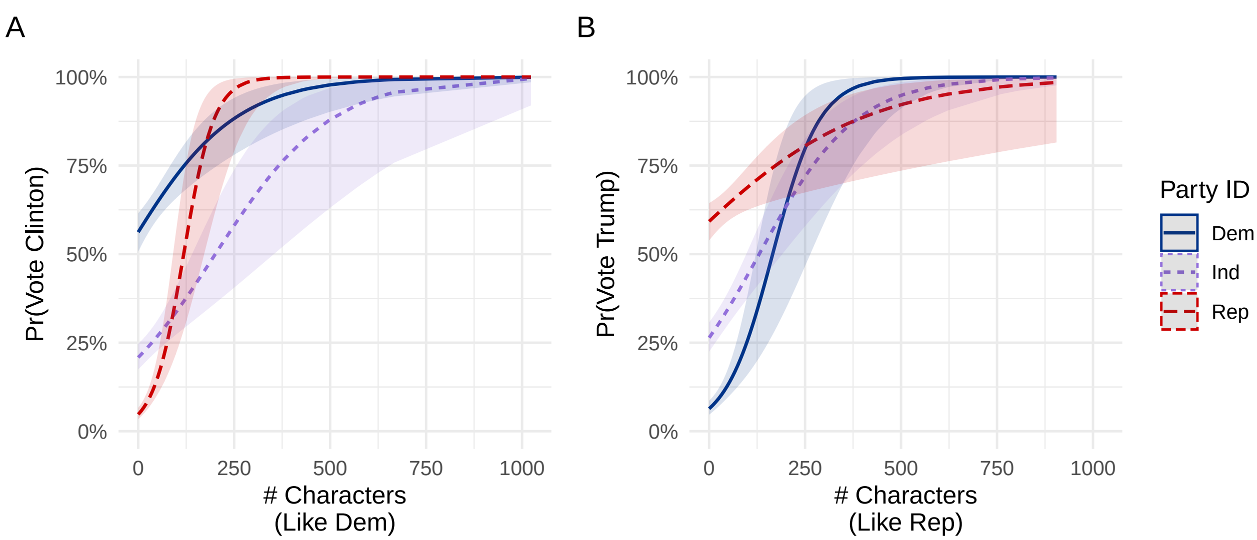


Figure 2: Marginal effects of number of characters on probability of supporting nominee, by party identification. Logistic regression models control for education, age, female, race, income, and political attention along with scales for racial resentment, hostile sexism and authoritarianism (see Table A.1 in Appendix).

Text Nonresponse as Partisanship: Evidence from ANES

Can writing nothing still convey meaning? When seen as static text with missing data, perhaps not. However, when writing is viewed as an activity, such as "ghosting" in a text exchange, the meaning may differ. When asked what they liked and disliked about each party's nominee, between 32% and 56% of subjects wrote zero characters, depending on the question. Though nonresponse is often considered uninformative, in the context of a cognitively and affectively demanding writing task, nonresponse might convey how challenging it is for subjects to write positively about the outparty nominee and/or negatively about the copartisan nominee. For example, only 21% of Democrats fail to write anything when asked what would make them vote for Clinton but 86% of Democrats write nothing when asked what would make them vote for Trump. Republicans show similar patterns (see Table A.2 in Appendix). Put colloquially, we might call this an, "If you can't say anything nice..." effect. In psychological terms, for many, it might more accurately be understood as an "emotional-overload" or emotion regulation effect (Gross 2015).

To test for informative nonresponse, I run models with the same controls as above but with the amount of nonresponse to two concordant partisan-affect questions as the key predictor. Figure 3 shows the marginal effect of nonresponse on likelihood of self-reporting support for the two major party candidates, by party identification. Figure 3 shows that with nothing more than simple count of nonresponse, it is possible to effectively predict how likely a subject is to support a particular candidate. In Figure 3 Panel A, we see that zero nonresponse for a Democrat is associated with an almost 70% likelihood of self-reporting support for Hillary Clinton. Conversely, in Panel A two nonresponses from a Democrat are associated with a less than 30% chance of self-reporting support for Clinton. In Panel B we see almost the exact same pattern for a Republican subject's predicted support Trump. Returning to the diagram in Figure 1, this result is consistent with pathways *iv* and *vi* in which more nonresponse is predictive of negative or more at least ambiguous feelings towards a nominee.

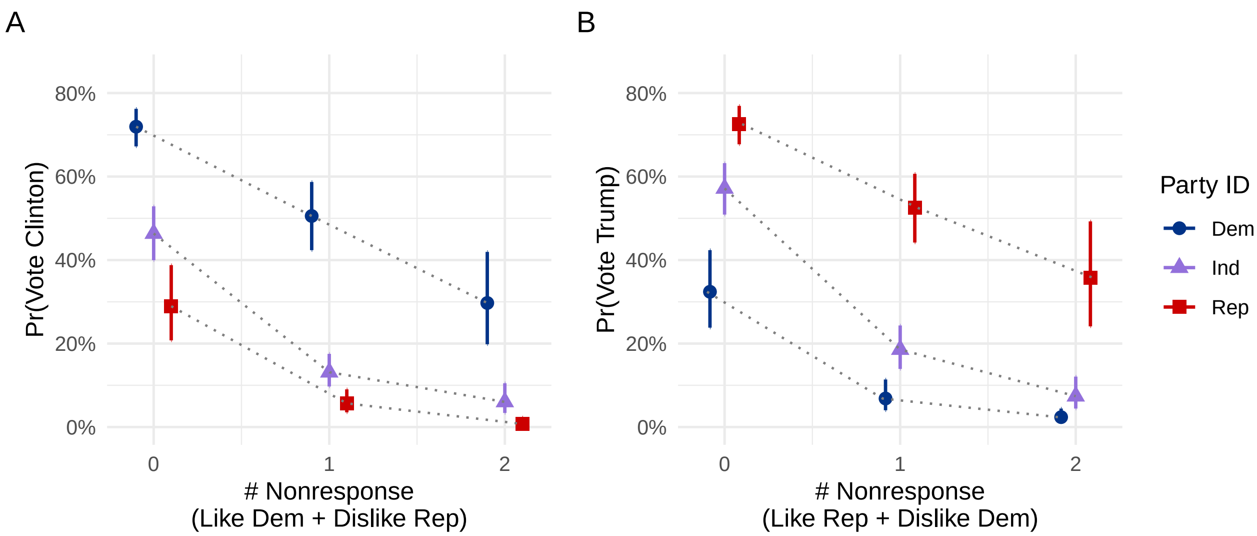


Figure 3: Marginal effects of nonresponse on probability of self-reporting support for nominee. Logistic regression models use interactions with party identification and same controls as in Figure 2 (see Table A.4 in Appendix).

Combined number of characters and nonresponse on vote choice

The results presented in Figure 3 suggest that counting nonresponse as zero characters—only slightly different from one character—significantly discounts the amount of information provided by nonresponse. Further, the results in Figure 3 point to some of the potential benefits of pooling multiple related questions. Building on these insights, the analysis in Figure 4 combines nonresponse and the number of characters from all four candidate-affect questions into a single scale I call the Partisanship Writing Scale.

To create the scale, I first subtract the Democratic-coded pair of nonresponse values from the Republican-coded pair (Equation 1). As nonresponse is reverse coded (i.e., high nonresponse suggests negative affect), this creates a scale in which negative values suggest a Democratic lean and positive values a Republican lean. I then subtract the total number of characters written in the Republican-coded pair of questions from the Democratic-coded pair (Equation 2). To put both measures on a common scale, I divide each respective scale by the maximum respective absolute value (i.e., 2 for nonresponses and about 2,000 for number of characters). The Partisanship Writing Scale is the sum of those two fractions (Equation 3).

$$\begin{aligned}
\text{Nonresponse Scale} &= \Big(non(\text{Like Dem}) + non(\text{Dislike Rep})\Big) - \Big(non(\text{Like Dem}) + non(\text{Dislike Rep})\Big) && (1)\\
\text{\# Characters Scale} &= \Big(nchar(\text{Like Rep}) + nchar(\text{Dislike Dem})\Big) \\
&\quad - \Big(nchar(\text{Like Dem}) + nchar(\text{Dislike Rep})\Big) && (2)\\
\text{Partisanship Writing Scale} &= \frac{\text{Nonresponse Scale}}{max(|\text{Nonresponse Scale}|)} + \frac{\text{\# Characters Scale}}{max(|\text{\# Characters Scale}|)} && (3)
\end{aligned}$$

Figure 4 shows that the combined Partisanship Writing Scale improves on the prior two approaches and provides comparatively precise predictions about self-reported vote choice, even after controlling for many relevant demographic characteristics. In both Panels A and B, moving from the most Democratic-leaning end of the scale to the most Republican-leaning is associated with a nearly 100 percentage point shift in support from one nominee to the other. Using a likelihood ratio test, I compare a reduced model with controls to a full model that adds the Partisanship Writing Scale. Results indicate the full model provides a significant improvement in explanatory power ($p < 0.001$, see Tables A.5 and A.6 in Appendix). Moving from the reduced to the full models, pseudo R^2 measures increase from approximately 14% to 24% in predicting support for Clinton and about 20% to 32% for Trump (see Tables A.9 and A.10 in Appendix).

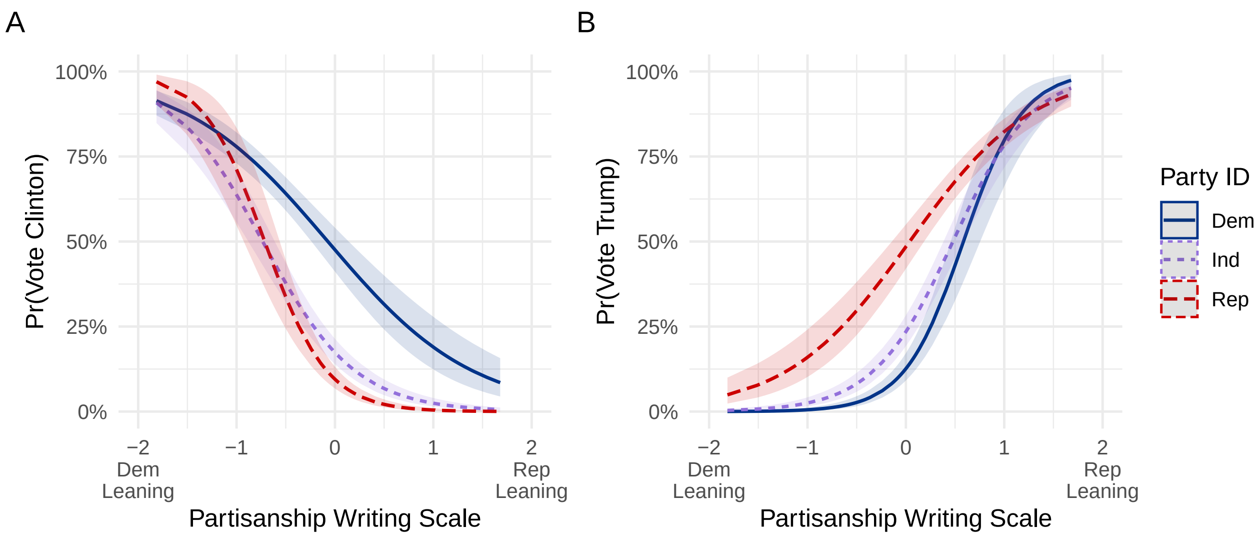


Figure 4: Marginal effects of Writing Partisanship Scale that combines number of characters and nonresponse on predicted probability of self-reporting support for nominee. Logistic regression models use interaction with party identification and same controls as in Figure 2 (see Table A.11 in Appendix).

Study 1b: Text as Validated Turnout

Can text metadata predict real-world outcomes of interest? In Study 1b, I test whether writing tasks can also be used to predict behavior, specifically turning out to vote. Common predictors of turnout such as income, education or paying attention to politics are only rough proxies for the likelihood to vote. Many college graduates, for example, spend far more time watching sports or other entertainment programming than news. A writing task, by contrast, could potentially offer more granular data on individual-level political engagement.

The primary analysis draws on an additional set of questions from the 2016 ANES in which subjects were asked four times to write about the “most important problems in America.” As the “most important problem” questions (henceforth, MIP) do not have an explicit partisan valence, I use a slightly different scale (Equation 4). First, I calculate the total number of characters written across all four questions. Second, to upweight nonresponse, I draw on Study 1a in which writing more appears to signal positive intensity, while writing nothing seems to signal a form of negative intensity. Consequently, rather than treating nonresponse as zero characters (i.e., nearly the same as one character), I count nonresponse as -25 characters. I opt for -25 as a reasonable negative approximation of the mean number of characters across the four questions, which is 31 (results are robust to other specifications). To begin the scale at zero, I also add 100 to the combined sum of upweighted nonresponse and total number of characters (e.g., if someone writes nothing four times, the MIP Writing Scale value would be 100 + (-25 x 4 nonresponses) + (0 characters) = 0).

Figure 5 shows the predicted probability of validated turnout against the number of characters with nonresponse upweighted after controlling for a standard battery of demographic measures.[1] The results presented in Figure 5 show that, after controlling for other variables, a subject who writes nothing has about a 70% predicted probability of turning out to vote. A subject who writes about 1,000 characters on the MIP Writing Scale has a greater than 80% probability of turning out, and, at 2,000 characters a greater than 90% likelihood of turning out. A likelihood ratio test comparing a reduced model with only controls to a full model that adds the MIP Writing Scale indicates the full model significantly improves the explanatory power of the model (p < 0.001, see Tables A.13 in Appendix). The full model increases pseudo R^2 estimates over the reduced model by approximately 4% (see Table A.15 in Appendix).

$$\text{Most Important Problem Writing Scale} \quad = \quad 100 + \sum_{n=1}^{4} \Big(-25 \times non(\text{MIP}_i) \Big) + \Big(nchar(\text{MIP}_i) \Big) \tag{4}$$

As a secondary analysis, I return to the candidate-affect questions about presidential nominees to test whether possible complicated feelings about a candidate, as measured in nonresponse, predict lower turnout. Figure 6 shows a noteworthy heterogeneous relationship between nonresponse, party identification and turnout. In Figure 6 Panels A and B, increased nonresponse from a copartisan is associated with a steep decrease in the predicted probability of turning out to vote. For example, in Panel A, Democrats who answer both questions (i.e., zero nonresponse) are predicted to turn out at a rate of about 73%, whereas those who fail to write anything for either question have a predicted turnout rate of about 55%. Further, in Panels A and B, more nonresponse among Independents is also associated with a significant—though less steep—decrease in the likelihood of turning out. For outpartisans, in contrast, the results in Panels A and B suggest increased nonresponse is associated with a slight *increase* in the likelihood of turning out (though, in Panel B, not a statistically significant increase). In short, among copartisans, nonresponse to the concordant candidate-affect questions appears to capture something like ambivalence that results in lower turnout. For outpartisans,

[1]For ease of visualization, Figure 5 does not show an interaction with party identification as the changes in slope by party were not statistically significant (for plot with interactions, see Figure A.2 in Appendix).

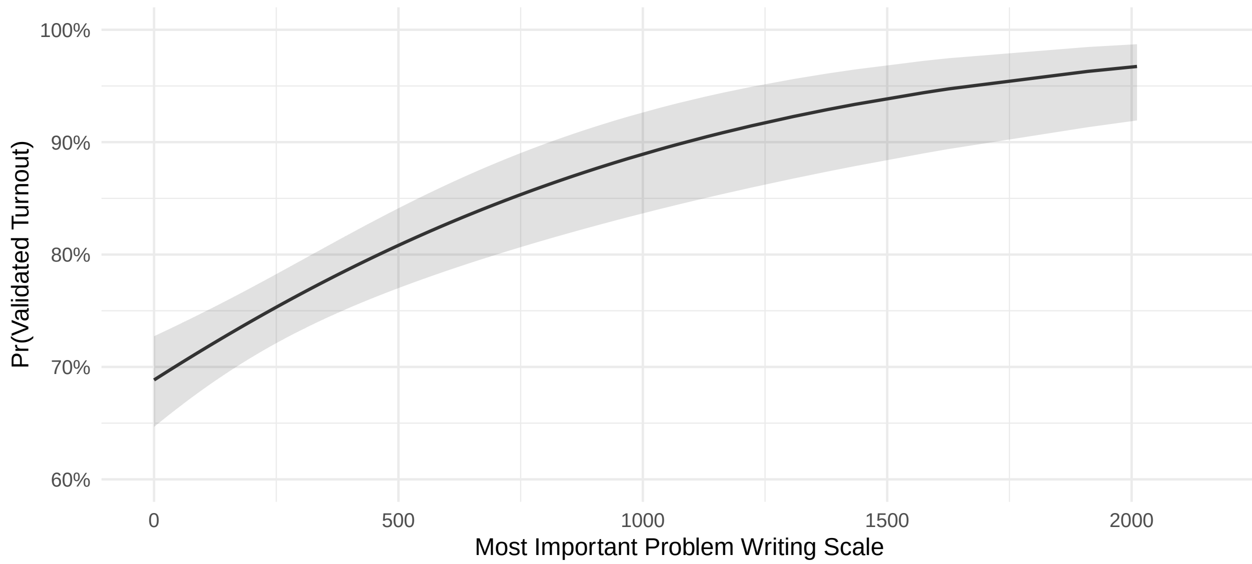


Figure 5: Marginal effects of MIP Writing Scale that combines total number of characters and weighted nonresponse on validated turnout. Logistic regression model controls for female, education, age, race, party identification, income, and political interest (see Table A.12 in Appendix).

increased nonresponse appears to reflect something like greater antipathy to the outparty nominee and/or heightened enthusiasm for the copartisan nominee that is associated with stable or modestly increased turnout.

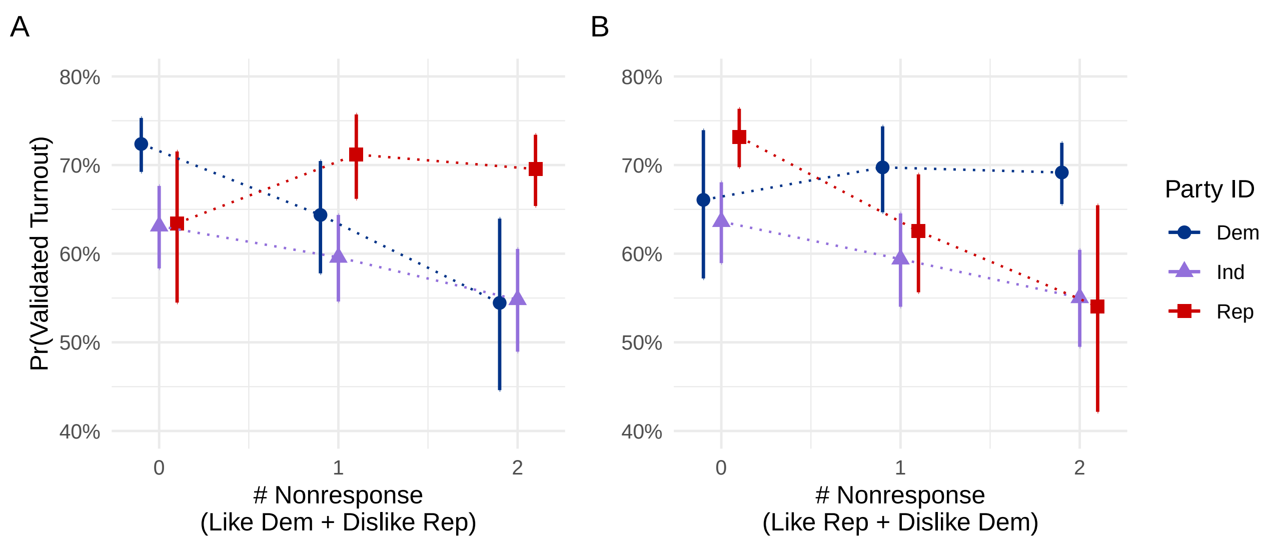


Figure 6: Marginal effects of nonresponse on validated turnout. Logistic regression models use interaction with party identification and same controls as in Figure 5 (see Table A.17 in Appendix).

Study 2a: Text as Instrument Validation

Do subjects who score lower on measures of racial resentment and hostile sexism actually behave in ways consistent with those results? As bias is often viewed negatively, a strategic subject might not respond honestly.

In Study 2a, I treat metadata about writing as an *outcome* and assess whether open-ended text responses can serve as a useful behavioral measure to validate self-reported attitudes and beliefs amid concerns about issues like social desirability bias.

The 2016 election was polarized along lines of race and gender (Schaffner, MacWilliams, and Nteta 2016; Valentino, Wayne, and Oceno 2018). Though a challenging trend for American politics, this division offers a useful way to test whether measures of racial resentment and hostile sexism correlate with observed partisan divides. Figure 7 Panels A and B present the results of two logistic regression models in which a survey instrument for racial resentment is used to predict the probability of any nonresponse to the two congruent candidate-affect questions. In Panel A we see that as racial resentment increases, the likelihood of any nonresponse increases when subjects are asked what they like about Clinton and dislike about Trump. Further, as racial resentment increases, subjects who identify as either Independents or Republicans are much more likely to exhibit both higher levels of nonresponse and significantly increased rates of nonresponse as compared with Democrats.

In contrast, in Panel B as racial resentment increases, the likelihood of any nonresponse decreases substantially when subjects are asked what would make them vote for Trump or against Clinton. Again, the levels and slopes vary significantly by party identification. The rate of change in the probability of any nonresponse is significantly lower for both Independents and Republicans compared to Democrats. In short, racial resentment is highly predictive of any nonresponse, and the levels and signs of the slopes vary consistently with other scholarship on racial attitudes and partisanship. Also, the flattest slopes in each plot are for the respective copartisan subjects and the steepest slopes are for the respective outpartisan subjects, consistent with evidence that racial attitudes help to explain some swing voting behavior (Schaffner, MacWilliams, and Nteta 2016; Valentino, Wayne, and Oceno 2018). Substantively similar results were found for any nonresponse versus hostile sexism (see Figure A.5 in Appendix).

While these results do not specifically validate that the racial resentment construct is capturing particular forms of bias, they are consistent with prior evidence that these measures were strongly correlated with partisan divides by attitudes on sex and race in 2016. Further, given that subjects were unlikely to view nonresponse as potentially revealing nonnormative attitudes, these results offer a useful behavioral measure to validate that the racial resentment and hostile sexism measures generally reflect sincere attitudes despite possible social desirability bias.

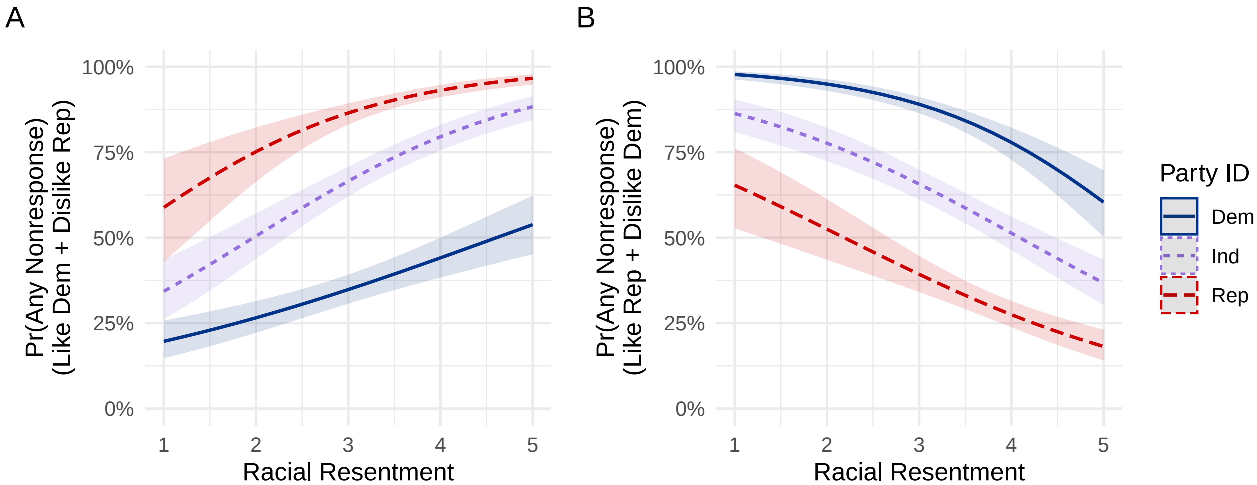


Figure 7: Marginal effects of racial resentment as a predictor of any nonresponse to congruent candidate-affect prompt pairs. Logistic regression models use interaction with party identification and same controls as in Figure 2 (see Table A.23 in Appendix).

Study 2b: Text as Multilingual Instrument Validation

Study 2a shows that attitudinal survey measures can be validated using text as behavior in a monolingual context. Metadata like number of characters and nonresponse plausibly reflect underlying features of human affect and cognition that operate independently of any specific language. To test whether measures like the number of characters work as a behavioral outcome in multilingual contexts, I run a similar set of tests to Study 2a but, in this case, drawing on the multilingual Afrobarometer (2016).

For Study 2b, I test whether subjects who self-report being more enthusiastic about democracy are sincere or simply offering what they perceive to be the socially acceptable response. The Afrobarometer asks subjects a democracy battery made up of eight questions about whether liberal democracy is important. The survey also asked subjects three times, “What, if anything, does ‘democracy’ mean to you?”[2] Importantly, all interviews were conducted face-to-face in native languages and, when appropriate, translated into English, French, or Portuguese. So, in addition to measuring whether a simple feature of text like the number of characters works across languages, these Afrobarometer data also provide evidence as to whether transcribed and, in many cases, translated text can work as a behavioral measure.

Figure 8 shows whether responses to the battery of democracy questions are predictive of the pooled

[2]Note, interviewers were instructed to, "Read the question in the language of the interview, but always state the word ‘democracy’ in English. Only translate ‘democracy’ into local language if respondent does not understand the term in the official national language. Record whether respondent understood word in English or required a local language translation. Be sure to ask ALL questions of ALL respondents, even if they have difficulty understanding the term ‘democracy’" (Afrobarometer 2014).

number of characters when asked "what democracy means to them?" (controlling for gender, education, age, language, and income-proxy). While intercepts differ across the three languages, the slopes are all positive, statistically significant and substantively similar. In short, the overall result suggests those who self-report valuing democracy in a survey are not simply giving a socially desirable answer and are significantly more likely to say more about what they value about democracy in response to open-ended prompts. More generally, the results in Figure 8 suggest that the number of characters can work well as a behavioral outcome measure across languages, modes, transcription and translation.

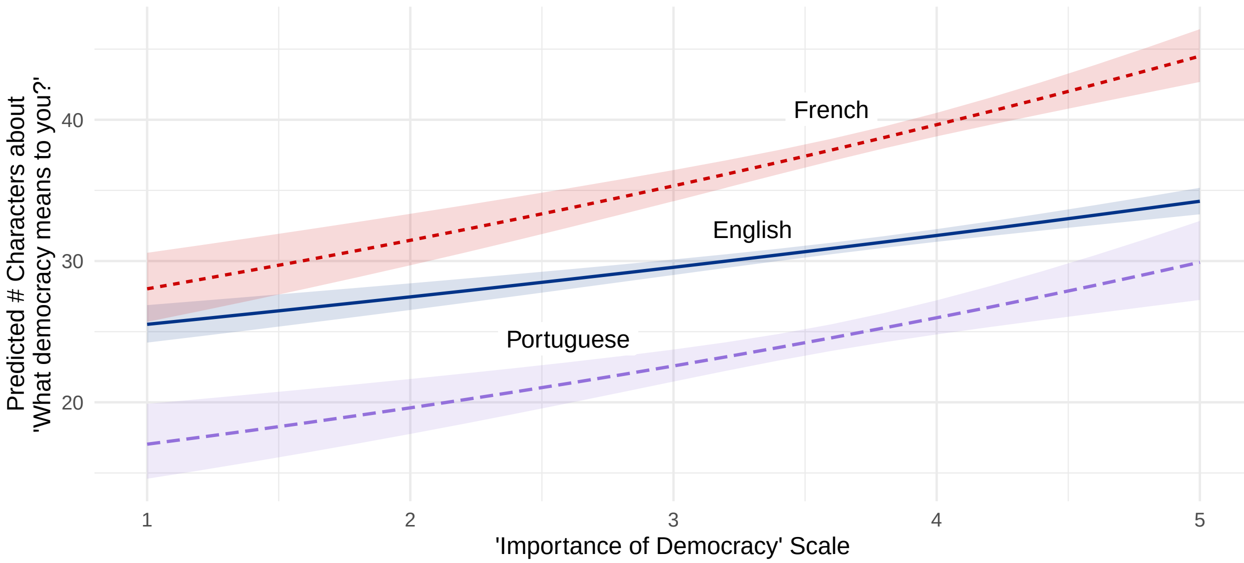


Figure 8: Marginal effect of battery of questions about importance of democracy on number of characters written in response to open-ended questions asking 'What democracy means to you?', by language. Negative binomial model includes controls for gender, education, age and income-proxy (see Table A.25 in Appendix).

Study 3a: Text as Manipulation Check of Social Exclusion

Can open-ended text prompts help reveal thoughts or feelings that are ambiguous or complex? Much of what people think or feel in a given moment is often opaque, even to themselves (Wilson 2004). Two types of hard-to-interpret experiences for any individual are (1) subtle forms of social exclusion and, (2) feelings of cross-pressure in which competing values or commitments clash. In this section, I look at two different studies that induce complex, perhaps even unconscious feelings in subjects and evaluate whether using metadata about writing as a measure of behavior can help further illuminate whether the experimental manipulation was successful.

Kuo, Malhotra, and Mo (2017), henceforth KMM, study how feelings of social exclusion might influence political attitudes, particularly among Asian Americans. As noted previously, KMM randomly induce feelings of social exclusion in white and Asian American subjects. Subjects then completed a survey and, across four different questions, wrote lists of things they liked and disliked about both the Democratic and Republican Parties. Results suggested that the joint effect of being Asian American and receiving a microaggression treatment, versus being white in the control condition, caused a negative shift in attitudes toward the Republican Party and more positive affect toward the Democratic Party.

As a form of manipulation check, KMM further asked subjects to “list as many US politicians [as] they could think of on the spot” (27). They note, “if the racial microaggression offends Asians, they may desire to compensate by showing how much they know about American politics in an attempt to feel less excluded and prove themselves as more ‘American’ ” (27). As evidence that the manipulation worked, they report that the joint effect of being Asian American and treated, relative to being white and in the control group, was that subjects listed names of an estimated 5.78 more US politicians and took approximately 88 more seconds to complete the survey (both results, $p < 0.05$).

Building on KMM’s theory and design, I run an additional manipulation check with the following assumptions: (1) on average for Asian Americans—but not whites—being asked to list things one likes and dislikes about Democrats and Republicans will have a similar ‘prove you’re an American’ effect to the US politician list test described above; (2) The additional affective and cognitive load of these tasks for some Asian Americans will result in a kind of ‘choking’ or ‘writer’s block’ effect in which more time on task does not result in more writing; (3) The absence of a ‘prove you’re an American’ cognitive load for white subjects will result in a relatively linear relationship between time on task and writing output, regardless of treatment condition.

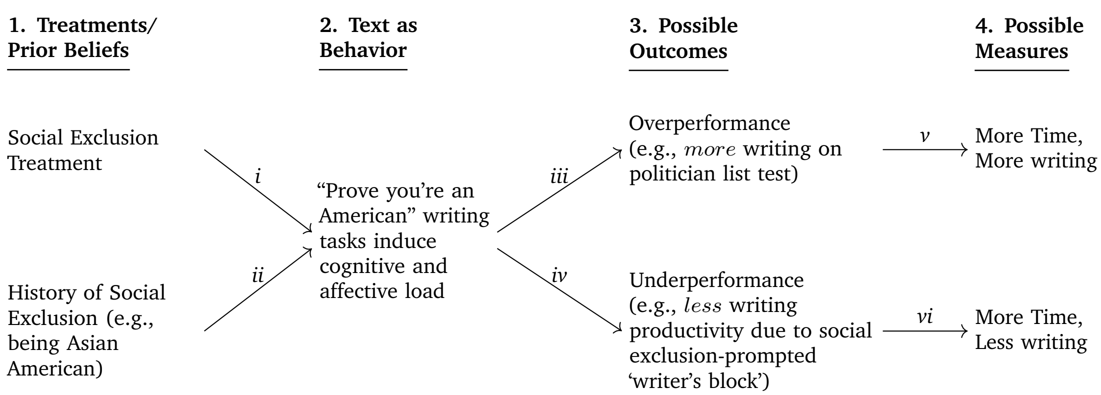


Figure 9: Hypothesized model of how (1) the joint effect of a social exclusion treatment and a history of exclusion, such as being Asian American, might combine to (2) induce complicated cognitive and affective reactions that (3) influence writing behavior to produce two potentially distinct outcomes (i.e., writing more or less) that can be detected by measuring (4) time to complete the survey and total amount of writing. Diagram does not include paths for control conditions in which subjects do not receive social exclusion treatment or do not have history of social exclusion.

Figure 9 presents a simple diagram of the process outlined by KMM along paths *iii* and *v*. In addition, the hypothesized ‘writer’s block’ process is outlined along paths *iv* and *vi*. The key insight suggested by

Figure 9 is there may be heterogeneous treatment effects that potentially induce related but distinct outcomes. Specifically, a desire to overperform may lead to increased time to complete the survey, in part as a function of writing more. The experience of underperforming, in contrast, may lead to increased time as a function of something like a 'choking' effect that leads to both taking more time and writing *less*.

To test the social exclusion 'writer's block' hypothesis, I pool the total number of characters written across all text responses and compare this measure to the total time to complete the survey. I do this under the assumption that all the various open-ended list prompts capture some component of the 'prove you're an American' effect in the microaggression treatment. Pooling all open-ended text responses has two additional benefits. First, the list response design invites subjects to provide very short replies (e.g., "Pelosi") so the number of characters in each list offers limited information about a possible affective or cognitive load compared to the pooled number of characters. Second, the timing data are only for the start and finish of the whole survey, not the time for each question. Consequently, the total amount of writing is more clearly related to the the total time spent on the survey.

Figure 10 plots the relationship between the total number of characters written across all writing prompts and time spent on the survey by race and condition. Consistent with the assumptions about a social exclusion affective and cognitive load, Figure 10 shows that while only two Asian American subjects in the control condition take 20 minutes or more to complete the survey, among treated Asian Americans, 12 take 20 minutes or more. Similarly, we can also see across both treated and control conditions, only one white subject takes 30 minutes or more to complete the survey, while five Asian American subjects do so, four of them in the treated condition. Further, the smoothed loess lines in Figure 10 show that for Asian Americans, there does appear to be a social exclusion 'writer's block' effect that results in *less* writing as time increases past about 25 minutes, particularly for treated subjects. For white subjects, however, there is no obvious significant social exclusion treatment effect on writing, and more time spent on the survey is almost linearly associated with more characters written for both treated and control subjects.

On its own, Figure 10 does not offer any kind of statistical test to assess whether the experimental manipulation worked. Further, where prior text as behavior analyses could use metadata like nonresponse or the total number of characters as relevant measures, with this study the total amount written fails to capture the 'writer's block' effect, particularly among treated Asian Americans. One way to incorporate a statistical test for the writing given the non-linearity in the total number of characters over time, is to treat the time measure as another kind of metadata for the writing tasks. KMM do precisely this for their politician list test. For this analysis I extend that approach to all five list-writing tasks under the assumption that they are the most time-intensive aspects of the survey. Under that assumption, variation in the time subjects took to complete the whole survey should serve as a reasonable approximation of the demands of those writing

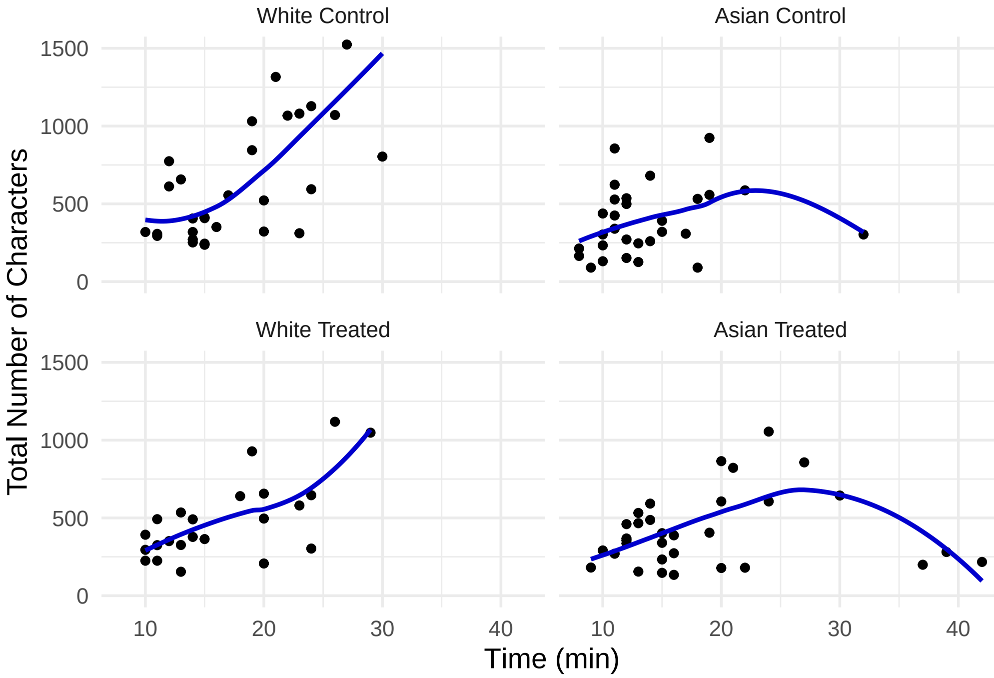


Figure 10: Scatter plots of number of characters written versus time to complete whole survey in minutes, by race and treatment condition with smoothed loess curves. Note: one white control subject who wrote more than 2000 characters is cropped for better visualization (loess curves remain unchanged).

prompts. Using time as a feature of writing has the key advantage of providing insight into the possible affective and cognitive load indpendent of the amount written.

I test for differential effects of the treatment by race on total time to complete the survey using two methods: a Wilcoxon rank-sum test and a negative binomial model that interacts treatment condition and race of subject. Table 2 presents the results of the Wilcoxon rank-sum test by race and treatment condition. The Wilcoxon test is appropriate because of the relatively small number of observations and the non-normal skew of the data. Further, as race is not randomly assigned, for this analysis I use a within-race test to estimate a causal effect of the experimental manipulation on time to complete the test. The results in Table 2 show that for white subjects, the social exclusion treatment does not induce a statistically significant difference in time to complete the survey ($p > 0.10$). In contrast, for Asian American subjects, the treatment does cause a statistically significant difference in time to completion ($p < 0.01$).

Table 2: Wilcoxon Rank Sum Test of Total Time in Study, by Race and Condition

Group	Wilcoxon W Statistic	p-value
White Treated vs White Control	416.0	0.2045
Asian-Am Treated vs Asian-Am Control	255.5	0.0026

I also estimate the joint effect of race and treatment condition on time with a negative binomial model and results are equivalent (see Table A.26 in Appendix). Though I use different statistical models than KMM, these results are substantively comparable to those published in their manipulation check section which raises another question: what value does text as behavior—in the form of both character counts and time—add to KMM's analysis? One contribution of considering two types of metadata about text is insight into two possibly distinct mechanisms by which the social exclusion treatment appears to work. KMM find that, consistent with a "prove you're an American" process, the joint effect of the social exclusion treatment and being Asian American results in listing significantly *more* politicians as compared to subjects who are white in the control condition. Looking across all writing prompts, however, provides suggestive evidence of an additional social exclusion effect in which some treated Asian American subjects write significantly *less*.

Study 3b: Text as Manipulation Check of Cross-Pressure

Another complex issue voters often confront is to be cross-pressured by different forces like ideology and identity. For example, a Catholic Democrat might favor liberal economic policies but also feel conflicted about whether to support a candidate who backs abortion rights. Detecting these kinds of cross-pressuring forces has been done widely in social science but often with survey instruments that rely on conscious thoughts in response to direct questions. Open-ended writing tasks can offer windows into a person's thought processes that may not be entirely conscious. For example, the high rates of nonresponse in Study 1a may reflect how, for many subjects, asking what they like about Hillary Clinton or Donald Trump may induce feelings of anger, disgust, sadness or even hatred as part of a complex rush of emotions.

To test for the possibility of writing as a measure of hard to articulate thoughts and feelings, I draw on data from a study by McGhee (2018) in which 1,050 Black subjects were randomly assigned to one of three experimental conditions and asked to pick between two hypothetical Black Democratic candidates, such as happens in urban primaries. In the control condition, the two candidates were similar on a list of demographic characteristics—both candidates were described as middle aged, Christian, male, straight and veterans—but one candidate supported more popular Black-oriented or liberal policies (e.g., "Strong advocate for criminal justice reform") while the second candidate backed more moderate and less conventionally Black-oriented policies (e.g., "Wants to reduce federal debt"). In the control condition, 85% of subjects opted for the more liberal candidate.

To induce a degree of cross-pressure, two treatment conditions varied one aspect of the more liberal candidate's identity while keeping the policy positions unchanged. In one treatment condition, the liberal candidate's religion was switched from Christian to Muslim, and was then preferred by 75% of subjects. In a second treatment condition, the liberal candidate's sexuality was changed from straight to gay, and was then selected by 68% of subjects (for more detail, see Figure A.7 in Appendix). In short, the treatment conditions appeared to successfully force some subjects to choose between ideological congruence (e.g., more liberal and/or Black-oriented policies) and identity congruence (e.g., shares my religion and/or not a member of a perceived out-group).

Subjects in this experiment were also asked to write a two to three sentence explanation about why they selected their chosen candidate and what they might change about the candidate. Building on the results presented in Study 1a, we might predict higher rates of nonresponse when subjects are confronted with more complex feelings such as bias towards an out-group or being cross-pressured between ideological and identity congruence. Further, we might also hypothesize that some cross-pressured subjects write *more* to explain or rationalize their choices.[3] Figure 11 presents the number of characters written by subjects under the three conditions. Looking first at nonresponse, one can see clearly that subjects in the cross-pressured or treated conditions (B and C) are entering zero characters at substantially higher rates than subjects in the non-cross-pressured control condition A. Figure 11 also shows that the longest responses are in conditions B and C, though it is not clear the right-tails of the distributions differ significantly.

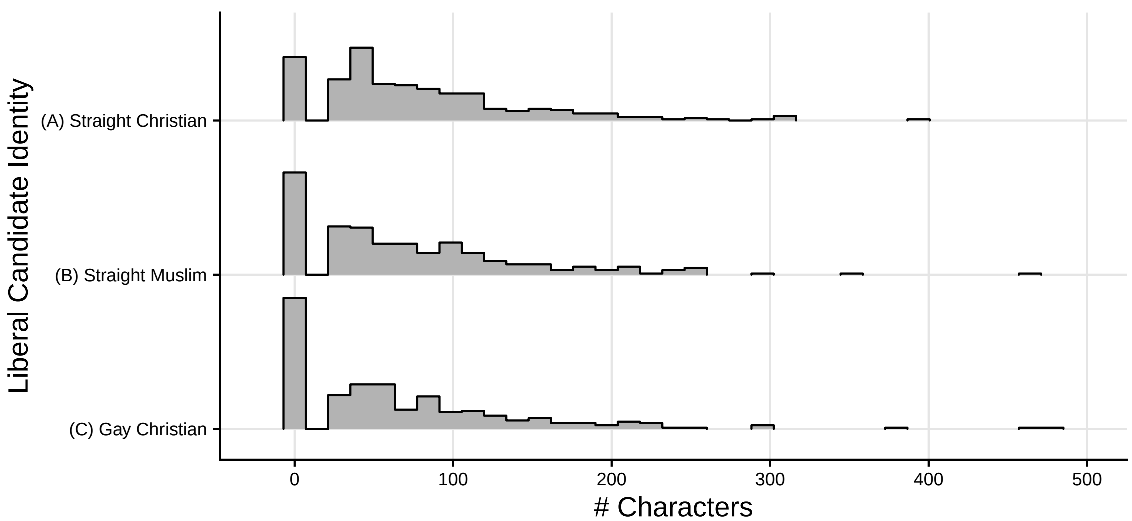


Figure 11: Ridgeline histogram plot of number of characters used in response to prompt about why subjects chose their candidate. Results are consistent with a cross-pressure effect that shows substantially higher rates of nonresponse (zero on x-axis) as subjects opt against hypothetical Muslim or gay candidates (rows B and C).

To test for these two cross-pressure effects (nonresponse and, potentially, overresponse), I run four models. The first two models use logistic regression to test for nonresponse, and find both treatment conditions are significantly predictive of nonresponse relative to the control condition. Figure 12 plots the marginal effects of these two models. Subjects in the "Straight Muslim" condition were about 81% more likely to enter zero characters when asked to explain their choice as compared to the "Straight Christian" condition. Subjects in

[3]McGhee (2018) highlights a number of longer replies in which subjects express feeling conflicted between "follow[ing] the Bible" and supporting a candidate whose "policies are better." To be clear, many subjects also appear unconflicted and explicitly state anti-Muslim or anti-gay views.

the "Gay Christian" condition were about 156% more likely to write zero characters compared to subjects assigned the control condition.

I also test for overresponse using negative binomial models estimating the total number of characters typed in the open-ended response by treatment condition. For both treatment conditions, I find a positive but not statistically significant relationship between receiving a cross-pressure treatment and the number of characters typed (see Table A.27 in Appendix).[4] In sum, the results in Figure 11 and Figure 12 suggest that under experimental conditions subjects can be cross-pressured and that those conflicted states of mind are detectable in higher rates of nonresponse but, at least in this study, not overresponse. More generally, these results demonstrate how text as behavior measures could serve as manipulation checks to show whether a treatment worked as intended, or as a behavioral measure of an otherwise difficult to reveal state of mind.

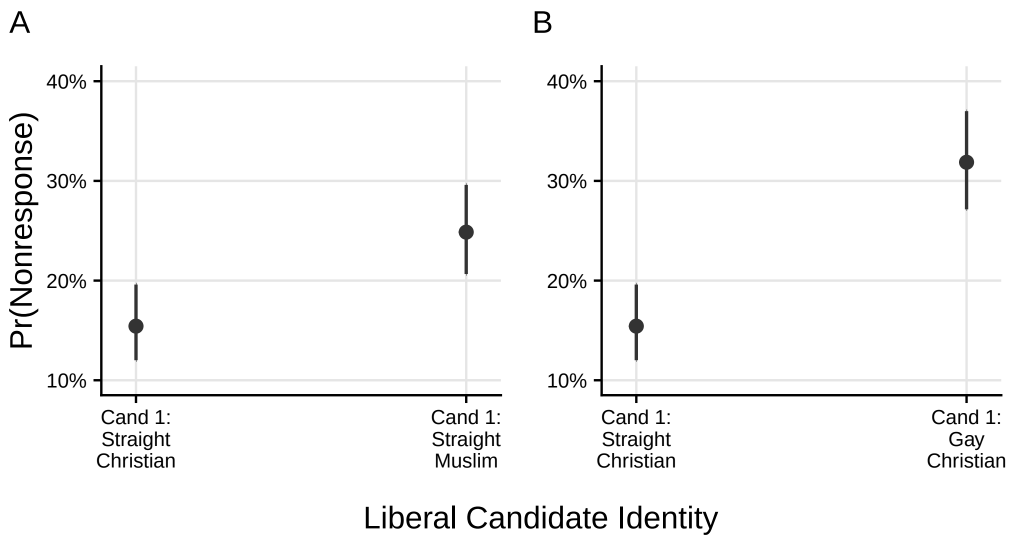


Figure 12: Marginal effects of liberal candidate identity on predicted probability of nonresponse in open-ended text when subjects asked why they supported their selected candidate, by experimental condition, using logistic regression (see Table A.27 in Appendix).

Discussion

While the six studies above show metadata from open-ended prompts can serve as useful measures of attitudes and behavior, at least five important questions remain. First, while the results above strongly suggest certain types of open-ended prompts are closely related to specific attitudes and behaviors, more research is needed to map and categorize which types of prompts are useful for measuring particular kinds of beliefs, attitudes and actions.

Second, a number of important questions remain about measurement, methods and instrument design.

[4]Zero-inflated methods that attempt model nonresponse and overresponse as two separate data generating processes produced similar results so the simpler models are presented for ease of interpretation.

While the studies above show nonresponse can be informative across a variety of research questions, I offer no strong theory about how to weight nonresponse when pooling it with other measures like the average total number of characters written across multiple related questions. While treating nonresponse as missing data or zero characters likely discounts informative responses, the solutions used here should be treated as preliminary. Approaches that model zero character responses as the result of a separate data-generating process may be more appropriate (e.g., zero-inflated models). Further, other forms of missing data such as attrition or saying "Don't know" may also potentially be types of informative nonresponse but are beyond the scope of this paper. Data visualizations, such as scatter plots with loess lines or ridgeline histogram plots (Wilke 2022), were shown to be effective for identifying important, if subtle, patterns in the textual data but I offer no clear framework for how to apply them when evaluating text as behavior. Technical and measurement questions, such as how best to write open-ended prompts or whether to include character limits remain unresolved. Additional types of tests, such as using text metadata to measure the salience of an issue over time with panel data (see example in Section A.8.1 in Appendix), also merit further development.

Third, none of the analyses in this paper were pre-registered and should all be treated as exploratory. Two recent, independent studies, however, have applied the text as behavior approach successfully. First, Cavaillé, Chen, and Straeten (n.d.) show that character counts in letter writing tasks on topics like the minimum wage and abortion can serve as meaningful behavioral outcomes for measures of preference intensity. Second, Mikkelborg (2023) ran a pre-registered study and found the number of characters from a candidate-affect writing prompt about Trump effectively predicted voting preferences for hypothetical candidates in a conjoint experiment.

Fourth, this analysis offers no particular insight or methods on how to address fraud detection with written responses. Some techniques, like scoring submissions for the number of words that appear in a dictionary, can help weed out junk responses. As generative artificial intelligence (AI) and large language models (LLM) such as ChatGPT becomes widespread, however, more sophisticated forms of fraud may emerge and require other forms of detection (Veselovsky, Ribeiro, and West 2023). To combat fraud, text as behavior methods may require subjects to be pre-screened or validated in some way as is done by many survey research firms. Alternatively, it may be necessary to try and detect fraud at the level of the subject via tools like attention checks or keystroke tracking rather than in the content of the text response. Veselovsky, Ribeiro, and West (2023), for example, use "Javascript to extract all keystrokes made by workers while performing the [writing] task, including copy and paste actions." Providing incentives can also improve the quality of open-ended survey responses (Li 2023). In addition, some measures like nonresponse are likely to be more robust to widespread adoption of generative AI, and speak to the value using instruments in which subjects have little insight into what is actually being measured.

Finally, while this analysis shows that text metadata can work across cultural contexts and modes, without more research we should remain cautious about the generalizability of these methods. There may be languages or cultures for which a simple measure like number of characters is a poor gauge. For example, if a culture values quiet reflection more than talkativeness, summing the number of characters in a response may be uninformative. In addition, the cognitive and affective dynamics of speaking are substantively different than that of writing and metadata measures such as character counts, time or nonresponse may not operate in equivalent ways (Benoit, Munger, and Spirling 2019). Further consideration is needed as to how culture, mode, and type of text source might influence metadata measures.

Conclusion

How to quantify human attitudes and behaviors is a fundamental question in social science. This study proposes an extension of "text as data" methods to include "text as behavior." Specifically, writing is sufficiently cognitively and affectively demanding that it should often be understood as both a means of communicating things like sentiment and, also, as a form of action. As shown in prior work and throughout this analysis, metadata about writing tasks—like the number of characters in a submission or the time to completion—can provide valuable insights into states of mind, even when those states may not be fully apparent to the subjects themselves. Measuring human behavior will always be a challenge but the findings in this paper suggest that treating writing tasks as effortful offers social scientists an additional useful method to reveal genuine preferences and behaviors. As Gloria Steinem once said, "I don't like writing. I like having written" (1976).

References

Afrobarometer. 2014. “Questionairre for Afrobarometer Round 6: The Quality of Democracy and Governance in Ghana.” Online at https://www.afrobarometer.org/wp-content/uploads/2022/02/gha_r6_questionnaire.pdf.

———. 2016. Merged Round 6 data (36 countries). Available at http://www.afrobarometer.org.

Alsaedi, Nasser, Pete Burnap, and Omer Rana. 2017. “Can We Predict a Riot? Disruptive Event Detection Using Twitter.” *ACM Trans. Internet Technol* 17 (2). https://doi.org/10.1145/2996183.

Ansolabehere, Stephen, and Brian Schaffner. 2017. “CCES Common Content, 2016.” Harvard Dataverse.

Aronson, Elliot, and J Merrill Carlsmith. 1968. “Experimentation in Social Psychology.” *The Handbook of Social Psychology* 2 (2): 1–79.

Benoit, Kenneth, Kevin Munger, and Arthur Spirling. 2019. “Measuring and Explaining Political Sophistication Through Textual Complexity.” *American Journal of Political Science* 63 (2): 491–508.

Bisgaard, Martin. 2019. “How Getting the Facts Right Can Fuel Partisan-Motivated Reasoning.” *American Journal of Political Science* 63 (4): 824–39.

Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. “Twitter Mood Predicts the Stock Market.” *Journal of Computational Science* 2 (1): 1–8. https://www.sciencedirect.com/science/article/pii/S187775031100007X.

Butler, Daniel M, and David E Broockman. 2011. “Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators.” *American Journal of Political Science* 55 (3): 463–77.

Campbell, Donald T, and Julian C Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Houghton Mifflin Company.

Cavaillé, Charlotte, Daniel L. Chen, and Karine Van der Straeten. n.d. “Who Cares? Measuring Differences in Preference Intensity.” Working papaer. Online at https://charlottecavaille.wordpress.com/working-papers/.

Collins, Jonathan E. 2021. “Does the Meeting Style Matter? The Effects of Exposure to Participatory and Deliberative School Board Meetings.” *American Political Science Review* 115 (3): 790–804.

Danescu-Niculescu-Mizil, Cristian, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. “A Computational Approach to Politeness with Application to Social Factors.” https://arxiv.org/abs/1306.6078.

Day, Martin V, Susan T Fiske, Emily L Downing, and Thomas E Trail. 2014. “Shifting Liberal and Conservative Attitudes Using Moral Foundations Theory.” *Personality and Social Psychology Bulletin* 40 (12): 1559–73.

Denscombe, Martyn. 2008. “The Length of Responses to Open-Ended Questions: A Comparison of Online and Paper Questionnaires in Terms of a Mode Effect." *Social Science Computer Review* 26 (3): 359–68. https://doi.org/10.1177/0894439307309671.

Eady, Gregory, Frederik Hjorth, and Peter Thisted Dinesen. 2022. "Do Violent Protests Affect Expressions of Party Identity? Evidence from the Capitol Insurrection." *American Political Science Review*, 1–7.

Enamorado, Benjamin Fifield, Ted, and Kosuke Imai. 2017. ""User's Guide and Codebook for the ANES 2016 Time Series Voter Validation Supplemental Data."" Technical report, Princeton University. Online at https://electionstudies.org/wp-content/uploads/2018/03/anes_timeseries_2016voteval_userguidecodebook.pdf.

Findley, Michael G., Kyosuke Kikuta, and Michael Denly. 2021. "External Validity." *Annual Review of Political Science* 24 (1): 365–93. https://doi.org/10.1146/annurev-polisci-041719-102556.

Gillion, Daniel Q. 2016. *Governing with Words: The Political Dialogue on Race, Public Policy, and Inequality in America*. Cambridge University Press.

Graham, Steve. 2018. "A Revised Writer(s)-Within-Community Model of Writing." *Educational Psychologist* 53 (4): 258–79.

Grimmer, Justin, Margaret E Roberts, and Brandon M Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.

Gross, James J. 2015. "Emotion Regulation: Current Status and Future Prospects." *Psychological Inquiry* 26 (1): 1–26. https://doi.org/10.1080/1047840X.2014.940781.

Haidt, Jonathan. 2012. *The Righteous Mind: Why Good People Are Divided by Religion and Politics*. New York: Pantheon Books.

Hayes, John R. 1996. "A New Framework for Understanding Cognition and Affect in Writing." In *The Science of Writing: Theories, Methods, Individual Differences and Applications*, edited by C. Michael Levy and Sarah Ransdell, 6–44. Lawrence Erlbaum Associates.

Jacobi, Tonja, and Dylan Schweers. 2017. "Justice, Interrupted: The Effect of Gender, Ideology, and Seniority at Supreme Court Oral Arguments." *Va. L. Review* 103: 1379.

Karpowitz, Christopher F, Tali Mendelberg, and Lee Shaker. 2012. "Gender Inequality in Deliberative Participation." *American Political Science Review* 106 (3): 533–47.

Kellogg, Ronald T. 1999. *The Psychology of Writing*. Oxford University Press.

Kramer, Adam DI, Jamie E Guillory, and Jeffrey T Hancock. 2014. "Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks." *Proceedings of the National Academy of Sciences* 111 (24): 8788–90.

Kuo, Alexander, Neil Malhotra, and Cecilia Hyunjung Mo. 2017. "Social Exclusion and Political Identity: The Case of Asian American Partisanship." *The Journal of Politics* 79 (1): 17–32.

Li, Yimeng. 2023. "Incentivizing Open-Ended Survey Responses." Presented at Polmeth 2023, July 6, 2023 at Stanford Univeristy.

Lowande, Kenneth, and Andrew Proctor. 2020. "Bureaucratic Responsiveness to LGBT Americans." *American Journal of Political Science* 64 (3): 664–81.

Maloney, Elizabeth. 2021. "The Gender Wage Gap: A Product of Misogyny and Gender Norms." Online at https://www.economics.uci.edu/files/docs/2021/gradjobmarket/maloney_elizabeth.pdf.

Matthews, Peter Hugoe. 2003. *Linguistics: A Very Short Introduction*. Oxford University Press.

McDermott, Rose. 2002. "Experimental Methodology in Political Science." *Political Analysis* 10 (4): 325–42. http://www.jstor.org/stable/25791696.

McGhee, Brandon. 2018. "The Blacker the Berry: The Black Church, Linked Fate, Marginalization, and the Electability of Black Candidates." Senior Thesis, Department of Politics, Princeton University. Online at http://arks.princeton.edu/ark:/88435/dsp010k225d77p.

Mikkelborg, Anna. 2023. "Do White Democrats (Still) Prefer Coethnic Representation?" Paper presented at Midwest Political Science Association.

Mosleh, Mohsen, Cameron Martel, Dean Eckles, and David Rand. 2021. "Perverse Downstream Consequences of Debunking: Being Corrected by Another User for Posting False Political News Increases Subsequent Sharing of Low Quality, Partisan, and Toxic Content in a Twitter Field Experiment." In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, USA: Association for Computing Machinery. https://doi.org/10.1145/3411764.3445642.

Mossholder, Kevin W, Randall P Settoon, Stanley G Harris, and Achilles A Armenakis. 1995. "Measuring Emotion in Open-Ended Survey Responses: An Application of Textual Data Analysis." *Journal of Management* 21 (2): 335–55.

Mummolo, Jonathan, and Erik Peterson. 2019. "Demand Effects in Survey Experiments: An Empirical Assessment." *American Political Science Review* 113 (2): 517–29.

Munger, Kevin. 2017. "Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment." *Political Behavior* 39 (3): 629–49.

Panteli, Niki. 2002. "Richness, Power Cues and Email Text." *Information & Management* 40 (2): 75–86.

Pennebaker, James W. 1997. "Writing about Emotional Experiences as a Therapeutic Process." *Psychological Science* 8 (3): 162–66.

Pennebaker, James W, Martha E Francis, and Roger J Booth. 2001. "Linguistic Inquiry and Word Count: LIWC 2001." *Mahway: Lawrence Erlbaum Associates* 71 (2001): 2001.

Roberts, Margaret E, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G Rand. 2014. "Structural Topic Models for Open-Ended Survey

Responses." *American Journal of Political Science* 58 (4): 1064–82.

Rude, Stephanie, Eva-Maria Gortner, and James Pennebaker. 2004. "Language Use of Depressed and Depression-Vulnerable College Students." *Cognition & Emotion* 18 (8): 1121–33.

Sakaki, Takeshi, Makoto Okazaki, and Yutaka Matsuo. 2010. "Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors." In *Proceedings of the 19th International Conference on World Wide Web*, 851–60. Association for Computing Machinery. https://doi.org/10.1145/1772690.1772777.

Schaffner, Brian. 2020. *The Acceptance and Expression of Prejudice During the Trump Era*. Elements in American Politics. Cambridge University Press.

Schaffner, Brian, Matthew MacWilliams, and Tatishe Nteta. 2016. "Explaining White Polarization in the 2016 Vote for President: The Sobering Role of Racism and Sexism." In *Conference on the US Elections of*, 8–9.

Steinem, Gloria. 1976. "As Quoted in Pocatellan Captures Flavor of Stanley Basin's History by Joan LaLiberte." *Idaho State Journal*.

Steinert-Threlkeld, Zachary C. 2017. "Spontaneous Collective Action: Peripheral Mobilization During the Arab Spring." *American Political Science Review* 111 (2): 379–403.

Stephens-Davidowitz, Seth. 2014. "The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data." *Journal of Public Economics* 118: 26–40.

Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. "Lexicon-Based Methods for Sentiment Analysis." *Computational Linguistics* 37 (2): 267–307.

ten Kleij, Frederieke, and Pieter A. D Musters. 2003. "Text Analysis of Open-Ended Survey Responses: A Complementary Method to Preference Mapping." *Food Quality and Preference* 14 (1): 43–52.

Thompson, Clive. 2010. "Clive Thompson: The Phone Call Is Dead." *Wired* 18 (8).

Tumasjan, Andranik, Timm O Sprenger, Philipp G Sandner, and Isabell M Welpe. 2010. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment." In *Fourth International AAAI Conference on Weblogs and Social Media*. Citeseer.

Valentino, Nicholas A, Carly Wayne, and Marzia Oceno. 2018. "Mobilizing Sexism: The Interaction of Emotion and Gender Attitudes in the 2016 US Presidential Election." *Public Opinion Quarterly* 82 (S1): 799–821.

van Zanden, Jan Luiten, Joerg Baten, Marco Mira d'Ercole, Auke Rijpma, and Marcel P. Timmer, eds. 2014. *How Was Life?: Global Well-Being Since 1820*. Paris: OECD Publishing. https://doi.org/https://doi.org/https://doi.org/10.1787/9789264214262-en.

Veselovsky, Veniamin, Manoel Horta Ribeiro, and Robert West. 2023. "Artificial Artificial Artificial Intelligence: Crowd Workers Widely Use Large Language Models for Text Production Tasks."

Voigt, Rob, Nicholas P Camp, Vinodkumar Prabhakaran, William L Hamilton, Rebecca C Hetey, Camilla M Griffiths, David Jurgens, Dan Jurafsky, and Jennifer L Eberhardt. 2017. "Language from Police Body

Camera Footage Shows Racial Disparities in Officer Respect." *Proceedings of the National Academy of Sciences* 114 (25): 6521–26.

Wilke, Claus O. 2022. *ggridges: Ridgeline Plots in "Ggplot2"*. https://wilkelab.org/ggridges/.

Wilson, Timothy D. 2004. *Strangers to Ourselves*. Cambridge, MA: Harvard University Press.

Winking, Jeffrey, and Nicholas Mizer. 2013. "Natural-Field Dictator Game Shows No Altruistic Giving." *Evolution and Human Behavior* 34 (4): 288–93.

Wright, Robert. 2017. *Why Buddhism Is True: The Science and Philosophy of Meditation and Enlightenment*. Simon; Schuster.

Yan, Alan N., and Rachel Bernhard. 2023. "The Silenced Text: Field Experiments on Gendered Experiences of Political Participation." *American Political Science Review*.

Zhang, Justine, Jonathan P Chang, Cristian Danescu-Niculescu-Mizil, Lucas Dixon, Yiqing Hua, Nithum Thain, and Dario Taraborelli. 2018. "Conversations Gone Awry: Detecting Early Signs of Conversational Failure." *arXiv Preprint arXiv:1805.05345*.

Ziegler, Jeffrey. 2022. "A Text-as-Data Approach for Using Open-Ended Responses as Manipulation Checks." *Political Analysis* 30 (2): 289–97.

%TC:ignore

A Appendix

A.1 Study 1a

A.1.1 Study 1a: Table of Self-reported Vote Choice vs Number of Characters in Positive Partisan-Affect Prompts

Table A.1: Self-Reported Vote Choice vs Number of Characters in Positive Partisan-Affect Prompts

	Dependent variable:	
	Vote Clinton	Vote Trump
	logistic	*logistic*
	(1)	(2)
# Characters (Like Dem)	0.01^{*}	
	(0.002)	
# Characters (Like Rep)		0.02^{*}
		(0.003)
Party: Independent	-1.58^{*}	1.66^{*}
	(0.13)	(0.18)
Party: Republican	-3.25^{*}	3.06^{*}
	(0.20)	(0.19)
Racial Resentment	-0.62^{*}	0.66^{*}
	(0.06)	(0.06)
Hostile Sexism	-0.30^{*}	0.23^{*}
	(0.06)	(0.07)
Authoritarianism	-0.16^{*}	0.10^{*}
	(0.05)	(0.05)
Education	0.06^{*}	0.02
	(0.03)	(0.03)
Age (yrs)	0.08^{*}	0.10^{*}
	(0.02)	(0.02)
Female	0.02	0.07
	(0.11)	(0.11)
Race: Black	0.56	−0.07
	(0.31)	(0.46)
Race: Hispanic	0.04	−0.09
	(0.30)	(0.38)
Race: Native American	0.48	−1.20
	(0.70)	(1.16)
Race: Other	0.17	0.36
	(0.37)	(0.42)
Race: White	−0.20	1.03^{*}
	(0.27)	(0.32)
Income	0.03^{*}	0.01^{*}
	(0.01)	(0.01)
Political Attenion	0.18^{*}	0.15^{*}
	(0.05)	(0.05)
# Chars (Like Dem) x Ind	−0.0004	
	(0.002)	
# Chars (Like Dem) x Rep	0.02^{*}	
	(0.005)	
# Chars (Like Rep) x Ind		-0.01^{*}
		(0.004)
# Chars (Like Rep) x Rep		-0.01^{*}
		(0.004)
Constant	1.25^{*}	-8.60^{*}
	(0.52)	(0.61)
Observations	3,203	3,203
Log Likelihood	-1,231.15	-1,227.35
Akaike Inf. Crit.	2,500.30	2,492.70
Note:		$^{*}p < 0.05$

A.1.2 Study 1a: Plot of Self-reported Vote Choice vs Number of Characters without Interaction with Party ID

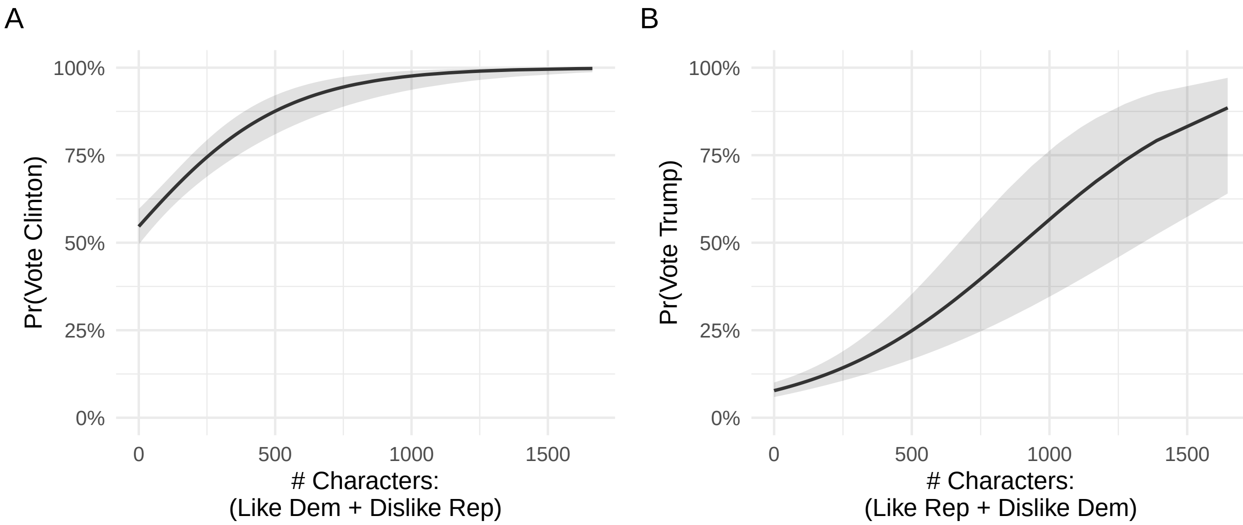


Figure A.1: Marginal effects of number of characters on probability of supporting nominee *without* plotting interaction with party identification. Models control for education, age, female, race, party ID, income, political attention, racial resentment, hostile sexism and authoritarianism.

A.1.3 Study 1a: Tables of Self-reported Vote Choice vs Nonresponse with Candidate-Affect Prompts

Table A.2: Nonresponse by Party ID and Individual Candidate-Affect Prompts

Party ID	Candidate-Affect	# Nonresponse	n	Percent
Democrat	Like Dem	0	1,139	79%
		1	311	21%
	Dislike Dem	0	473	33%
		1	977	67%
	Like Rep	0	210	14%
		1	1,240	86%
	Dislike Rep	0	1,262	87%
		1	188	13%
Independent	Like Dem	0	547	40%
		1	820	60%
	Dislike Dem	0	909	66%
		1	458	34%
	Like Rep	0	587	43%
		1	780	57%
	Dislike Rep	0	961	70%
		1	406	30%
Republican	Like Dem	0	181	15%
		1	1,050	85%
	Dislike Dem	0	1,081	88%
		1	150	12%
	Like Rep	0	970	79%
		1	261	21%
	Dislike Rep	0	550	45%
		1	681	55%
Other	Like Dem	0	71	32%
		1	151	68%
	Dislike Dem	0	136	61%
		1	86	39%
	Like Rep	0	83	37%
		1	139	63%
	Dislike Rep	0	138	62%
		1	84	38%

Table A.3: Nonresponse by Party ID and Congruent Candidate-Affect Prompt Pairs

Party ID	Candidate-Affect	# Nonresponse	n	Percent
Democrat	Like Dem + Dislike Rep	0	1,070	74%
		1	261	18%
		2	119	8%
	Like Rep + Dislike Dem	0	138	10%
		1	407	28%
		2	905	62%
Independent	Like Dem + Dislike Rep	0	515	38%
		1	478	35%
		2	374	27%
	Like Rep + Dislike Dem	0	539	39%
		1	418	31%
		2	410	30%
Republican	Like Dem + Dislike Rep	0	151	12%
		1	429	35%
		2	651	53%
	Like Rep + Dislike Dem	0	903	73%
		1	245	20%
		2	83	7%
Other	Like Dem + Dislike Rep	0	68	31%
		1	73	33%
		2	81	36%
	Like Rep + Dislike Dem	0	74	33%
		1	71	32%
		2	77	35%

A.1.4 Study 1a: Table of Self-Reported Vote Choice vs Nonresponse

Table A.4: Self-reported vote choice vs nonresponse

	Dependent variable:	
	Vote Clinton	Vote Trump
	(1)	(2)
# Nonresponse 1 (Like Dem + Dislike Rep)	-0.92^{*}	
	(0.18)	
# Nonresponse 2 (Like Dem + Dislike Rep)	-1.80^{*}	
	(0.28)	
# Nonresponse 1 (Like Rep + Dislike Dem)		-1.88^{*}
		(0.35)
# Nonresponse 2 (Like Rep + Dislike Dem)		-2.98^{*}
		(0.39)
Party: Independent	-1.09^{*}	1.02^{*}
	(0.14)	(0.24)
Party: Republican	-1.84^{*}	1.71^{*}
	(0.23)	(0.23)
Racial Resentment	-0.50^{*}	0.50^{*}
	(0.06)	(0.06)
Hostile Sexism	-0.22^{*}	0.19^{*}
	(0.07)	(0.07)
Authoritarianism	-0.13^{*}	0.09
	(0.05)	(0.05)
Education	0.02	-0.01
	(0.03)	(0.03)
Age (yrs)	0.09^{*}	0.11^{*}
	(0.02)	(0.02)
Female	0.002	0.17
	(0.11)	(0.11)
Race: Black	0.52	0.37
	(0.32)	(0.48)
Race: Hispanic	-0.08	0.10
	(0.32)	(0.39)
Race: Native American	0.63	-0.79
	(0.73)	(1.18)
Race: Other	0.18	0.36
	(0.39)	(0.43)
Race: White	-0.17	0.92^{*}
	(0.29)	(0.33)
Income	0.03^{*}	0.01
	(0.01)	(0.01)
Political Attenion	0.19^{*}	0.09
	(0.05)	(0.05)
# Nonresponse 1 (Like Dem + Dislike Rep) x Ind	-0.82^{*}	
	(0.26)	
# Nonresponse 2 (Like Dem + Dislike Rep) x Ind	-0.80	
	(0.42)	
# Nonresponse 1 (Like Dem + Dislike Rep) x Rep	-0.99^{*}	
	(0.37)	
# Nonresponse 2 (Like Dem + Dislike Rep) x Rep	-2.20^{*}	
	(0.68)	
# Nonresponse 1 (Like Rep + Dislike Dem) x Ind		0.11
		(0.40)
# Nonresponse 2 (Like Rep + Dislike Dem) x Ind		0.16
		(0.48)
# Nonresponse 1 (Like Rep + Dislike Dem) x Rep		1.00^{*}
		(0.40)
# Nonresponse 2 (Like Rep + Dislike Dem) x Rep		1.42^{*}
		(0.48)
Constant	1.59^{*}	-5.50^{*}
	(0.55)	(0.66)
Observations	3,203	3,203
Log Likelihood	-1,134.58	-1,125.75
Akaike Inf. Crit.	2,313.15	2,295.49
Note:		$^{*}p < 0.05$

A.1.5 Study 1a: Tables of Likelihood Ratio Test for Addition of Partisan Writing Scale

Models used in Table A.5

- Model 1: Vote Clinton Party ID + Racial Resentment + Hostile Sexism + Authoritarianism + Survey Mode + Education + Age + Female + Race + Income + Political Attention
- Model 2: Vote Clinton Partisan Scale x Party ID + Racial Resentment + Hostile Sexism + Authoritarianism + Survey Mode + Education + Age + Female + Race + Income + Political Attention

Table A.5: Table of Likelihood Ratio Test for Vote Dem vs Models with and without Partisan Writing Scale Interacted with Party ID

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
3186	2548.67			
3183	2156.34	3	392.33	<0.00001

Models used in Table A.6

- Model 1: Vote Trump Party ID + Racial Resentment + Hostile Sexism + Authoritarianism + Survey Mode + Education + Age + Female + Race + Income + Political Attention
- Model 2: Vote Trump Partisan Scale x Party ID + Racial Resentment + Hostile Sexism + Authoritarianism + Survey Mode + Education + Age + Female + Race + Income + Political Attention

Table A.6: Table of Likelihood Ratio Test for Vote Rep vs Models with and without Addition of Partisan Writing Scale Interacted with Party ID

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
3186	2542.76			
3183	2063.37	3	479.39	<0.00001

Models used in Table A.7

- Model 1: Vote Clinton Party ID + Racial Resentment + Hostile Sexism + Authoritarianism + Survey Mode + Education + Age + Female + Race + Income + Political Attention
- Model 2: Vote Clinton Partisan Scale + Party ID + Racial Resentment + Hostile Sexism + Authoritarianism + Survey Mode + Education + Age + Female + Race + Income + Political Attention

Table A.7: Table of Likelihood Ratio Test for Vote Dem vs Models with and without Partisan Writing Scale, No Interaction with Party ID

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
3186	2548.67			
3185	2186.25	1	362.42	<0.00001

Models used in Table A.8

- Model 1: Vote Trump Party ID + Racial Resentment + Hostile Sexism + Authoritarianism + Survey Mode + Education + Age + Female + Race + Income + Political Attention
- Model 2: Vote Trump Partisan Scale + Party ID + Racial Resentment + Hostile Sexism + Authoritarianism + Survey Mode + Education + Age + Female + Race + Income + Political Attention

Table A.8: Table of Likelihood Ratio Test for Vote Rep vs Models with and without Partisan Writing Scale, No Interaction with Party ID

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
3186	2542.76			
3185	2091.45	1	451.31	<0.00001

A.1.6 Study 1a: Tables of Pseudo R^2 for Full and Reduced Models

Table A.9: Table of Pseudo R^2 for Full vs Reduced Models with and without Partisan Writing Scale, Respectively, Interacted with Party ID

	McFadden	McFadden Adj	Cox Snell	Nagelkerke	Aldrich Nelson	Veall Zimmermann
Full Model	0.49	0.48	0.47	0.65	0.39	0.69
Reduced Model	0.39	0.39	0.40	0.55	0.34	0.60
Percent Change	23.75	23.88	17.08	17.08	14.50	14.50

Table A.10: Table of Pseudo R^2 for Full vs Reduced Models with and without Partisan Writing Scale, Respectively, Interacted with Party ID

	McFadden	McFadden Adj	Cox Snell	Nagelkerke	Aldrich Nelson	Veall Zimmermann
Full Model	0.49	0.48	0.46	0.64	0.38	0.68
Reduced Model	0.37	0.36	0.37	0.52	0.32	0.57
Percent Change	32.09	32.43	23.39	23.39	19.86	19.86

A.1.7 Study 1a: Table of Self-Reported Vote Choice vs Writing Partisanship Scale

Table A.11: Self-Reported Vote Choice vs Writing Partisanship Scale

	Dependent variable:	
	Vote Clinton	Vote Trump
	logistic	*logistic*
	(1)	(2)
Writing Partisanship Scale	-1.36^{*}	3.31^{*}
	(0.16)	(0.34)
Party: Independent	-1.46^{*}	0.75^{*}
	(0.16)	(0.20)
Party: Republican	-2.15^{*}	1.87^{*}
	(0.21)	(0.21)
Racial Resentment	-0.39^{*}	0.35^{*}
	(0.06)	(0.07)
Hostile Sexism	-0.20^{*}	0.09
	(0.07)	(0.07)
Authoritarianism	-0.16^{*}	0.02
	(0.05)	(0.05)
Mode: Web	-0.15	-0.02
	(0.12)	(0.13)
Education	0.08^{*}	0.05
	(0.03)	(0.03)
Age (yrs)	0.08^{*}	0.10^{*}
	(0.02)	(0.02)
Female	-0.03	0.18
	(0.12)	(0.12)
Race: Black	0.31	0.67
	(0.32)	(0.50)
Race: Hispanic	-0.22	0.29
	(0.31)	(0.41)
Race: Native American	0.39	-1.51
	(0.74)	(1.25)
Race: Other	0.20	0.37
	(0.39)	(0.45)
Race: White	-0.15	0.86^{*}
	(0.28)	(0.35)
Income	0.03^{*}	0.02^{*}
	(0.01)	(0.01)
Political Attenion	0.23^{*}	0.09
	(0.05)	(0.06)
Partisanship Scale x Ind	-0.77^{*}	-0.84^{*}
	(0.23)	(0.38)
Partisanship Scale x Rep	-1.80^{*}	-1.71^{*}
	(0.38)	(0.38)
Constant	-0.30	-6.36^{*}
	(0.57)	(0.67)
Observations	3,203	3,203
Log Likelihood	-1,078.17	-1,031.69
Akaike Inf. Crit.	2,196.34	2,103.37
Note:		$^{*}p < 0.05$

A.2 Study 1b

A.2.1 Study 1b: Plot of Validated Turnout vs Most Important Problem Writing Scale interacted with Party ID

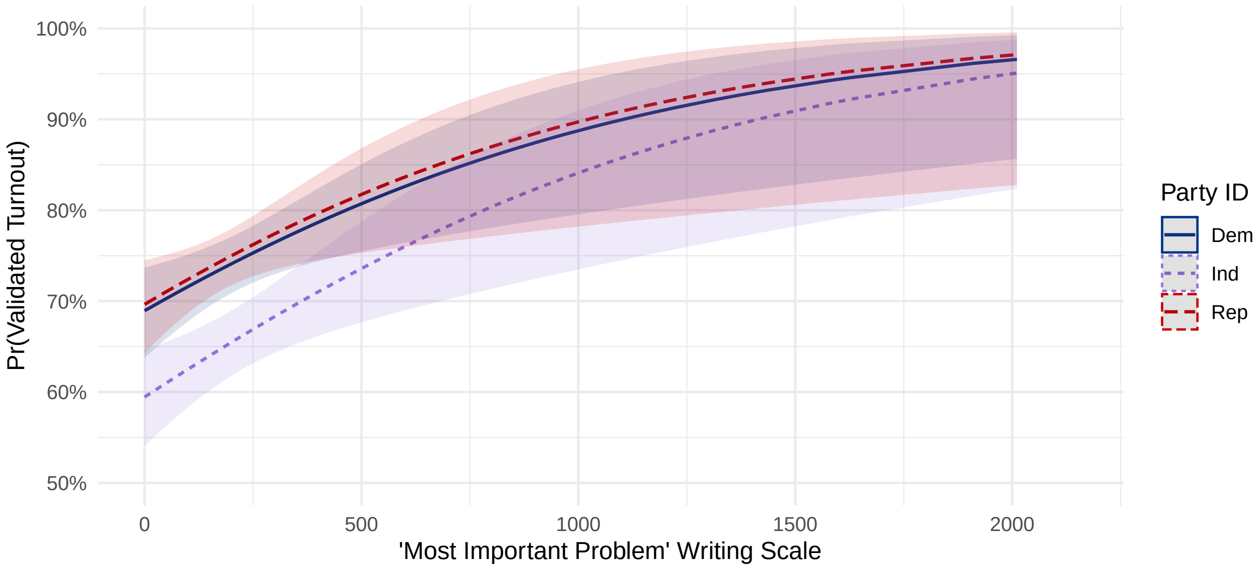


Figure A.2: Marginal effects of validated turnout versus number of characters combined with weighted nonresponse for 'most important problem' questions. Interaction with party identification not shown due to minimal statistical and substantive differences by partisanship. Model controls include party ID, female, education, age, race, income, and political interest.

A.3 Study 1b

A.3.1 Study 1b: Table of Validated Turnout vs Most Important Problem Writing Scale, with and without Interactions with Party ID

Table A.12: Validated Turnout vs Most Important Problem Writing Scale Combining Number of Characters and Weighted Nonresponse

	Dependent variable:		
	Validated Turnout		
	logistic		
	(1)	(2)	(3)
Most Imp Prob Scale (# Chars + Nonresp)	0.001^{*}	0.001^{*}	0.001^{*}
	(0.0002)	(0.0003)	(0.0004)
Party: Independent		-0.41^{*}	-0.41^{*}
		(0.09)	(0.15)
Party: Republican		0.05	0.03
		(0.10)	(0.16)
Mode: Web		-0.04	-0.04
		(0.09)	(0.09)
Female		0.22^{*}	0.22^{*}
		(0.08)	(0.08)
Education		0.11^{*}	0.11^{*}
		(0.02)	(0.02)
Age (yrs)		0.15^{*}	0.15^{*}
		(0.01)	(0.01)
Race: Hispanic		-0.16	-0.16
		(0.16)	(0.16)
Race: Other		-0.23	-0.23
		(0.18)	(0.18)
Race: White		0.09	0.09
		(0.13)	(0.13)
Income		0.04^{*}	0.04^{*}
		(0.01)	(0.01)
Political Attenion		0.13^{*}	0.13^{*}
		(0.04)	(0.04)
Most Imp Prob Scale x Ind			0.0000
			(0.001)
Most Imp Prob Scale x Rep			0.0001
			(0.001)
Constant	0.26^{*}	-2.95^{*}	-2.95^{*}
	(0.05)	(0.25)	(0.26)
Observations	4,270	3,758	3,758
Log Likelihood	-2,777.04	-2,160.33	-2,160.32
Akaike Inf. Crit.	5,558.07	4,346.65	4,350.64
Note:			$^{*}p < 0.05$

A.3.2 Study 1b: Tables of Likelihood Ratio Test for Addition of Partisan Writing Scale

Models used in Table A.13

- Model 1: Validated Turnout Party ID + Survey Mode + Female + Education + Age + Race + Income + Political Attention
- Model 2: Validated Turnout MIP Scale + Party ID + Survey Mode + Female + Education + Age + Race + Income + Political Attention

Table A.13: Likelihood Ratio Test Comparing Full Model with MIP Writing Scale to Reduced Model without MIP Writing Scale, No Interaction with Party ID

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
3746	4346.48			
3745	4320.65	1	25.83	<0.00001

Models used in Table A.14

- Model 1: Validated Turnout Party ID + Survey Mode + Female + Education + Age + Race + Income + Political Attention
- Model 2: Validated Turnout MIP Scale x Party ID + Survey Mode + Female + Education + Age + Race + Income + Political Attention

Table A.14: Likelihood Ratio Test Comparing Full Model with MIP Writing Scale to Reduced Model without MIP Writing Scale, Interacted with Party ID

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
3746	4346.48			
3743	4320.64	3	25.84	0.00001

A.3.3 Study 1b: Table of Pseudo R^2 for Full and Reduced Models

Table A.15: Table of Pseudo R^2 for Full vs Reduced Models with and without MIP Writing Scale, Respectively, No Interaction with Party ID

	McFadden	McFadden Adj	Cox Snell	Nagelkerke	Aldrich Nelson	Veall Zimmermann
Full Model	0.118	0.112	0.142	0.195	0.133	0.235
Reduced Model	0.112	0.108	0.136	0.187	0.128	0.226
Percent Change	4.690	4.520	4.340	4.340	4.070	4.070

Table A.16: Table of Pseudo R^2 for Full vs Reduced Models with and without MIP Writing Scale, Respectively, Interacted with Party ID

	McFadden	McFadden Adj	Cox Snell	Nagelkerke	Aldrich Nelson	Veall Zimmermann
Full Model	0.12	0.11	0.14	0.20	0.13	0.24
Reduced Model	0.11	0.11	0.14	0.19	0.13	0.23
Percent Change	4.69	3.77	4.34	4.34	4.07	4.07

A.3.4 Study 1b: Table of Validated Turnout vs Nonresponse in Candidate-Affect Prompts

Table A.17: Validated turnout vs nonresponse in candidate-affect prompts

	Dependent variable:	
	Validated Turnout	
	logistic	
	(1)	(2)
# Nonresponse 1 (Like Dem + Dislike Rep)	-0.37^{*}	
	(0.16)	
# Nonresponse 2 (Like Dem + Dislike Rep)	-0.78^{*}	
	(0.22)	
# Nonresponse 1 (Like Rep + Dislike Dem)		0.17
		(0.23)
# Nonresponse 2 (Like Rep + Dislike Dem)		0.14
		(0.21)
Party: Independent	-0.43^{*}	-0.11
	(0.13)	(0.22)
Party: Republican	-0.41^{*}	0.34
	(0.21)	(0.21)
Mode: Web	-0.10	-0.10
	(0.09)	(0.09)
Education	0.11^{*}	0.12^{*}
	(0.02)	(0.02)
Age (yrs)	0.15^{*}	0.15^{*}
	(0.01)	(0.01)
Female	0.21^{*}	0.25^{*}
	(0.08)	(0.08)
Race: Black	0.32	0.38
	(0.23)	(0.23)
Race: Hispanic	0.19	0.23
	(0.23)	(0.23)
Race: Native American	-0.03	-0.002
	(0.55)	(0.55)
Race: Other	0.25	0.27
	(0.27)	(0.27)
Race: White	0.47^{*}	0.44^{*}
	(0.20)	(0.20)
Income	0.04^{*}	0.04^{*}
	(0.01)	(0.01)
Political Attenion	0.13^{*}	0.12^{*}
	(0.04)	(0.04)
# Nonresponse 1 (Like Dem + Dislike Rep) x Ind	0.22	
	(0.21)	
# Nonresponse 2 (Like Dem + Dislike Rep) x Ind	0.44	
	(0.26)	
# Nonresponse 1 (Like Dem + Dislike Rep) x Rep	0.73^{*}	
	(0.27)	
# Nonresponse 2 (Like Dem + Dislike Rep) x Rep	1.06^{*}	
	(0.30)	
# Nonresponse 1 (Like Rep + Dislike Dem) x Ind		-0.35
		(0.27)
# Nonresponse 2 (Like Rep + Dislike Dem) x Ind		-0.50
		(0.26)
# Nonresponse 1 (Like Rep + Dislike Dem) x Rep		-0.66^{*}
		(0.28)
# Nonresponse 2 (Like Rep + Dislike Dem) x Rep		-0.98^{*}
		(0.33)
Constant	-2.81^{*}	-3.24^{*}
	(0.32)	(0.36)
Observations	3,758	3,758
Log Likelihood	-2,161.16	-2,161.60
Akaike Inf. Crit.	4,362.32	4,363.20
Note:		$^{*}p < 0.05$

A.3.5 Study 1b: Plot of Validated Turnout vs Partisanship Writing Scale

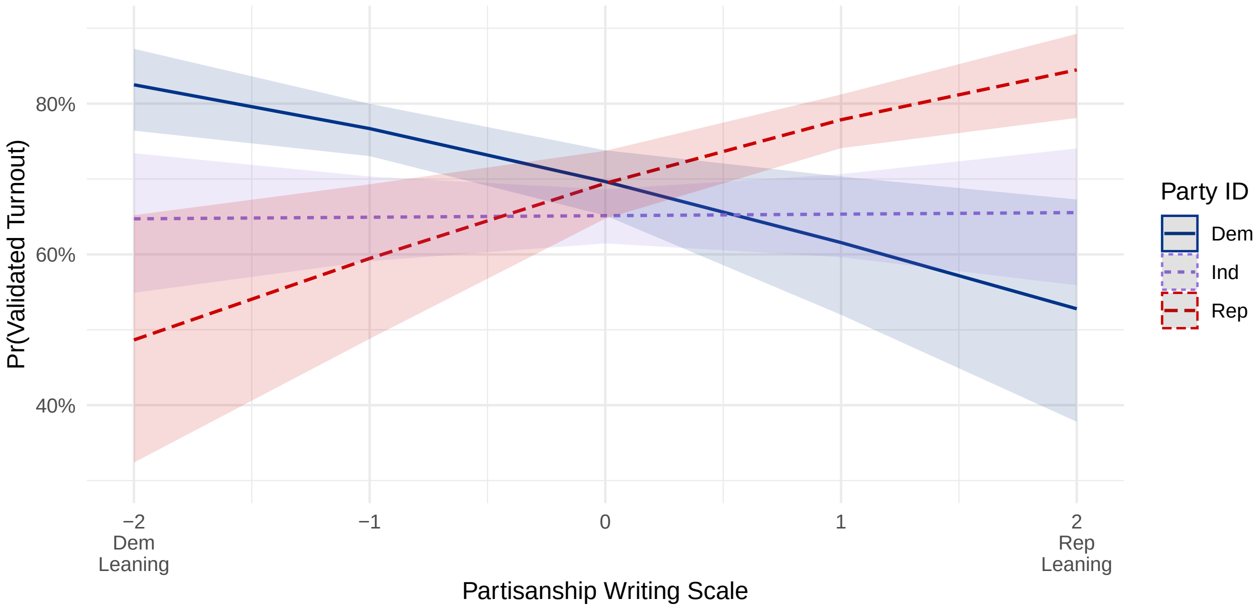


Figure A.3: Marginal effect plot of validated turnout versus Partisanship Writing Scale interacted with party identification.

Table A.18: Validated Turnout vs Partisanship Writing Scale, by Party ID

	Dependent variable:
	Validated Turnout
	logistic
Writing Partisanship Scale	-0.36^{*}
	(0.12)
Party: Independent	-0.21
	(0.11)
Party: Republican	-0.01
	(0.14)
Mode: Web	-0.13
	(0.08)
Female	0.23^{*}
	(0.08)
Education	0.12^{*}
	(0.02)
Age (yrs)	0.15^{*}
	(0.01)
Race: Hispanic	-0.12
	(0.16)
Race: Other	-0.19
	(0.18)
Race: White	0.13
	(0.13)
Income	0.04^{*}
	(0.01)
Political Attenion	0.13^{*}
	(0.04)
Partisanship Scale x Ind	0.37^{*}
	(0.15)
Partisanship Scale x Rep	0.80^{*}
	(0.18)
Constant	-2.96^{*}
	(0.26)
Observations	3,758
Log Likelihood	-2,163.28
Akaike Inf. Crit.	4,356.56
Note:	$^{*}p < 0.05$

A.3.6 Study 1b: Plots of Validated Turnout vs Candidate-Affect Prompts

In addition to the Partisan Writing Scale, I attempt to weight nonresponse via a similar method to the Most Important Problem Writing Scale by assigning nonresponse a negative value at about the mean number of characters for that battery of questions, in this case -50. Equation 5 shows how nonresponse and the number of characters from all four candidate-affect questions are combined in what I call the Candidate-Affect Writing Scale or simply "# Characters + Weighted Nonresponse."

$$\text{\# Characters + Weighted Nonresponse} \quad = \quad 200 + \sum_{n=1}^{4} \Big(-50 \times non(\text{Candidate-Affect}_i)\Big) + \Big(nchar(\text{Candidate-Affect}_i)\Big) \tag{5}$$

Figure A.4 presents four plots predicting validated turnout using either just the number of characters (Panels A and C) or the combination of character counts and weighted nonresponse (Panels B and D). In addition, Panels C and D show the same results as Panels A and B but moderated by party identification. While overall the writing measures are good predictors of turnout, as with earlier analyses, there is noteworthy heterogeneity by party identification, particularly for Republicans, in how writing predicts turnout.

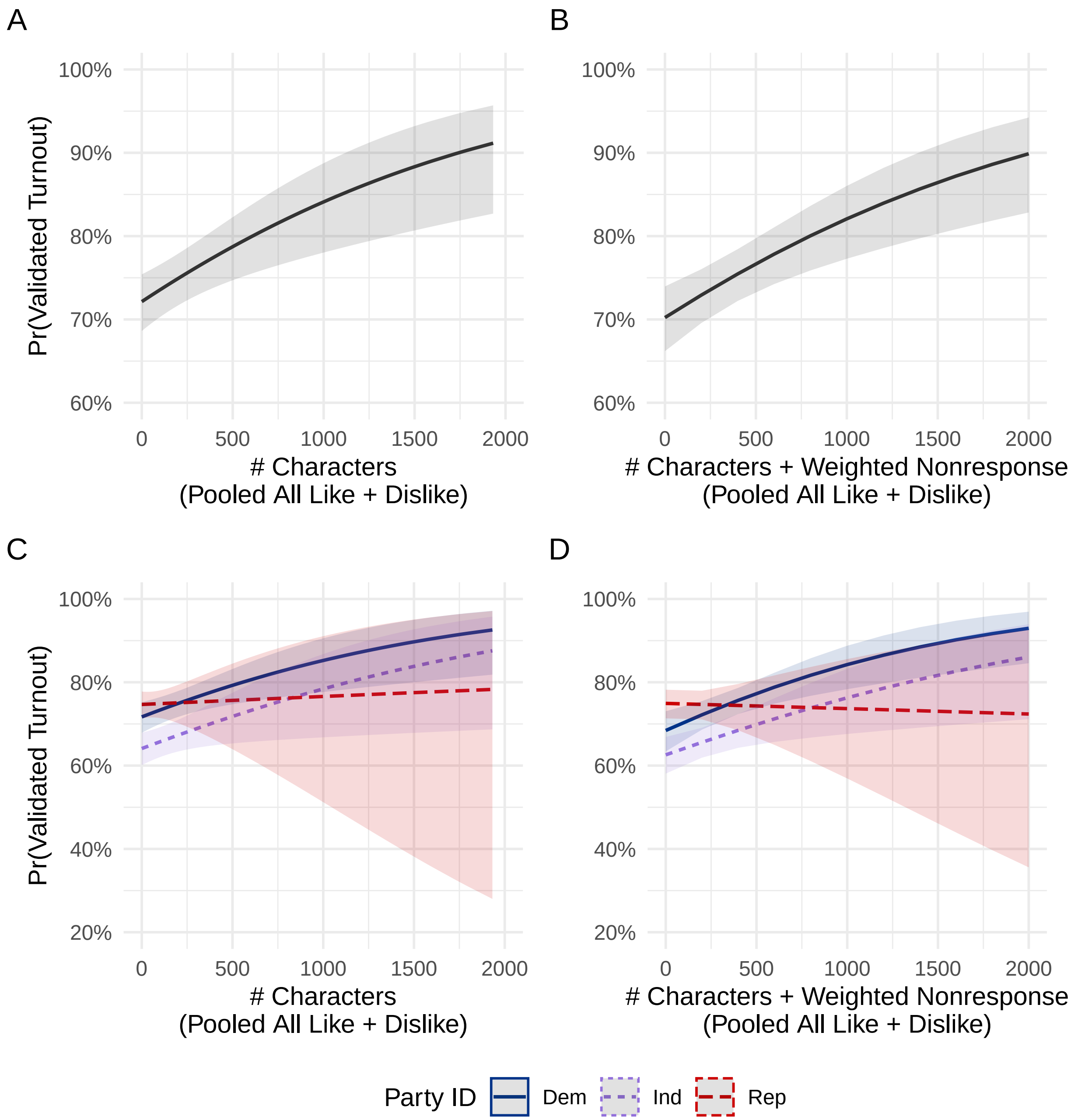


Figure A.4: Four plots predicting validated turnout using the number of characters (Panels A and C) or the combination of character counts and weighted nonresponse (Panels B and D). In addition, Panels C and D show the same results as Panels A and B but moderated by party identification.

Table A.19: Validated Vote vs Number of Characters *without* Nonresponse for Congruent Candidate-Affect Prompt Pairs

	Dependent variable:		
	Validated Turnout		
	logistic		
	(1)	(2)	(3)
# Chars (Pooled Like + Dislike)	0.001^{*}	0.001^{*}	0.001^{*}
	(0.0002)	(0.0002)	(0.0003)
Party: Independent		-0.37^{*}	-0.35^{*}
		(0.09)	(0.11)
Party: Republican		0.10	0.15
		(0.10)	(0.11)
Mode: Web		0.02	0.01
		(0.09)	(0.09)
Female		0.23^{*}	0.22^{*}
		(0.08)	(0.08)
Education		0.11^{*}	0.11^{*}
		(0.02)	(0.02)
Age (yrs)		0.15^{*}	0.15^{*}
		(0.01)	(0.01)
Race: Hispanic		-0.15	-0.15
		(0.16)	(0.16)
Race: Other		-0.22	-0.22
		(0.18)	(0.18)
Race: White		0.12	0.12
		(0.13)	(0.13)
Income		0.04^{*}	0.04^{*}
		(0.01)	(0.01)
Political Attenion		0.14^{*}	0.14^{*}
		(0.04)	(0.04)
# Chars (Pooled) x Ind			-0.0001
			(0.0004)
# Chars (Pooled) x Rep			-0.001
			(0.001)
Constant	0.47^{*}	-2.93^{*}	-2.94^{*}
	(0.04)	(0.25)	(0.25)
Observations	4,270	3,758	3,758
Log Likelihood	-2,791.91	-2,166.45	-2,165.84
Akaike Inf. Crit.	5,587.82	4,358.89	4,361.67
Note:			$^{*}p < 0.05$

Table A.20: Validated Vote vs Number of Characters *with* Weighted Nonresponse for Congruent Candidate-Affect Prompt Pairs

	Dependent variable:		
	Validated Turnout		
	logistic		
	(1)	(2)	(3)
# Chars + Nonresp (Pooled Like + Dislike)	0.001^{*}	0.001^{*}	0.001^{*}
	(0.0001)	(0.0002)	(0.0003)
Party: Independent		-0.34^{*}	-0.26
		(0.09)	(0.13)
Party: Republican		0.16	0.32^{*}
		(0.11)	(0.14)
Mode: Web		0.02	0.02
		(0.09)	(0.09)
Female		0.22^{*}	0.22^{*}
		(0.08)	(0.08)
Education		0.11^{*}	0.11^{*}
		(0.02)	(0.02)
Age (yrs)		0.15^{*}	0.15^{*}
		(0.01)	(0.01)
Race: Hispanic		-0.15	-0.15
		(0.16)	(0.16)
Race: Other		-0.21	-0.21
		(0.18)	(0.18)
Race: White		0.13	0.13
		(0.13)	(0.13)
Income		0.04^{*}	0.04^{*}
		(0.01)	(0.01)
Political Attenion		0.14^{*}	0.13^{*}
		(0.04)	(0.04)
# Chars + Nonresp (Pooled) x Ind			-0.0002
			(0.0003)
# Chars + Nonresp (Pooled) x Rep			-0.001^{*}
			(0.0005)
Constant	0.41^{*}	-2.99^{*}	-3.05^{*}
	(0.04)	(0.25)	(0.26)
Observations	4,270	3,758	3,758
Log Likelihood	-2,790.13	-2,164.99	-2,162.93
Akaike Inf. Crit.	5,584.25	4,355.98	4,355.87
Note:			$^{*}p < 0.05$

A.4 Study 2a

A.4.1 Study 2a: Table: Any Nonresponse vs Racial Resentment

Table A.21: Predicted probability of any nonresponse versus racial resentment

	Dependent variable:	
	Vote Clinton	Vote Trump
	logistic	*logistic*
	(1)	(2)
Racial Resentment	0.39^{*}	-0.84^{*}
	(0.07)	(0.10)
Party: Independent	0.48	-2.16^{*}
	(0.33)	(0.42)
Party: Republican	1.41^{*}	-3.43^{*}
	(0.49)	(0.49)
Hostile Sexism	0.31^{*}	-0.26^{*}
	(0.06)	(0.06)
Authoritarianism	0.15^{*}	-0.04
	(0.04)	(0.04)
Mode: Web	0.80^{*}	0.63^{*}
	(0.10)	(0.10)
Education	-0.13^{*}	-0.09^{*}
	(0.02)	(0.02)
Age (yrs)	-0.004	-0.01
	(0.01)	(0.01)
Female	-0.05	0.15
	(0.10)	(0.10)
Race: Black	-0.07	1.02^{*}
	(0.30)	(0.38)
Race: Hispanic	-0.28	0.44
	(0.29)	(0.31)
Race: Native American	0.08	0.81
	(0.69)	(0.78)
Race: Other	0.12	-0.39
	(0.34)	(0.35)
Race: White	0.35	-0.52
	(0.26)	(0.27)
Income	-0.01	0.005
	(0.01)	(0.01)
Political Attenion	-0.09^{*}	-0.27^{*}
	(0.05)	(0.05)
Racial Resentment x Ind	0.28^{*}	0.24^{*}
	(0.10)	(0.12)
Racial Resentment x Rep	0.36^{*}	0.30^{*}
	(0.14)	(0.14)
Constant	-2.17^{*}	7.07^{*}
	(0.52)	(0.61)
Observations	3,203	3,203
Log Likelihood	-1,475.50	-1,447.22
Akaike Inf. Crit.	2,989.01	2,932.45
Note:		$^{*}p < 0.05$

A.4.2 Study 2a: Number of Characters vs Hostile Sexism

Figure A.5 Panels A and B present the results of two logistic regression models in which a survey instrument for hostile sexism is used to predict the probability of any nonresponse to the two concordant candidate-affect questions. In Figure A.5 Panel A we see that as scores increase on a standard survey instrument for hostile sexism, the probability of an nonresponse increases significantly in response to questions about what subjects like about the Democratic nominee, Hillary Clinton, and dislike about the Republican nominee, Donald Trump. Further, while the intercepts vary significantly by party identification, the slopes are not statistically significanlly different. Conversely, in Panel B as the hostile sexism measure increases, subjects are significantly less likely to write nothing in response to prompts about what they like about the Republican nominee and dislike about the Democratic nominee. Again, while the intercepts are significantly different by party, the slopes are not statistically significantly different.

A.4.3 Study 2a: Hostile Sexism Question Battery

The hostile sexism questions used in the 2016 ANES were posed on a five-point scale from Agree strongly to Disagree strongly. The four questions were:

1. Many women interpret innocent remarks or acts as being sexist.
2. Most women fail to appreciate fully all that men do for them.
3. Women seek to gain power by getting control over men.
4. Once a woman gets a man to commit to her, she tries to put him on a tight leash.

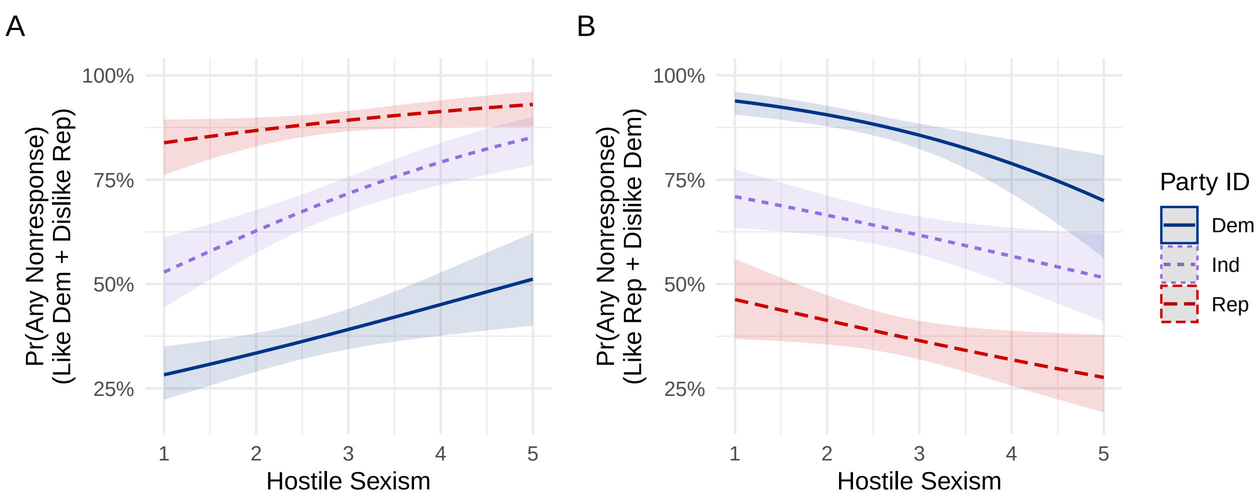


Figure A.5: Marginal effects of hostile sexism as a predictor of writing on partisan congruent like / dislike nominee prompts, by party ID. Model controls same as in Figure A.5.

A.4.4 Study 2a: Table for Any Nonresponse vs Hostile Sexism, by Party Identification

Table A.22: Any Nonresponse vs Hostile Sexism

	Dependent variable:	
	Like Dem + Dislike Rep	Like Rep + Dislike Dem
	logistic	*logistic*
	(1)	(2)
Hostile Sexism	0.24^{*}	-0.47^{*}
	(0.08)	(0.12)
Party: Independent	0.88^{*}	-2.10^{*}
	(0.33)	(0.41)
Party: Republican	2.59^{*}	-3.15^{*}
	(0.43)	(0.45)
Racial Resentment	0.56^{*}	-0.63^{*}
	(0.05)	(0.05)
Authoritarianism	0.14^{*}	-0.04
	(0.04)	(0.04)
Mode: Web	0.80^{*}	0.64^{*}
	(0.10)	(0.10)
Education	-0.13^{*}	-0.09^{*}
	(0.02)	(0.02)
Age (yrs)	-0.003	-0.01
	(0.01)	(0.01)
Female	-0.05	0.15
	(0.10)	(0.10)
Race: Black	-0.01	1.06^{*}
	(0.30)	(0.38)
Race: Hispanic	-0.28	0.44
	(0.29)	(0.31)
Race: Native American	0.03	0.80
	(0.69)	(0.78)
Race: Other	0.14	-0.39
	(0.34)	(0.35)
Race: White	0.35	-0.53^{*}
	(0.26)	(0.27)
Income	-0.01	0.004
	(0.01)	(0.01)
Political Attenion	-0.09	-0.27^{*}
	(0.05)	(0.05)
Hostile Sexism x Ind	0.16	0.26
	(0.12)	(0.14)
Hostile Sexism x Rep	-0.01	0.27
	(0.15)	(0.15)
Constant	-2.57^{*}	6.99^{*}
	(0.52)	(0.61)
Observations	3,203	3,203
Log Likelihood	-1,479.37	-1,448.11
Note:		$^{*}p < 0.05$

A.4.5 Study 2a: Table for Any Nonresponse vs Racial Resentment, by Party Identification

Table A.23: Any Nonresponse vs Racial Resentment

	Dependent variable:	
	Any Nonresp. (Like Dem + Dislike Rep)	Any Nonresp. (Like Rep + Dislike Dem)
	logistic	*logistic*
	(1)	(2)
Racial Resentment	0.39^{*}	-0.84^{*}
	(0.07)	(0.10)
Party: Independent	0.48	-2.16^{*}
	(0.33)	(0.42)
Party: Republican	1.41^{*}	-3.43^{*}
	(0.49)	(0.49)
Hostile Sexism	0.31^{*}	-0.26^{*}
	(0.06)	(0.06)
Authoritarianism	0.15^{*}	-0.04
	(0.04)	(0.04)
Mode: Web	0.80^{*}	0.63^{*}
	(0.10)	(0.10)
Education	-0.13^{*}	-0.09^{*}
	(0.02)	(0.02)
Age (yrs)	-0.004	-0.01
	(0.01)	(0.01)
Female	-0.05	0.15
	(0.10)	(0.10)
Race: Black	-0.07	1.02^{*}
	(0.30)	(0.38)
Race: Hispanic	-0.28	0.44
	(0.29)	(0.31)
Race: Native American	0.08	0.81
	(0.69)	(0.78)
Race: Other	0.12	-0.39
	(0.34)	(0.35)
Race: White	0.35	-0.52
	(0.26)	(0.27)
Income	-0.01	0.005
	(0.01)	(0.01)
Political Attenion	-0.09^{*}	-0.27^{*}
	(0.05)	(0.05)
Racial Resentment x Ind	0.28^{*}	0.24^{*}
	(0.10)	(0.12)
Racial Resentment x Rep	0.36^{*}	0.30^{*}
	(0.14)	(0.14)
Constant	-2.17^{*}	7.07^{*}
	(0.52)	(0.61)
Observations	3,203	3,203
Log Likelihood	-1,475.50	-1,447.22
Akaike Inf. Crit.	2,989.01	2,932.45
Note:		$^{*}p < 0.05$

A.4.6 Study 2a: Racial Resentment Question Battery

The racial resentment questions used in the 2016 ANES are asked on a five-point scale from Agree strongly to Disagree strongly. The four questions are:

1. Irish, Italians, Jewish and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors.
2. Generations of slavery and discrimination have created conditions that make it difficult for blacks to work their way out of the lower class.
3. Over the past few years, blacks have gotten less than they deserve.
4. It's really a matter of some people not trying hard enough, if blacks would only try harder they could be just as well off as whites.

A.5 Study 2b

A.5.1 Study 2b: Table for Afrobarometer Summary Statistics

Table A.24: Frequency table of Languages within Afrobarometer

Language	n	Percent
English	34,838	65%
French	14,116	26%
Portuguese	4,707	9%
NA	274	1%
Total	53,935	-

A.5.2 Study 2b: Table for Afrobarometer Regression Results

Table A.25: Number of characters about democracy means vs democracy battery interacted with language

	Dependent variable:			
	Number of Characters: 'What Democracy Means to You?'			
	English	French	Portuguese	Interaction
	(1)	(2)	(3)	(4)
Democratic Importance	0.07*	0.13*	0.16*	0.07*
	(0.01)	(0.01)	(0.04)	(0.01)
Lang: French				0.05
				(0.07)
Lang: Portuguese				−0.47*
				(0.11)
Gender	−0.16*	−0.20*	−0.08	−0.16*
	(0.01)	(0.02)	(0.05)	(0.01)
Education	0.12*	0.05*	0.10*	0.09*
	(0.004)	(0.004)	(0.01)	(0.003)
Age (yrs)	0.001	0.003*	0.01*	0.002*
	(0.0005)	(0.001)	(0.002)	(0.0004)
Income	−0.002	−0.01*	0.01	−0.005*
	(0.002)	(0.002)	(0.01)	(0.001)
Dem Impt x French				0.04*
				(0.02)
Dem Impt x Portuguese				0.07*
				(0.03)
Constant	2.98*	3.19*	2.09*	3.03*
	(0.05)	(0.07)	(0.19)	(0.05)
Observations	30,140	13,370	3,459	46,969
Log Likelihood	-134,543.30	-61,231.97	-14,127.98	-210,641.30
Akaike Inf. Crit.	269,098.60	122,476.00	28,267.95	421,302.50
Note:				*$p < 0.05$

A.6 Study 3a: Manipulation Check for Social Exclusion and Asian Americans

Table A.26: Negative binomial models of time vs social exclusion treatment separately for (1) Asian Americans; (2) Whites; and (3) together with an interaction term.

	Dependent variable:		
	Time (Minutes)		
	Asian-Americans	Whites	Asian-Am + White
	(1)	(2)	(3)
Social Exclusion Treatment	0.31*	−0.09	−0.09
	(0.10)	(0.09)	(0.10)
Race: Asian (vs White)			−0.28*
			(0.09)
Treated x Asian			0.41*
			(0.13)
Constant	2.62*	2.90*	2.90*
	(0.07)	(0.06)	(0.06)
Observations	61	53	114
Log Likelihood	-194.77	-165.59	-360.15
Akaike Inf. Crit.	393.53	335.17	728.29
Note:			$^{*}p < 0.05$

Figure A.6 presents the estimated joint effect of race and treatment condition on time with a negative binomial model. Once again, for whites, there is no significant change in time to completion for treated versus control subjects. In contrast, for Asian American subjects, there is a significant increase in time to completion.

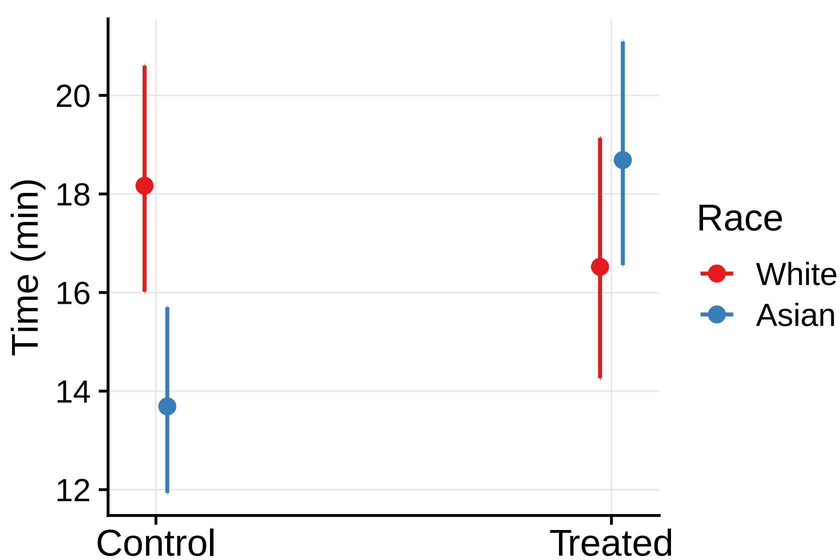


Figure A.6: Marginal effect of predicted total time on survey in minutes versus treatment condition, by race, with Negative Binomial model and standard set of controls.

A.7 Study 3b: Manipulation Check for African American Cross-Pressure

Figure A.7 presents bar plots showing which candidate was preferred in each of the conditions. In the control condition (Panel A) in which the identities of the two candidates were similar and the only distinguishing traits were policy, more than 80% of African American subjects preferred the more liberal candidate. In Panel B, the liberal candidate's religion is described as "Muslim," while all other traits are held constant, and we can see a modest but statistically significant shift in support to the more moderate candidate who is still described as "straight" and "Christian." In Panel C, the liberal candidate's sexuality was described as "gay" while his religion remained "Christian." Under this condition, substantially more subjects switch to support the politically moderate candidate described as "straight" and "Christian."

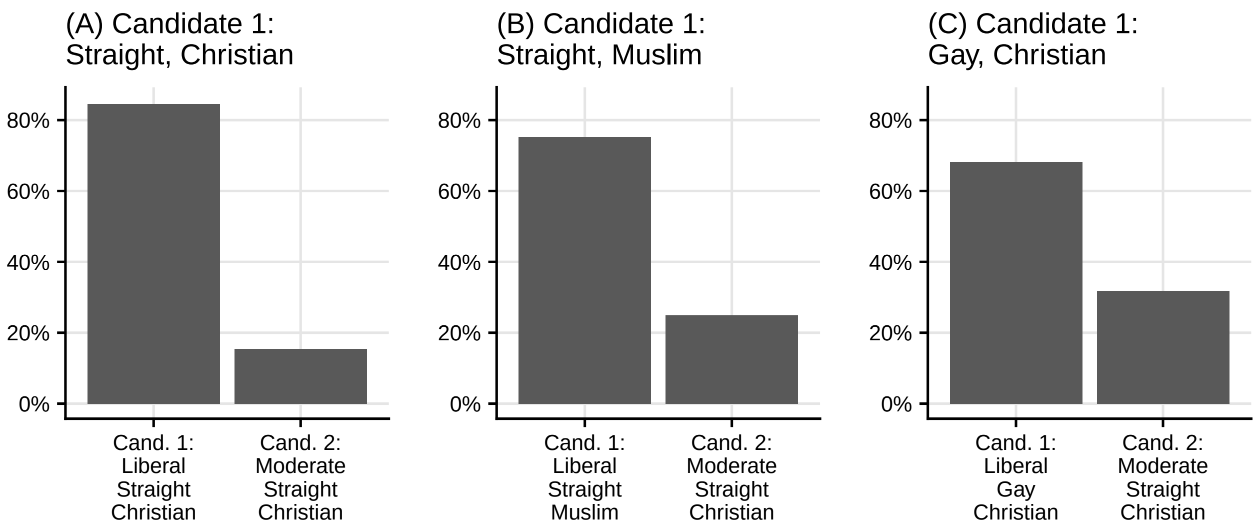


Figure A.7: Bar plot of support among African American subjects for two hypothetical Black candidates with varying ideology and demographic characteristics. In Panel A, the control condition, both candidates are described as 'straight' and 'Christian,' and more than 80% of subjects prefer the liberal candidate. In Panel B, the liberal candidate's religion is switched to 'Muslim' and a statistically significant number of voters opt for the moderate candidate. In Panel (C), the liberal candidate's sexual orientation is switched to 'Gay' and a larger number of subjects opt for the moderate candidate.

Figure A.8 presents the forced choice voting options in the control condition.[5] In the Gay Christian condition, every field was the same as in the control condition but the sexuality row was changed to Gay for Candidate 1. In the Straight Muslim Condition, all fields were the same as in the control condition but religion

[5]Candidate 1's platform includes a typographic error that the candidate 'Seeks to implement new work initiatives to reduce the employment rate,' rather than 'reduce the *un*employment rate.' I opt to ignore the error for two reasons. First, in the control condition, approximately 84.6% of subjects supported Candidate 1 suggesting the error had little to no influence on subjects' interpretation of the hypothetical candidate's commitment to Black-oriented policies. Second, the error was present in all the conditions and, therefore, is unlikely to have been a source of variation in the amount of writing or nonresponse across conditions.

was changed to Muslim for Candidate 1. The description of Candidate 2 remains the same in all conditions.

	Candidate 1	Candidate 2
Religion	Christian	Christian
Race/Ethnicity	Black	Black
Sexuality	Straight	Straight
Age	37	40
Gender	Male	Male
Military Service	Served in U.S. military	Served in U.S. military
Party Affiliation	Democrat	Democrat
Policy Beliefs	✓ Strong advocate for criminal justice reform and addressing police brutality ✓ Seeks to implement new work initiatives to reduce the employment rate ✓ Wants to increase funding for social security and social welfare programs that benefit families and children	✓ Wants to work with fellow legislators to reduce the federal debt ✓ Wants to combat domestic and international terrorism ✓ Strong advocate for comprehensive immigration reform

Figure A.8: The forced choice voting options in the control condition. In the Gay Christian condition, sexuality is changed to Gay for Candidate 1. In the Straight Muslim Condition, religion is changed to Muslim for Candidate 1. The description of Candidate 2 remains the same in all conditions.

To statistically test for these two cross-pressure effects (nonresponse and, potentially, overresponse), I run four models presented in Table **??**. The first two models test for nonresponse, here coded as entering zero characters in the open-ended prompt. Table **??** Models (1) and (2) show that with logistic regression, both treatment conditions are significantly predictive of nonresponse. The coefficient in Table **??** Model (1) indicates that subjects in the "Straight Muslim" condition were about 81% more likely to enter zero characters at the open prompt as compared to the control condition. The results in Model 2 suggest that subjects in the "Gay Christian" condition were about 156% more likely to engage in nonresponse versus subjects given the "Straight Christian" condition. Figure 12 plots the marginal effects presented in Models (1) and (2).

In Table **??** Models (3) and (4) present negative binomial models estimating the total number of characters typed in the open-ended response by treatment condition. Models (3) and (4) show a positive but not statistically significant relationship between the randomly assigned treatment condition and the number of characters typed. This suggests there is not a statistically significant tendency towards overresponse.[6]

[6]Zero-inflated methods that attempt model nonresponse and overresponse as two separate data generating processes produced similar results so the simpler models are presented for ease of interpretation.

Table A.27: Odds ratio for treatment vs control conditions on nonresponse and number of characters used in open response about why subject chose preferred candidate (using Logistic and Negative Binomial models). NOTE: for ease of interpretation, coefficients are exponentiated. Standard errors are transformed by multiplying the original log-odds standard error by the respective exponentiated coefficient. The original p-values are used untransformed. Coefficients that are larger than about two times the presented standard error may still have p-values greater than 0.05.

	Dependent variable:			
	Zero Characters		Number of characters	
	logistic		*negative binomial*	
	(1)	(2)	(3)	(4)
Straight Muslim vs Straight Christian	1.81^{*}			
	(0.35)			
tr_gay		2.56^{*}		
		(0.48)		
tr_muslim_fctStraight Muslim			0.92	
			(0.10)	
tr_gay_fctStraight Christian				1.17
				(0.14)
Constant	0.18^{*}	0.18^{*}	75.79^{*}	64.68^{*}
	(0.03)	(0.03)	(5.69)	(5.35)
Observations	708	692	708	692
Log Likelihood	-351.29	-364.58	-3,661.43	-3,499.90
Akaike Inf. Crit.	706.58	733.16	7,326.85	7,003.80
Note:				$^{*}p < 0.05$

A.8 Study 5

A.8.1 Study 5: Text as Events in Time: Evidence from CCES and Political Engagement

Another possible application of the text as behavior approach could be to detect contextual effects, such as in response to events over time. Event detection with textual data is common. For example, prior research has used the Twitter stream as a "social sensor" to detect earthquakes (Sakaki, Okazaki, and Matsuo 2010), protest activity (Steinert-Threlkeld 2017), and individual-level behavior in response to major events (Eady, Hjorth, and Dinesen 2022). With survey data, however, the use of text related to events is much less common. The 2016 Cooperative Congressional Election Survey (Ansolabehere and Schaffner 2017) included both a pre- and a post-election wave as well as some experiments in which white subjects were asked to respond to open-ended prompts about six different groups (Schaffner 2020). As the 2016 election was highly polarized around issues of race (Schaffner, MacWilliams, and Nteta 2016), I pool the open-ended responses into two groups: an explicitly racial group (i.e, Blacks, Mexicans, and Whites) and a facially non-racial group (i.e., politicians, the middle class and young people). I then calculate a total number of characters written by the subject for each group. For the purposes of this analysis, the relevant "treatment" is not an experimental manipulation conducted in the survey but, rather, whether the subject was part of the pre- or post-election wave. While this study was not a panel design with the same subjects taking the survey in each wave, an equivalent analysis could be conducted with panel data.

Figure A.9 presents the estimated joint effect of election timing and political interest on the amount of writing about different groups. The political interest scale goes from a measure of one (low) to four (high). The plots present the results of two negative binomial models in which the outcome is the number of characters for a pooled group and the predictor is an interaction term for election timing and political interest. Panel A shows that the total number of characters written about the three non-racial groups does not change significantly between the pre- and post-election period for subjects with either low or high political interest. In Panel B, by contrast, we see that for subjects with low political interest, the total amount of writing about the three racial groups decreases modestly but not statistically significantly while, for those with high political interest, there is a significant increase in the amount of writing. In short, the joint effect of election timing and political interest has no effect on the amount of writing about non-racial groups but a significant effect on the amount of writing about racial groups ($p < 0.01$). These results suggest a heightened salience of racial groups, but not non-racial groups, in the post-election cohort, and only for those at the higher end of the political interest scale.

A narrow application of this result might be as a manipulation check to confirm an "election treatment" is inducing hypothesized changes in subject behavior perhaps as moderated by something like political interest.

A slightly broader application might treat the "costly" behavior of writing as an outcome of interest on its own. For example, this result might reasonably be interpreted as evidence that an energizing effect of elections could extended to how subjects would behave when writing messages on Facebook or Twitter.

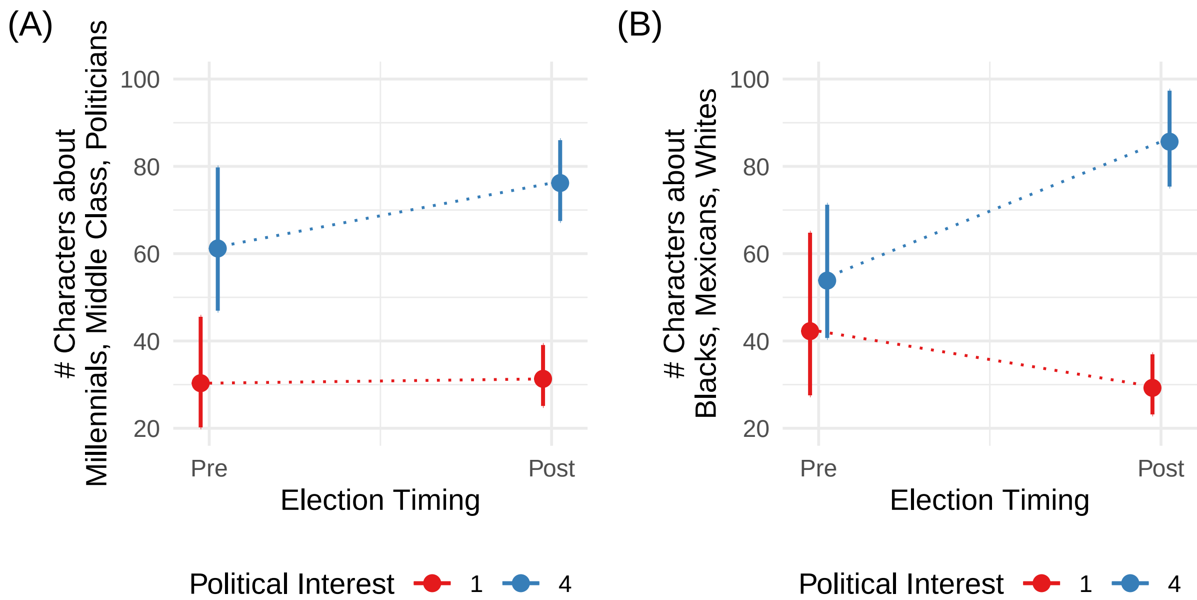


Figure A.9: Marginal effect of pre- vs post-election timing and political interest on number of characters written about three non-racial groups and three racial groups in 2016. Model controls for racial resentment, education, age, income, gender and union membership.

A.9 Study 5: CCES Text by Group Alternate Plot

Table A.28: Election timing on six open text prompts, interacted with political interest

	Dependent Variable: Number of Characters					
	Blacks	Mexicans	Whites	Politicians	Middle	Millenials
	(1)	(2)	(3)	(4)	(5)	(6)
Timing (Pre vs Post)	-0.78^*	-0.36	-0.27	-0.17	0.08	0.19
	(0.32)	(0.31)	(0.32)	(0.31)	(0.30)	(0.32)
Political Interest	-0.02	0.08	0.08	0.21^*	0.17	0.27^*
	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Timing x Pol Int	0.37^*	0.26^*	0.24^*	0.14	0.05	0.04
	(0.10)	(0.10)	(0.10)	(0.10)	(0.09)	(0.10)
Constant	3.01^*	2.42^*	2.36^*	2.23^*	2.25^*	1.84^*
	(0.30)	(0.28)	(0.29)	(0.28)	(0.27)	(0.29)
Pseudo R^2	0.009	0.010	0.009	0.009	0.005	0.009
Observations	1,191	1,191	1,191	1,191	1,191	1,191
Note:						$^*p < 0.05$

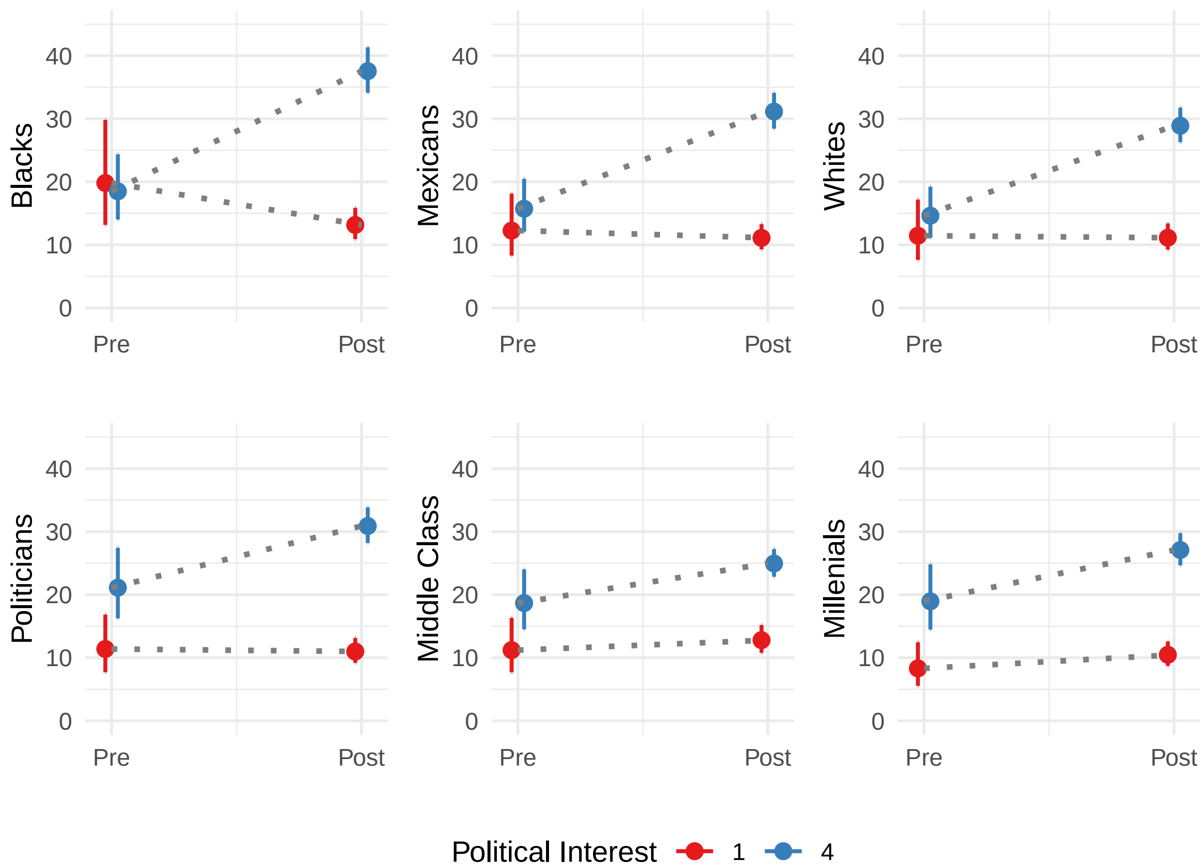
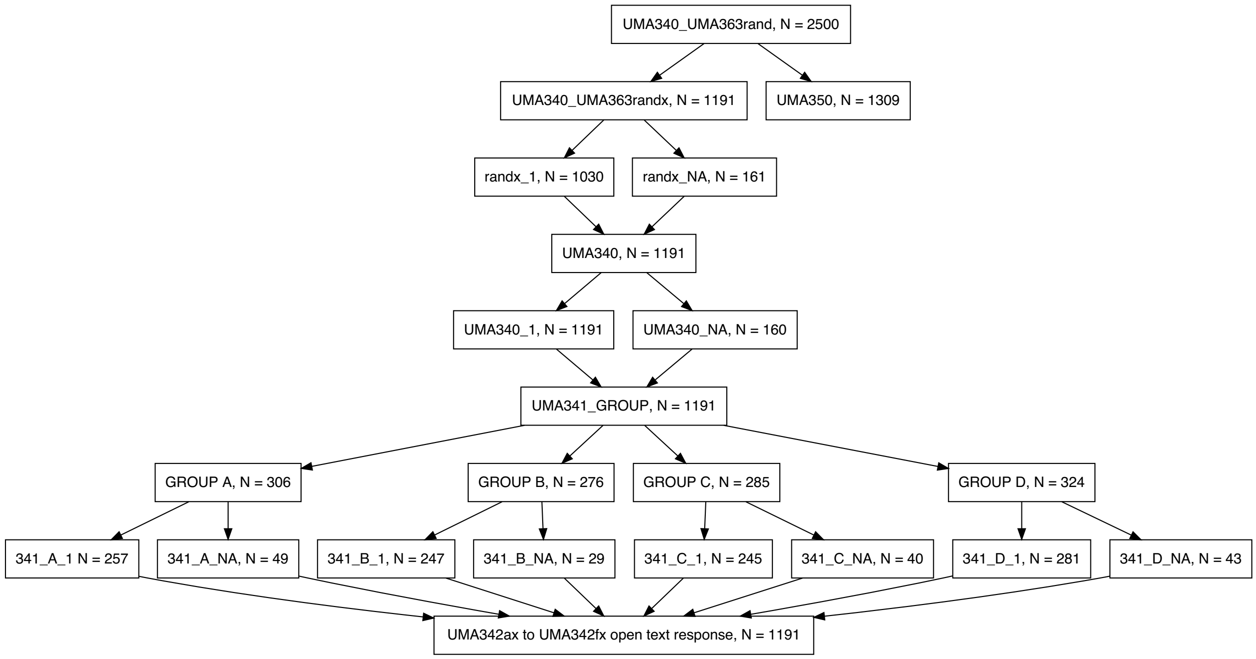


Figure A.10: Plot of joint effect of political interest and election timing (pre- vs post-) on number of characters written in open-ended prompts about each group. For responses about the three racial groups, there's not much difference by political interest in the pre-election period but significant separation in the post-election period. For the three non-racial groups, there is already sufficient separation by political interest in the pre-period that there is no significant additional joint effect of political interest and election timing (though for writing about both politicians and millennials, there is a significant political interest effect).

A.10 CCES Flowchart



%TC:endignore