Research Article

Psychological Science
2021, Vol. 32(1) 14–26

DOI: 10.1177/0956797620960397
www.psychologicalscience.org/PS

Moral-Language Use by U.S. Political Elites

Sze-Yuh Nina Wang and Yoel Inbar
Department of Psychology, University of Toronto

Abstract

We used a distributed-language model to examine the moral language employed by U.S. political elites. In Study 1, we analyzed 687,360 Twitter messages (tweets) posted by accounts belonging to Democratic and Republican members of Congress from 2016 to 2018. In Study 2, we analyzed 2,630,688 speeches given on the floor of the House and Senate from 1981 to 2017. We found that partisan differences in moral-language use shifted over time as the parties gained or lost political power. Overall, lower political power was associated with greater use of moral language for both Democrats and Republicans. On Twitter, Democrats used more moral language in the period after Donald Trump won the 2016 presidential election. In Congressional transcripts, both Democrats and Republicans used more of most kinds of moral language when they were in the minority.

Keywords

political ideology, moral language, political psychology, natural-language analysis, open data

Received 10/3/19; Revision accepted 7/12/20

Political elites use language to shift the public debate and to persuade and motivate voters (Lakoff, 2004). Moral language is particularly powerful, as voters respond especially strongly to moral language from elites (Clifford & Jerit, 2013). Once people connect an issue to their moral values, they are more likely to act to be sure their side prevails, be it by persuasion, voting, protesting, or—in extreme cases—violence (Skitka & Mullen, 2002). Indeed, morality is so strong a motivator that it can sometimes justify any means to a moral end (Fiske & Rai, 2015; Skitka, 2010).

Despite the significant consequences of morality, the use of moral language by political elites is poorly understood. Research has largely focused on ideological differences in moral-language use using the theoretical framework of moral-foundations theory (Haidt & Graham, 2007), which posits five moral domains (foundations) said to constitute the basic building blocks of morality across cultures. The individualizing foundations of *harm* and *fairness* concern individual rights and well-being, whereas the binding foundations of *loyalty, authority*, and *purity* concern adherence to norms that maintain group cohesion. In the United States and elsewhere, both liberal and conservative members of the public rate the individualizing foundations as morally relevant, although liberals endorse them somewhat more strongly. In contrast, the binding foundations are endorsed much more strongly by conservatives than liberals (Graham, Haidt, & Nosek, 2009).

For elite communication, however, there is not a consistent pattern of ideological differences in moral-language use. For example, in U.S. Senate debates on abortion from 1989 to 2006, Republicans used more purity-related language and more fairness-related language than did Democrats (Sagi & Dehghani, 2014). In four months of Twitter posts from 2014, more liberal members of Congress used more moral language related to every foundation except authority, which was referenced more by conservative legislators (Sterling & Jost, 2017). However, in *New York Times* op-ed articles on stem-cell research (published 1999–2010), harm-related language was used more by writers taking the liberal position (i.e., supporting stem-cell research), and purity-related language, though rare overall, was used exclusively by writers taking the conservative position (Clifford & Jerit, 2013). Finally, in political texts (including State of the Union addresses, speeches from the House and Senate floors, and party platforms), associations

Corresponding Author:
Yoel Inbar, Department of Psychology, University of Toronto
E-mail: yoel.inbar@utoronto.ca

between ideology and moral-language use were inconsistent and dependent on the specific dictionaries used to assess moral language (Frimer, 2020).

These mixed results might arise for several reasons. First, dictionary-based word counting was used in many studies to assess moral-language use (this was the case for every study described above except Sagi & Dehghani, 2014). Especially when dictionary words are uncommon, word counting can produce unreliable results that are highly sensitive to the exact words included (Garten et al., 2018). Second, researchers often examined rhetoric around specific issues, which may introduce issue-specific differences (e.g., liberals may use moral language for some issues more often than conservatives, and vice versa). Finally, no research has examined whether moral-language use by elites changes over time, as features of the political environment change. One particularly important feature of the environment is political power—that is, whether a political party controls the executive or legislative branches of government. It may be that as parties gain and lose control of branches of government, their members change the moral language they use. If this is the case, then research examining moral-language use in different time periods would not necessarily be expected to show consistent results.

Political Power and Moral Language

There are several reasons to think that less power might lead to changes in moral language and, specifically, to greater use of such language. First, research on social identity (Tajfel & Turner, 1986) has found that threatened groups (in this case, the minority party) will be particularly motivated to defend their group identity (Branscombe & Wann, 1992). Using moral language is one way to do this: It can bolster a threatened group identity by elevating in-group values to sacred principles and casting out-group members as irredeemably wicked (Fiske & Rai, 2015; Graham & Haidt, 2012). Second, research on integrative complexity (Tetlock, 1986) has found that minority-party members generally show lower integrative complexity in their public statements than do majority-party members (Suedfeld, 2010). Although integrative complexity and morality are distinct, they share some characteristics: Low integrative complexity is characterized by black-and-white thinking and a rejection of compromise, just as morality is (Baron & Spranca, 1997; Tetlock, Kristel, Elson, Green, & Lerner, 2000). Finally, parties have a strategic incentive to use moral language when they need to mobilize their supporters to regain a majority, because moral language is particularly effective at increasing message transmission (Brady, Wills, Burkart, Jost, & Van Bavel, 2019; Brady, Wills, Jost, Tucker, & Van Bavel, 2017) and motivating supporters (Ryan, 2014; Skitka & Bauman, 2008).

The Current Research

In the current research, we examined (a) whether U.S. political elites show ideological differences in moral-language use and (b) whether these differences vary over time—in particular, whether they change as parties gain and lose power. In Study 1, we examined the complete set of Twitter posts made by all accounts belonging to U.S. members of Congress in the 2-year period beginning in January 2016. In Study 2, we examined all speeches made on the floor of the House and Senate in the 97th to 114th Congresses (1981–2017). In each corpus, we assessed how much Democrats and Republicans used language relevant to the domains proposed by moral-foundations theory (harm, fairness, loyalty, authority, and purity) and how this varied with shifts in political power.

Study 1

In Study 1, we analyzed the public tweets posted by accounts belonging to members of the U.S. Congress (i.e., the House of Representatives and the Senate). Nearly every member of Congress has a Twitter account, and many have a large number of followers (users who have subscribed to see another user's tweets). In our data set, for example, the most-followed account (Bernie Sanders) had more than 7 million followers, but even the median account had more than 20,000. Consequently, Twitter is a powerful tool for politicians to rally supporters and communicate with the public, and tweets from elected representatives are a natural way to examine the moral language used by political elites.

Method

Data and code availability. All R scripts are available at https://github.com/yoelinbar/moral-language, and all data are available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FQ8MIL. Because of copyright restrictions, the full tweet text is not included. Full text for any tweet can be retrieved using Twitter's application programming interface (API) and the globally unique identifier (GUID) for the tweet in the data file.

Tweet collection and cleaning. We obtained the official Twitter accounts for members of the 114th and 115th Congresses (from https://gwu-libraries.github.io/sfm-ui/posts/2017-05-23-congress-seed-list) and verified the list with Web searches. Using a set of Python scripts and the

Tweepy library, we used Twitter's API to collect all original tweets (i.e., omitting those marked by Twitter as retweets) posted by these accounts from January 1, 2016, to January 31, 2018, inclusive. We cleaned tweets using a Python script that removed manual retweets (i.e., text following the strings "RT:" or "MT:"), URLs, and punctuation. We removed tweets from eight accounts that had posted fewer than 50 tweets each because we thought they might cause estimation problems in our mixed-effects models. (All these removals are documented in the R script.) Dictionary similarity scores (see below) could not be computed for 2,701 cleaned tweets (these tweets were almost entirely nonwords that had been hashtags, e.g., "VA10," "WhyWeMarch," or "FlyEaglesFly"). Our final sample comprised 687,360 tweets from 578 unique accounts: 385,206 from Democratic members of Congress (n = 261) and 302,154 from Republicans (n = 317).

Moral-language analysis. A major difficulty in examining naturally occurring moral-language use is that it is difficult to measure at scale. Human coders are considered the gold standard of natural-language analysis, but human-coding hundreds of thousands of texts is infeasible. Automated text analysis by word counting (as implemented by, e.g., the Linguistic Inquiry and Word Count [LIWC] software) is an attractive alternative to human coding but is limited in different ways. Most problematically, researchers must specify the entire dictionary of words or word stems that match a concept. If the dictionary coverage is too narrow, important parts of the concept will be omitted, and the text will not match when it should. But if the coverage is too broad, the text will match when it should not. And because word-counting software does not know what words mean, only what they look like, including more terms can easily lead to false positives (e.g., "happ*" matches "happy" and "happiness" but also "happened" and "happenstance"). This problem is exacerbated for short texts, such as tweets (which were until recently limited to 140 characters and are currently limited to 280), in which the large majority of texts might contain no dictionary words at all (Garten et al., 2018). Especially for dictionaries comprising mainly low-frequency words (as is the case for the moral-foundations dictionaries), researchers' subjective judgments about whether a term should be included can have a large influence on results.

In the present research, we avoided these limitations by using a text-analysis technique built on a distributed-language model. Distributed-language models, which derive from a long tradition of research in computational linguistics, cognitive psychology, and computer science (Landauer & Dumais, 1997; Osgood, Suci, & Tannenbaum, 1957; Rogers & McClelland, 2004; Salton, Wong, & Yang, 1975), encode word meaning as a point in a many-dimensional space. Semantically similar words (e.g., *car* and *automobile*) are close to each other in this space. Each word's location in space is described by an n-dimensional vector of real numbers (where n = the number of dimensions in the space), and mathematical operations on these vectors are meaningful. The method we used, called *distributed dictionary representations* (DDR; Garten et al., 2018), combines a small researcher-specified dictionary of terms representing a concept with a distributed-language model. Averaging the normalized vectors for each word in the dictionary produces a single composite representation of the dictionary's meaning. In the same way, the words in a text (in this case, a tweet) can be combined into a composite representation. Finally, the similarity between the text and the dictionary can be computed by cosine similarity (scores range from −1 to 1, with higher values reflecting greater similarity). Importantly, the text need not contain any of the dictionary words for similarity to be computed. For example, a text containing "car" might be considered highly similar to a dictionary consisting of "automobile," "auto," and "vehicle." This means that dictionaries can be short but still cover a concept well, and texts can be short but still be considered similar to a dictionary. This eliminates the two major shortcomings of word-counting approaches. In validation studies comparing automated methods to human coders, this method was substantially more accurate than word counting (Garten et al., 2018).

We used moral-foundations dictionaries developed and validated by Garten et al. (2018) in combination with a publicly available distributed-language representation (https://code.google.com/p/word2vec/) trained on approximately 100 million words from Google News articles (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). We used the *ddr* Python module developed by Garten et al. (2018) to compute the semantic similarity between each tweet and short dictionaries representing positive and negative aspects ("virtue" and "vice") of each of the moral foundations (see Table 1). For each tweet, we therefore had 10 values representing the similarity between the tweet and the positive and negative aspects of each foundation.

Results

Language-model validation. In a validation study looking specifically at moral-language use on Twitter, DDR with these dictionaries showed the highest agreement with human coders of any automated method tested (Garten et al., 2018). In particular, it substantially outperformed word counting using the full moral-foundations dictionary (Graham et al., 2009).[1] We conducted a similar validation comparing a subset of the tweets in our sample

Table 1. Central Concerns for Each of the Five Moral Foundations and the Distributed Dictionary Representations (DDR) Seed Words Used to Measure Positive ("Virtue") and Negative ("Vice") Moral Language Pertaining to Each Foundation

		DDR seed words	
Moral foundation	Concerns	Virtue	Vice
Harm	Kindness, compassion, protection of the helpless or innocent	Kindness, compassion, nurture, empathy	Suffer, cruel, hurt, harm
Fairness	Justice, rights, cooperation, reciprocity	Fairness, equality, justice, rights	Cheat, fraud, unfair, injustice
Loyalty	Patriotism, heroism, fidelity, self-sacrifice	Loyal, solidarity, patriot, fidelity	Betray, treason, disloyal, traitor
Authority	Respect, duty, leadership, magnanimity	Authority, obey, respect, tradition	Subversion, disobey, disrespect, chaos
Purity	Chastity, piety, sacredness, self-restraint	Purity, sanctity, sacred, wholesome	Impurity, depravity, degradation, unnatural

with ratings given by human coders, and we found that DDR with these dictionaries showed substantial agreement with human ratings (mean F_1 score = .865; for full details, see the Supplemental Methods section in the Supplemental Material available online). Table S7 in the Supplemental Material shows, separately for Democrats and Republicans, the tweets in our sample that DDR identified as expressing each foundation and aspect most strongly.

Analytic strategy. We used DDR to compute the similarity of each tweet to each foundation and aspect; these similarity scores were our primary dependent variables. To account for nonindependence of observations (i.e., multiple tweets from the same username), we used mixed-effects models for most of our analyses. Except where otherwise noted, we used the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) with restricted maximum likelihood estimation to fit our models and the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017) to test significance of model coefficients using Satterthwaite-approximated degrees of freedom. These analyses were conducted in R (Version 3.3.2; R Core Team, 2016).

We fitted separate models for each foundation and aspect. Thus, for each analysis, we fitted 10 separate models. To model nonindependence between multiple tweets from the same account, we first added random person intercepts in addition to the fixed effects. Examining residual autocorrelation function (ACF) plots from these models showed that there was also substantial autocorrelation across days (i.e., model errors from day n were correlated with those from day $n + 1$). To account for this autocorrelation, we used the "splines" function in R to model a third-order polynomial effect of time (i.e., days since January 1, 2016) that was allowed to vary randomly by individual (see Verbyla, Cullis, Kenward, & Welham, 1999). We also added random intercepts for each day. We initially attempted to fit these models to the full set of tweets for each foundation and aspect, but no model converged. We therefore simplified the data by computing per-day average similarity scores for each account; that is, for each foundation and aspect, we treated the mean similarity score of all tweets posted on a given day by a given account as one observation. These models converged and showed acceptably low levels of autocorrelation in the residuals.

Our key tests were implemented by dummy variables for party, time period, or their interaction. Because multilevel models are designed to accommodate varying numbers of observations between groups (see Snijders & Bosker, 2012, p. 56), tests of these dummy variables are unbiased despite the fact that we have different numbers of observations between different individuals, parties, or time periods.

Partisan differences in moral-language use. For these analyses, we created a dummy variable for party (1 = Democrat, 0 = Republican); statistical tests are of the dummy coefficient. Table S1 in the Supplemental Material shows tests of the effects of party on moral-language use for each moral foundation and aspect. For every foundation and aspect, Democrats used moral language more than Republicans, all ps < .001 (see Fig. 1). To measure effect size, we standardized moral-language scores so that the dummy coefficients can be interpreted as the difference in scores between Democrats and Republicans, expressed in standard deviations (Lorah, 2018). Coefficients (bs) ranged from 0.19 (loyalty-virtue) to 0.39 (fairness-vice). Table S2 and Figure S1 in the Supplemental Material show that this

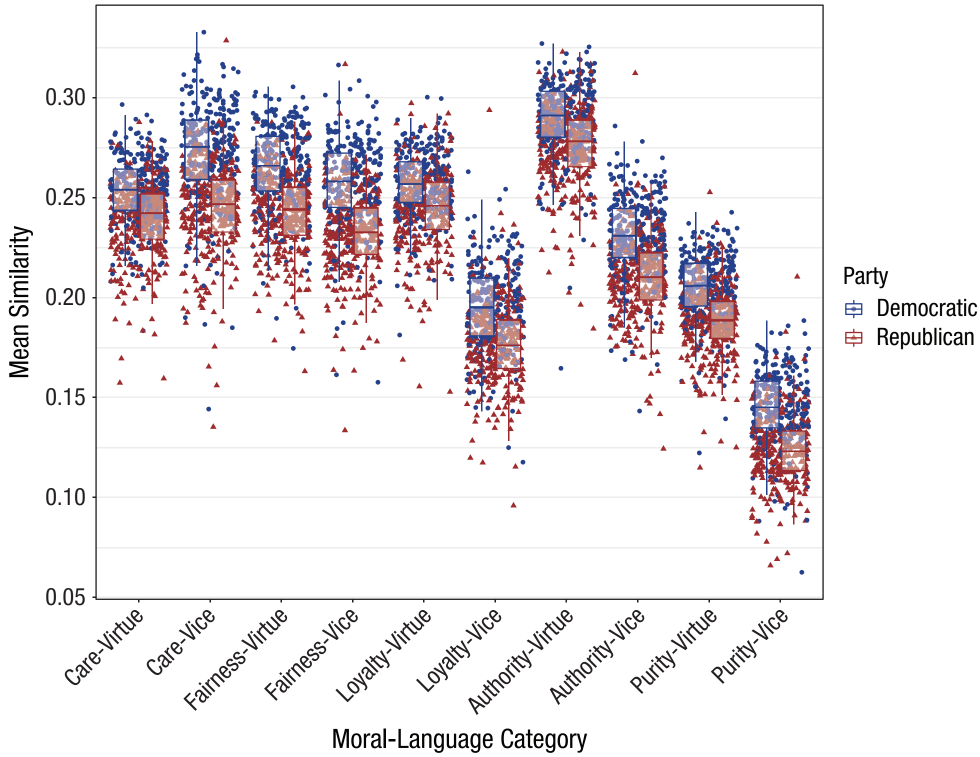

Fig. 1. Mean similarity score in Study 1 for each moral-language category for tweets by Republicans (red triangles) and Democrats (blue circles). Within a moral-language category, each point represents the average of all tweets from one member of Congress. Points are jittered for legibility. Overlaid boxes show the 25th (bottom line), 50th (middle line), and 75th (top line) percentiles. Whiskers extend to 1.5 times the interquartile range.

pattern is consistent when predicting moral-language use from each individual's dynamic, weighted nominal three-step estimation (DW-NOMINATE) score, a continuous measure of ideology based on Congressional voting records (Poole & Rosenthal, 1997).

Changes in political power. Republicans were the majority party in the House and Senate from 2016 to 2018. However, the presidency changed from Democratic to Republican control following the 2016 presidential election. We therefore tested the association between political power and moral-language use by adding a dummy variable to our mixed-effects models that coded whether each tweet was posted before or after the 2016 presidential election. Moral-language use by both Democrats and Republicans was higher in the postelection time period (i.e., after November 8, 2016). However, this difference between time periods was larger for Democrats than Republicans (ps < .001) for every foundation and aspect except harm-virtue (where it was equally large for both parties). For Democrats, coefficients (bs) for pre- versus postelection differences ranged from 0.17 (harm-virtue) to 0.45 (harm-vice and fairness-vice), all ps < .001. In contrast, for Republicans, the largest coefficient increase was 0.22 (fairness-virtue), and in several cases pre- versus postelection differences were very small (bs < 0.10), though still statistically significant. Table S3 in the Supplemental Material shows statistical tests from mixed-effects models of the differences in moral language between time periods and their interactions with party.

We next tested whether moral-language use showed a discontinuity specifically on election day (November 8, 2016) with an interrupted time-series analysis (Wagner, Soumerai, Zhang, Ross-Degnan, 2002). These analyses, which are reported in detail in the Supplemental Material and depicted in Figure 2, show positive and significant discontinuities for Democrats for every foundation except harm-virtue. The pattern for Republicans was much more mixed. For some foundations and aspects, moral-language use increased, but for others it did not change significantly or even decreased.

Topic models. Finally, we conducted an exploratory analysis of whether different topics of discussion were associated with the use of particular kinds of moral language and with political ideology. We did this using a structural topic model (Roberts et al., 2014), a variation of the popular latent Dirichlet allocation (Blei, Ng, & Jordan, 2003) that allows for the inclusion of additional document-level metadata that may alter either the prevalence or the content

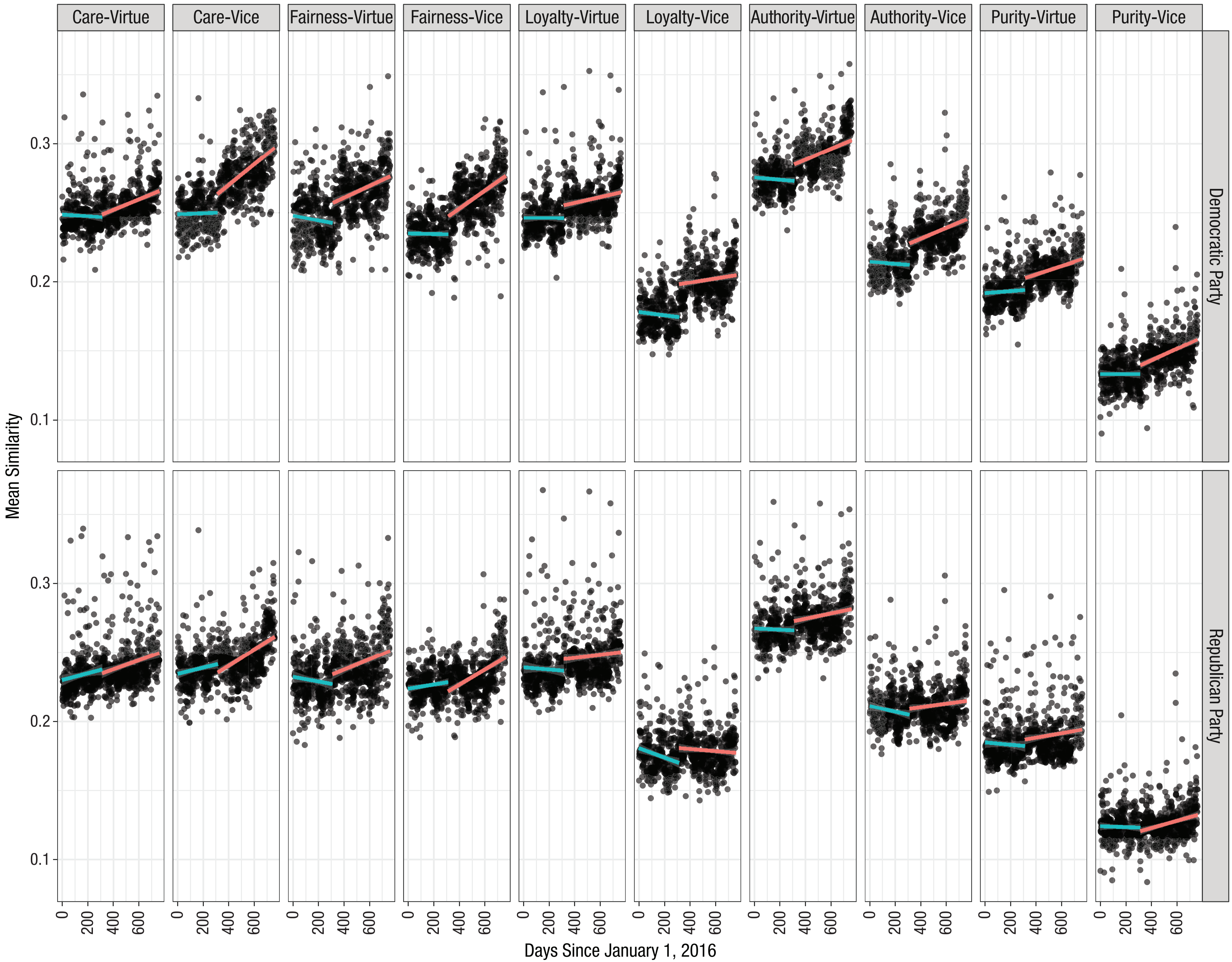

Fig. 2. Mean similarity score in Study 1 for each moral-language category before and after the election, separately for Democrats and Republicans. Each point represents the mean similarity for tweets from 1 day. Green and orange lines are separate linear fits of the time trend in scores before and after the election, respectively. The vertical gap between lines shows the discontinuity at November 8, 2016.

of topics. We fitted our structural topic model using the *stm* package in R (Roberts, Stewart, & Tingley, 2019). Our final model had 77 topics. We included political affiliation and a dummy-coded variable of whether the tweet was made before or after the 2016 presidential election as topical prevalence covariates, along with the tweet's similarity score for each of the 10 moral foundations and aspects. These covariates and their interactions were significant for almost all topics, and an inspection of the top words and tweets for each topic revealed that the topics were interpretable and reflected meaningful facets of political discussion (see Table S8 in the Supplemental Results). This suggests that, as expected, there are differences in the topics for which moral language is used, the topics that are discussed by Democrats and Republicans, and the topics that were discussed before and after the election. These differences are plotted in Figures S4 through S15 in the Supplemental Material. Figure S3 in the Supplemental Material shows the semantic coherence (a measure of topic quality) of these topics.

Discussion

Posts from Democratic lawmakers' accounts consistently used more moral language relevant to all moral foundations. Partisan differences increased after the election of Donald Trump because of a substantial increase in moral-language use by Democrats in the postelection period. Thus, ideological differences in moral-language use were inconsistent with moral-foundations theory (which predicts that authority, loyalty, and purity should be referenced more by Republicans than Democrats), and they varied over time. In Study 2, we examined the relationship between political power and moral-language use in a different corpus over a substantially longer time period.

Study 2

In Study 2, we analyzed the U.S. Congressional Record, which contains transcripts of the debates and proceedings of the U.S. Congress. Congressional transcripts are commonly used to study the speech of legislators (e.g., Gentzkow, Shapiro, & Taddy, 2019; Thomas, Pang, & Lee, 2006). Unlike tweets, congressional transcripts are available over a much longer period of time. This allowed us to test the effects of shifting political power on moral-language use as the parties gained and lost control of the presidency, House, and Senate.

Method

Data and code availability. We used a data set of the U.S. Congressional Record that parses the text by speech and includes metadata on speeches and their speakers (Gentzkow et al., 2019). We restricted our analyses to the 97th Congress (beginning January 3, 1981) onward, because our hypotheses concerned moral-language use by the minority party. Between the 72nd Congress (beginning March 4, 1931) and the 97th, Democrats had nearly unbroken control over both the House and Senate. Thus, for these years our analyses would have confounded minority status and party. Additionally, before the civil rights era in the 1960s, the ideological composition of the two parties was different enough that analyses of partisan differences would not be particularly meaningful across longer periods of time (Brady & Stewart, 1982).

The data set we used ends with the 114th Congress (ending January 3, 2017), so all our analyses were of the 97th to 114th Congresses (i.e., January 3, 1981–January 3, 2017). The full-text Congressional speeches data set is available at https://data.stanford.edu/congress_text. R code for all analyses is available at https://github.com/yoelinbar/moral-language.

Moral-language analysis. As in Study 1, we used DDR (Garten et al., 2018) to compute the semantic similarity between each speech and the positive and negative aspects of each moral foundation. We again found good agreement (mean F_1 score = .791) between human coders and the DDR scores (see the Supplemental Material for details).

Results

Analytic strategy. Our analytic strategy was similar to that of Study 1. We again used mixed-effects models to account for nonindependence of observations. These models included random intercepts for speakers and for Congresses (in the modern era, a single Congress lasts 2 years and consists of two sessions in consecutive years). We again fitted separate models for each foundation and aspect, resulting in 10 models for each analysis. Initial models failed to converge, so we computed the per-Congress average similarity scores for each speaker (in other words, we treated all speeches given by an individual in a single Congress as a single observation). After this change, every model converged and showed acceptably low levels of autocorrelation in the residuals.

Partisan differences in moral-language use. As in Study 1, we created a dummy variable for party (1 = Democrat, 0 = Republican). Table S9 in the Supplemental Material shows tests from mixed-effect models of the effect of party affiliation on moral-language use for each moral foundation and aspect. As in Study 1, when there were partisan differences in moral-language use, Democrats tended to use more of it. However, differences were much less consistent than in Study 1. Overall, Democrats used significantly more moral language in two categories, harm-virtue (b = 0.15) and fairness-virtue (b = 0.17), and there were no significant effects of political party in other categories. This suggests that the strong partisan differences we observed in Study 1 were specific to that political environment.

Effects of political power. To test the association between power and moral-language use, we created dummy variables representing whether or not a speaker's political party held control of their chamber of Congress (1 = yes, 0 = no). We ran mixed models assessing the effects of affiliation and political control separately for the House and Senate. Results of these mixed models are shown in Tables S10 and S11 in the Supplemental Material. We found largely consistent patterns for the effects of political control. In the Senate, the minority party used more moral language for all foundations and aspects except fairness-virtue and authority-vice (where there were no significant effects) and authority-virtue (which was used significantly more by the majority party). In the House, the minority party used more moral language for all foundations and aspects except purity-vice (where there were no significant effects), fairness-virtue (which was used significantly more by the majority party), and authority-virtue (which was used significantly more by the majority party). These results are shown in Figures 3 and 4. To determine whether party affiliation moderated the effects of political control, we compared models that contained only main effects of political control and party with models that also included an interaction term between the two. Likelihood-ratio tests comparing each set of models found no significant improvements in model fit from the inclusion of an interaction term for any model, indicating that the effects of political power on moral-language use was not moderated by party.

We also tested for cross-chamber effects—that is, whether party control of the other chamber of Congress affected a speaker's moral-language use, controlling for minority status in the speaker's own chamber. We found a significant cross-chamber effect for authority-virtue in the Senate, and no other significant effects in the Senate. Cross-chamber effects were significant for five moral-language categories in the House, but effect sizes were small (ranging from b = 0.055 to b = 0.083). Likewise, we tested whether the speaker's party controlled the presidency and found significant effects for four categories: The minority party used more moral language related to fairness-virtue, fairness-vice, and authority-virtue, and the majority party used more moral language related to loyalty-virtue, again with small effect sizes (b = 0.048 to b = 0.072). Full results of these models are in the Supplemental Material.

Overall, then, relationships between party control and moral-language use were most consistent and strongest for control of the speaker's own chamber of Congress: When speakers were in the minority in their chamber, they tended to use moral language more.

General Discussion

We examined ideological differences in moral-language use and whether these differences vary over time as political power shifts. In Study 1, in which we examined all Twitter posts from members of Congress from January 2016 to 2018, we found that Democrats used more moral language than Republicans across all moral foundations, including binding foundations, and that these differences increased after the election of Donald Trump. In Study 2, in which we examined 36 years of Congressional transcripts, we found that Democrats used two kinds of moral language more overall (harm-virtue and fairness-virtue) but also that partisan differences in the use of most moral language fluctuated over time because minority-party legislators in a chamber used more moral language of most kinds.

Theoretical implications

Assessing the moral concerns of laypeople via self-report consistently shows that conservative respondents see the binding moral foundations (loyalty, authority and purity) as much more morally relevant than do liberal respondents (Graham et al., 2009). On this basis, one would expect that conservative elites would reference the binding foundations more than liberal elites do. Some investigations of elite moral-language use have found this pattern, at least in part (e.g., Clifford & Jerit, 2013; Graham et al., 2009, Study 4), but others have not (e.g., Frimer, 2020; Sagi & Dehghani, 2014; Sterling & Jost, 2017).

Together with this previous research, our findings suggest that self-report ratings of relevance cannot be straightforwardly extrapolated to the moral language used by elites. However, our results suggest some promising possibilities for theoretical refinement. We found that moral-language use changes over time as parties' political fortunes rise and fall, so it is likely too simplistic to treat observed ideological differences in morality as static. For a better understanding of ideological differences in moral language, theorists need to acknowledge that any differences may shift over time and that differences at a specific time may not be generalizable to other time periods. Ideally, theorists should model the effects of specific influences as they change over time, as we did here for political power. For example, liberals and conservatives might use different moral

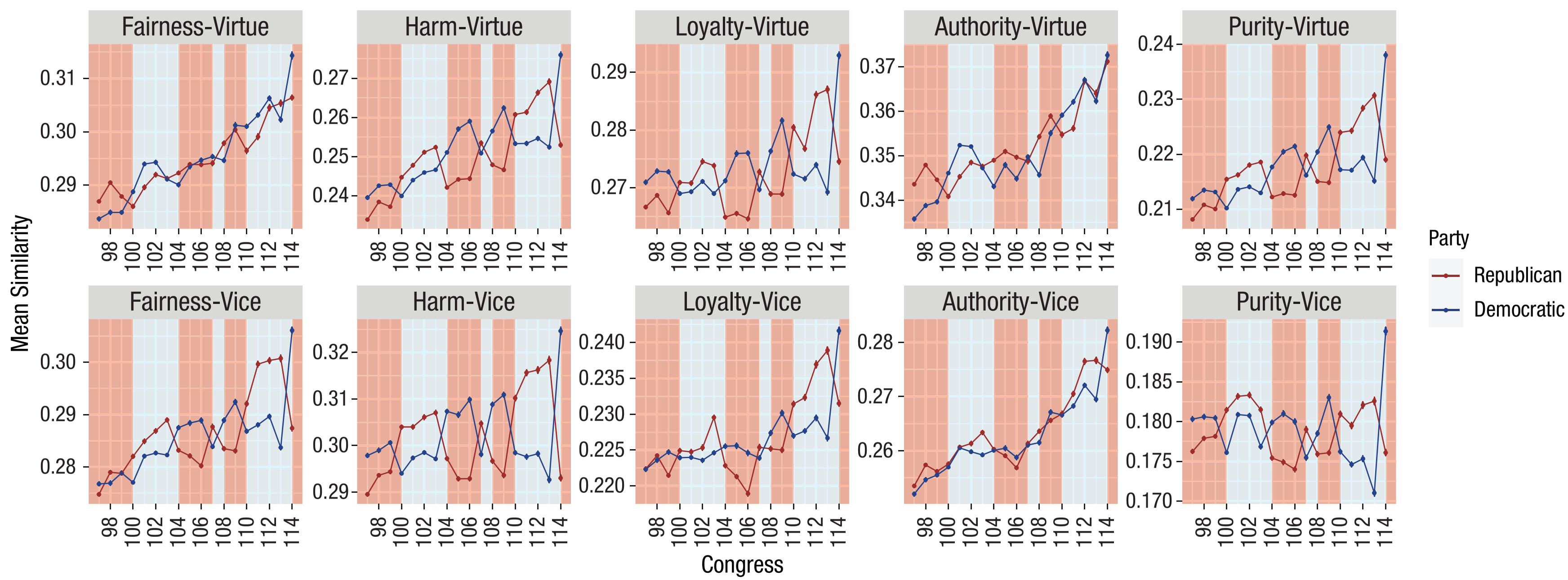

Fig. 3. Mean loading in Study 2 for each moral-language category, separately for Democratic and Republican Senators in the 97th Congress through the 114th Congress. Shading marks majority control of the Senate.

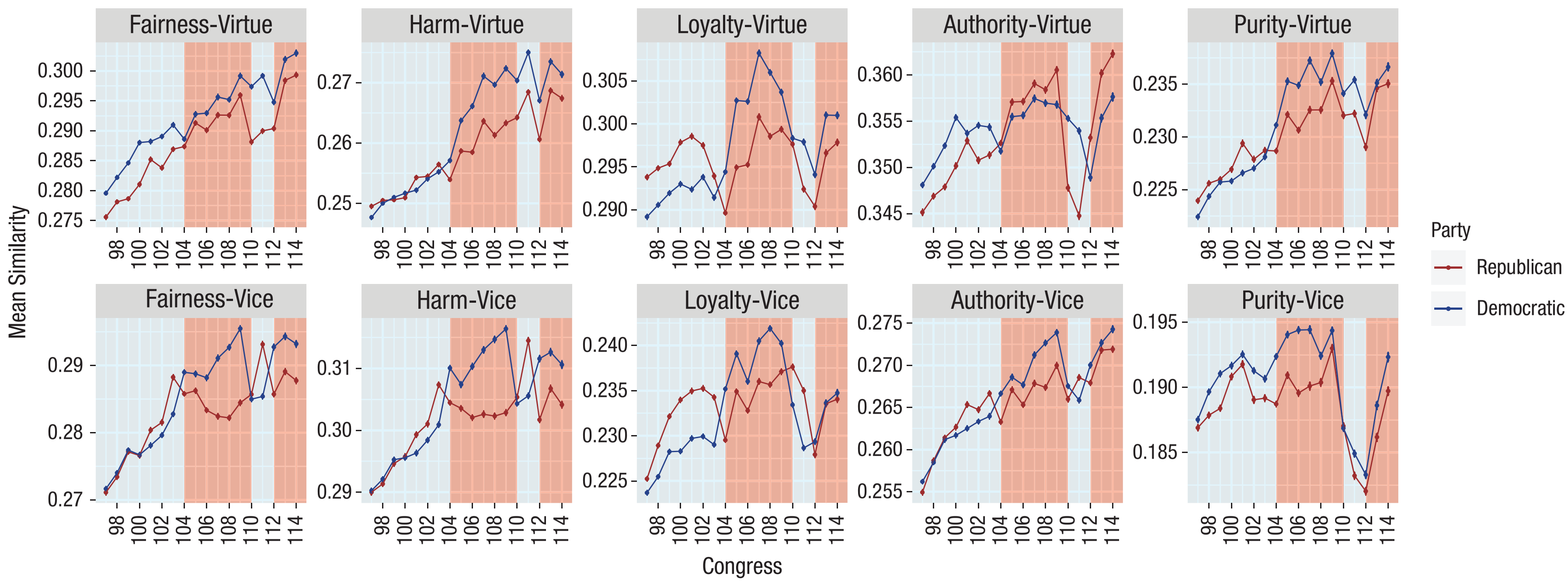

Fig. 4. Mean loading in Study 2 for each moral-language category, separately for Democratic and Republican Representatives in the 97th Congress through the 114th Congress. Shading marks majority control of the House.

language in response to domestic crises, conflicts with foreign adversaries, or power struggles among different branches of government. Of course, it might also be that liberals and conservatives use moral language in response to these events similarly (as we found to be the case for political power). In either case, however, our theoretical understanding will be enhanced by treating moral-language use as dynamic rather than static.

Open questions

In Study 2, we observed the strongest and most consistent power effects for party control of legislators' own chambers. Effects for control of the presidency were much smaller and less consistent. This contrasts with Study 1, in which we observed significant increases in the moral language used by Democrats across all categories after the 2016 election. This suggests that there may be something unique about the Trump presidency—perhaps the fact that Trump, himself, is especially polarizing (Dugan, 2018) or the fact that Clinton was widely expected to win—that increases moral-language use beyond the level one would expect in a more normal political environment. It is also possible that the medium of communication—Twitter posts, rather than Congressional debates and proceedings—may be responsible for some of these differences. Both of these questions are promising ones for future research.

Although minority-party members generally used more moral language across most categories, this was not true for all types of moral language in both chambers. Although it was uncommon for the majority party to use more moral language, one exception in both the House and Senate was the authority-virtue category. This category includes concepts such as respect for authority figures and obedience; its greater use by the majority party therefore seems reasonable. However, in future research, topic modeling or similar approaches should be employed to better understand what is being discussed when moral language is used and how this changes over time.

Future directions

For our Twitter data set, we conducted an exploratory analysis (reported in the results of Study 1) using structural topic models to determine whether different topics of discussion were associated with particular kinds of moral language. We found that different topics were indeed associated with different kinds of moral language for Democrats and Republicans (see also Sterling & Jost, 2017). Moreover, Democratic and Republican lawmakers differed in how likely they were to tweet about specific topics, and topic prevalence changed over time (see Figs. S2–S14 in the Supplemental Material). These preliminary results suggest that tracing the rise and fall in prevalence of particular topics over time, and the different moral language these topics evoke from liberals and conservatives, is a promising avenue for understanding ideological differences in morality.

Transparency

Action Editor: Leaf Van Boven
Editor: D. Stephen Lindsay
Author Contributions
Both authors jointly designed and performed the research, analyzed the data, and wrote the manuscript. Both authors approved the final manuscript for submission.
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This work was supported by Social Sciences and Humanities Research Council of Canada Grant No. 430-2017-00836.
Open Practices
All data have been made publicly available via Harvard Dataverse and can be accessed at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FQ8MIL. R code for all analyses is available at https://github.com/yoelinbar/moral-language. The design and analysis plans for the studies were not preregistered. This article has received the badge for Open Data. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iD

Yoel Inbar https://orcid.org/0000-0002-3176-3727

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620960397

Note

1. In terms of F_1, a commonly used measure of classification accuracy that takes into account both false positives and false negatives, DDR achieved a mean F_1 of .496, whereas word counting achieved a mean F_1 of .275 (Garten et al., 2018).

References

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, *70*, 1–16.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). doi:10.18637/jss.v067.i01

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Brady, D., & Stewart, J. (1982). Congressional party realignment and transformations of public policy in three realignment eras. *American Journal of Political Science*, *26*, 333–360.

Brady, W. J., Wills, J. A., Burkart, D., Jost, J. T., & Van Bavel, J. J. (2019). An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General*, *148*, 1802–1813.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences, USA*, *114*, 7313–7318.

Branscombe, N. R., & Wann, D. L. (1992). Physiological arousal and reactions to outgroup members during competitions that implicate an important social identity. *Aggressive Behavior*, *18*, 85–93.

Clifford, S., & Jerit, J. (2013). How words do the work of politics: Moral foundations theory and the debate over stem cell research. *The Journal of Politics*, *75*, 659–671.

Dugan, A. (2018). Party, ideology and race play key roles in Trump approval. *Gallup*. Retrieved from http://news.gallup.com/poll/226091/party-ideology-race-play-key-roles-trump-approval.aspx

Fiske, A. P., & Rai, T. S. (2015). *Virtuous violence*. Cambridge, England: Cambridge University Press.

Frimer, J. A. (2020). Do liberals and conservatives use different moral language? Two replications and six extensions of Graham, Haidt, and Nosek's (2009) moral text analysis. *Journal of Research in Personality*, *84*, Article 103906. doi:10.1016/j.jrp.2019.103906

Garten, J., Hoover, J., Johnson, K. M., Boghrati, R., Iskiwitch, C., & Dehghani, M. (2018). Dictionaries and distributions: Combining expert knowledge and large scale textual data content analysis. *Behavior Research Methods*, *50*, 344–361.

Gentzkow, M., Shapiro, J. M., & Taddy, M. (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, *87*, 1307–1340.

Graham, J., & Haidt, J. (2012). Sacred values and evil adversaries: A moral foundations approach. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (p. 11–31). Washington, D.C.: American Psychological Association.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*, 1029–1046.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*, 98–116.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). doi:10.18637/jss.v082.i13

Lakoff, G. (2004). *Don't think of an elephant! Know your values and frame the debate*. Hartford, VT: Chelsea Green Publishing.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education*, *6*, Article 8. doi:10.1186/s40536-018-0061-2

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (*Vol. 26*). Retrieved from https://papers.nips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Champaign, IL: University of Illinois Press.

Poole, K. T., & Rosenthal, H. (1997). *Congress: A political-economic history of roll call voting*. Oxford, England: Oxford University Press.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R package for structural topic models. *Journal of Statistical Software*, *91*(2). doi:10.18637/jss.v091.i02

Roberts, M., Stewart, B., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., . . . Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*, 1064–1082.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Ryan, T. J. (2014). Reconsidering moral issues in politics. *The Journal of Politics*, *76*, 380–397.

Sagi, E., & Dehghani, M. (2014). Measuring moral rhetoric in text. *Social Science Computer Review*, *32*, 132–144.

Salton, G., Wong, A., & Yang, C.-S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, *18*, 613–620.

Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass*, *4*, 267–281.

Skitka, L. J., & Bauman, C. W. (2008). Moral conviction and political engagement. *Political Psychology*, *29*, 29–54.

Skitka, L. J., & Mullen, E. (2002). Understanding judgments of fairness in a real-world political context: A test of the value protection model of justice reasoning. *Personality and Social Psychology Bulletin*, *28*, 1419–1429.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, England: SAGE.

Sterling, J., & Jost, J. T. (2017). Moral discourse in the Twitterverse: Effects of ideology and political sophistication on language use among U.S. citizens and members of Congress. *Journal of Language and Politics*, *17*, 195–221.

Suedfeld, P. (2010). The cognitive processing of politics and politicians: Archival studies of conceptual and integrative complexity. *Journal of Personality*, *78*, 1669–1702.

Tajfel, H., & Turner, J. C. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago, IL: Nelson-Hall.

Tetlock, P. E. (1986). A value pluralism model of ideological reasoning. *Journal of Personality and Social Psychology*, *50*, 819–827.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*, 853–870.

Thomas, M., Pang, B., & Lee, L. (2006). Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In D. Jurafsky & E. Gaussier (Eds.), *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* (pp. 327–335). Sydney, Australia: Association for Computational Linguistics

Verbyla, A. P., Cullis, B. R., Kenward, M. G., & Welham, S. J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *48*, 269–311.

Wagner, A. K., Soumerai, S. B., Zhang, F., & Ross-Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics*, *27*, 299–309.