

Partisan Conflict and Congressional Outreach

Partisan criticism generates most engagement in social media; most liberal and conservative legislators, party leadership more likely to 'go negative'

FOR MEDIA OR OTHER INQUIRIES:

Solomon Messing, Director of Data Labs
Rachel Weisel, Communications Manager

202.419.4372

www.pewresearch.org

About Pew Research Center

Pew Research Center is a nonpartisan fact tank that informs the public about the issues, attitudes and trends shaping America and the world. It does not take policy positions. The Center conducts public opinion polling, demographic research, content analysis and other data-driven social science research. It studies U.S. politics and policy; journalism and media; internet, science and technology; religion and public life; Hispanic trends; global attitudes and trends; and U.S. social and demographic trends. All of the Center's reports are available at www.pewresearch.org. Pew Research Center is a subsidiary of The Pew Charitable Trusts, its primary funder.

© Pew Research Center 2017

Table of Contents

About Pew Research Center	2
Methodological note	4
Terminology	5
Overview	7
1. Who is going negative? Leadership and very liberal or conservative members most critical of other side	14
2. How the public reacted on Facebook	29
3. Partisan language in congressional outreach	32
Acknowledgments	39
Methodology	40
Data collection	40
Content analysis	50

Methodological note

This report describes key findings from an analysis of all official press releases and Facebook posts identifiable via internet and archival sources, issued by members of the 114th Congress between Jan. 1, 2015, and April 30, 2016, prior to either party's 2016 presidential nominating convention. A total of 94,521 press releases were collected directly from members' websites and supplemented with additional press releases collected from LexisNexis searches. The 108,235 Facebook posts from members' official accounts were collected directly from the social media platform's application programming interface (API) for public pages.

To analyze these two collections of congressional statements, Pew Research Center used a combination of machine-learning techniques and traditional content analysis methods. The Center sampled 7,000 press releases and Facebook posts. Researchers and paid content coders then manually classified the sample to create a set of training data. Next, the Center employed an automated approach that approximates human judgments to classify the rest of the documents. Researchers used machine-learning models that estimated the relationship between words used in the documents and human classification decisions to predict the most likely classification for all documents that were not coded by hand. Additional detail can be found in the methodology section.

This approach allowed the Center to provide a more comprehensive account of congressional communications than would be possible with a much smaller sample of documents classified by hand. For example, researchers were able to estimate the proportion of communications containing expressions of indignant disagreement *for every U.S. senator and representative* during this period, allowing for an examination of the relationship between indignation and the political attributes of particular members and their districts.

Nonetheless, this process, which was relatively new for the Center, necessarily entails some degree of error in concept operationalization, human judgment, machine classification and statistical estimation. Because of this, the statistics included in this report should be understood as estimates. Just as survey results need to be understood in the context of a margin of error, there is a degree of error associated with the results presented here. Estimates of error are provided in the form of standard error intervals and confidence bands, which are attempts to quantify the uncertainty surrounding each set of statistics. Estimates of reliability in human and error in machine classification are provided in the methodology section at the end of the report.

Terminology

“Statement” refers to a single press release or Facebook post issued by a member of Congress. In this report, the term “statement” does *not* refer to a single sentence or assertion.

“Disagreement” refers to statements that explicitly oppose or disagree with any of the following domestic political actors: President Barack Obama or his administration; Democrats or liberals; or Republicans or conservatives. Such statements do *not* include critiques limited to policy, but rather must identify other political actors directly. Thus, criticism of the Affordable Care Act (ACA or “Obamacare”) is not classified as disagreement unless it explicitly blames President Obama or his administration for some negative outcome or associates the policy directly with political opponents.

“Indignant Disagreement” refers to statements that go beyond disagreement as defined above by also expressing emotions such as anger, resentment, or annoyance. All statements that contain indignant disagreement also contain disagreement more broadly as defined above.

“Criticism” refers to statements containing any form of disagreement (including indignant disagreement).

“Constituent Benefits” refers to favorable legislative outcomes and/or government spending that directly benefit an elected official’s home state or district. Benefits may accrue to residents of other districts as well.

“Bipartisan” refers to messages that suggest that Democrats and Republicans are – or should be – working together in good faith to achieve a shared goal.

“Conservative,” “liberal,” and **“moderate”** are defined using a measure called DW-NOMINATE, which places members of the U.S. House and Senate on a liberal-to-conservative ideology scale according to their roll-call voting history in each legislative session of Congress. The scale ranges from -1 (very liberal) to 1 (very conservative). A score closer to the extremes of the measure indicates a more consistently liberal or conservative voting record in that session. For example, a Republican who never votes with Democrats would receive a score closer to 1 than a Republican who votes in favor of bipartisan legislation.¹ When a member is referred to as “conservative” or “liberal,” these labels are based on their DW-NOMINATE score. The term

¹ See Poole, Keith T., and Howard Rosenthal. 1985. “A Spatial Model for Legislative Roll Call Analysis.” *American Journal of Political Science*.

“moderate” refers to members whose scores are near the middle of the scale, and “very liberal” and “very conservative” refers to members with scores closer to either end of the scale.

“District competitiveness” refers to the chance that a candidate from one party could beat a candidate from the other party at the ballot box. That chance is lower in states and districts where one party tends to win elections by a large margin. This report measures voters’ preferences and district competitiveness using 2012 presidential election results. Less competitive districts are defined as those where either President Obama or Mitt Romney received a large majority of the votes; more competitive districts are defined as those in which the presidential vote was closer to 50-50.

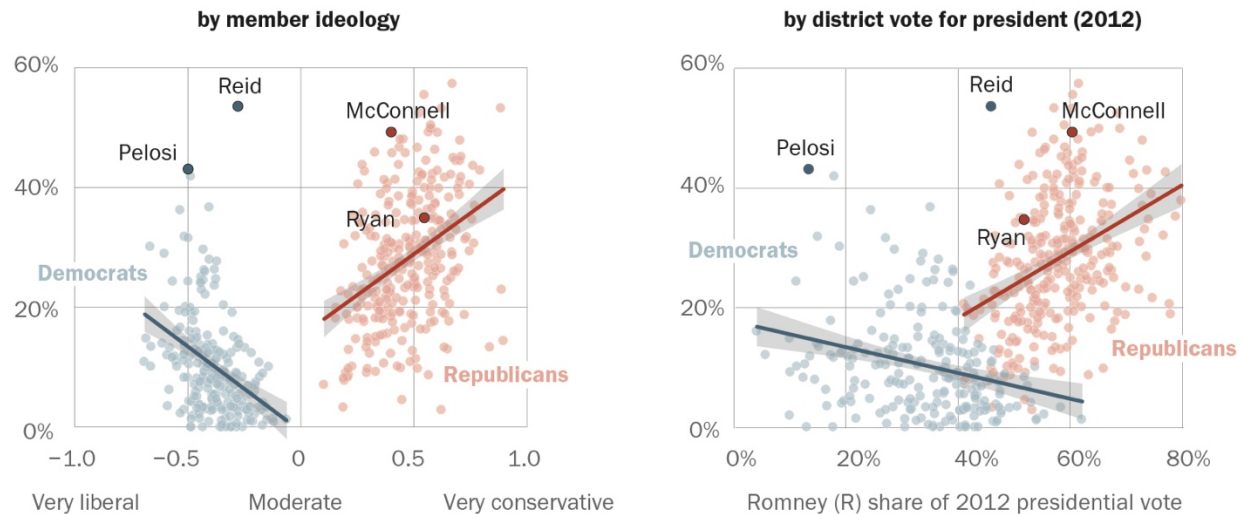
“Engagement” refers to the attention a Facebook post receives, as measured by likes, comments, and shares. Each of these metrics can provide a sense of how much a member’s audience on Facebook reacts to particular statements that they make.

Overview

Hostility in political discourse was thrown into sharp relief in 2016 by a contentious presidential campaign. A new Pew Research Center analysis of more than 200,000 press releases and Facebook posts from the official accounts of members of the 114th Congress uses methods from the emerging field of computational social science to quantify how often legislators themselves “go negative” in their outreach to the public. Overall, the study finds that the most aggressive forms of disagreement are relatively rare in these channels. For the average member, 10% of press releases and 9% of Facebook posts expressed disagreement with the other party in a way that conveyed anger, resentment or annoyance. But there are distinct patterns among those who voiced political discord: congressional leaders, those with more partisan voting records and those who are elected in districts that are solidly Republican or Democratic were the most likely to go negative. Republicans, who did not control the presidency during the 114th Congress, were much more likely to voice disagreement than were Democrats. The analysis represents the Center’s most extensive use of the tools and methods of data science to date.

Most liberal and conservative legislators, party leaders, legislators in safer seats more likely to express disagreement

% of press releases expressing disagreement with other party



Note: Each point in the figure represents a member of Congress. Party leaders of the 114th Congress are shown as a point of reference. The leaders identified include Nancy Pelosi, Harry Reid, Mitch McConnell, and Paul Ryan. Estimated ideology based on roll call votes (DW-NOMINATE). The fitted line shows the relationship between each legislator's proportion of expressed disagreement and their ideology (left) or district-level Romney vote share (right), for each party (estimated via ordinary least squares regression). The shaded region is that estimate's 95 percent confidence band, an attempt to quantify the range of uncertainty in this relationship.

Source: Pew Research Center analysis of data from congressional websites, Lexis-Nexis, Voteview.org, and Jacobson, Gary (2015). See Methodology section for details.

"Partisan Conflict and Congressional Outreach"

PEW RESEARCH CENTER

The study also finds that critical posts on Facebook get more likes, comments, and shares.² Posts that contained "indignant disagreement" – defined here as a statement of opposition that conveys annoyance, resentment or anger – averaged 206 more likes,³ 66 more shares and 54 more comments than those that contained no disagreement at all. Other research suggests that, faced with divisive policy rhetoric, audiences tend to adopt the stances of their party leaders.⁴

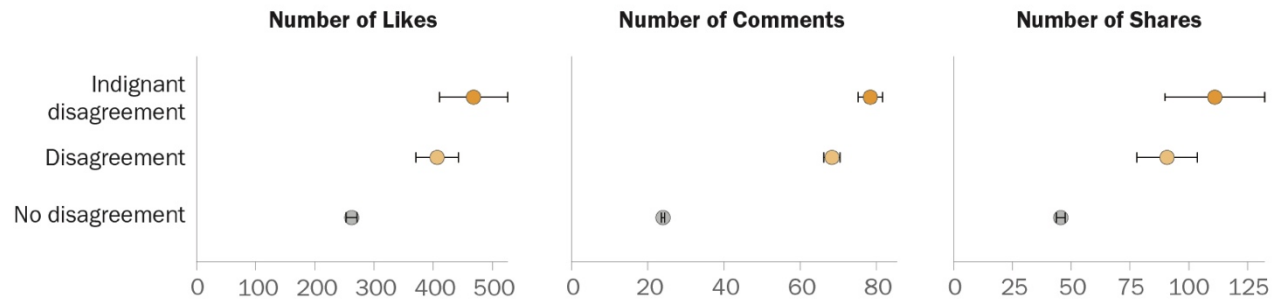
² This finding is robust to member-level statistical controls. See Chapter 2 for details.

³ This study reports on "likes" only; other "reactions" such as "wow," "sad," and "angry" are not analyzed here.

⁴ See Gabriel Lenz, *Follow the Leader: How Voters Respond to Politicians' Policies and Performance*, University of Chicago Press.

Critical posts get more likes, comments, and shares than other posts

Average number of likes, comments, and shares per Facebook post containing ...



Note: Lines indicate the standard error, an attempt to quantify the uncertainty surrounding each estimate. The “disagreement” and “indignant disagreement” categories are not mutually exclusive: statements that contain indignant disagreement are a subset of those that contain disagreement more broadly.

Source: Pew Research Center analysis of data from Facebook OpenGraph API. See Methodology section for details. “Partisan Conflict and Congressional Outreach”

PEW RESEARCH CENTER

On the other hand, members also discussed legislative bipartisanship – emphasizing the importance of working across the aisle – in 21% of press releases, more than three times as often as on Facebook, where just 6% of posts contained bipartisan content, on average. The most moderate legislators and those in districts with the most partisan competition were the most likely to emphasize bipartisanship. During the period under study here, Facebook posts focused on bipartisanship averaged substantially less engagement than those containing disagreement.

The findings come at a time when many Americans are struggling to make sense of the extreme partisanship and [polarization on display in Washington](#). Some scholarly evidence suggests that in recent years, a decades-long trend toward increasing partisanship in Congress may be influencing the growth of similar divisions among the American public, though it’s difficult to nail down exactly why the public has become more polarized.⁵

As of 2016, a Pew Research Center study found that [more than half of Americans who identify as Republicans or Democrats](#) had a “very unfavorable” view of the opposing party, up from around 20% in the early 1990s. And a [2014 report](#) found that “Republicans and Democrats are more

⁵ Hetherington, Marc J. “Resurgent Mass Partisanship: The Role of Elite Polarization,” *American Political Science Review*; Nicholson, Stephen P. “Polarizing Cues,” *American Journal of Political Science*.

divided along ideological lines – and partisan antipathy is deeper and more extensive – than at any point in the last two decades.”

At the same time, the number of moderates in Congress has [dwindled since the 1970s](#). Yet, in 2014, [56% of Americans said that they preferred leaders who compromise](#) over those who stick to their positions. In 2015, [57% reported feeling frustrated](#) with the federal government. The result is a paradox in American politics: Voters generally say they want a functioning government with legislators willing to compromise, but [polarization in Congress](#) – and partisan antipathy among members of the public – continues to rise.

The results presented here come from an analysis of more than a year’s worth of press releases and Facebook posts by members of Congress. While these channels don’t cover all the ways representatives communicate with their constituents – there are also town halls, media appearances, other social media outlets and plenty of discussion on the floors of the House and Senate – they represent two key channels for officials that can be captured and studied in their entirety.⁶ Studying Facebook posts also makes it possible to measure how much a member’s audience interacts with the posts by looking at likes, comments and shares.

To generate these findings, Pew Research Center used a combination of human coders and machine-learning methods to classify 94,521 press releases and 108,235 Facebook posts issued by members of Congress based on their substantive content: the events, topics and issues raised and discussed in each kind of document. The documents were classified as expressing disagreement; “indignant disagreement,” a type of disagreement that also expresses anger, resentment or annoyance; support for bipartisanship; or descriptions of constituent benefits.

Using machine-learning methods to sort through large amounts of written material has a key advantage: once trained, a computer can look through much more data far more quickly than a team of humans. This type of analysis – still relatively new to the Center – is in its infancy in terms of potential uses for answering key questions in social science, and is full of promise.

At the same time, using computers to code written language is inherently an error-filled process. Because of the complexity of the assigned task (attempting to identify nuanced themes such as “bipartisanship” and “indignation”) and the inevitable possibility of human error associated with the training process, a machine-learning system will miss instances it should have caught, and incorrectly classify some others. Just as survey results need to be understood in the context of a

⁶ Due to the way Twitter makes its data available, a full census of Twitter posts issued by members of Congress is costly to collect. Data from Twitter are not examined in this initial report.

margin of sampling error, there is a degree of error associated with the results presented here. In the cases where this is quantifiable, it is laid out in the methodology at the end of the report.

The period under study in this report spans the beginning of the 114th Congressional session in January 2015 through April 30, 2016, prior to either party's presidential nominating convention. It also incorporates information about legislators and legislative districts compiled from outside sources. In particular, this report uses a measure called "[DW-NOMINATE](#)," which compares a given legislator's voting record to every other lawmakers' to estimate their ideology: where they fall on the liberal to conservative spectrum.

While this analysis is based on a particular time period within a particular Congress, and thus is by definition a snapshot in time, it serves as an exploratory effort on behalf of Pew Research Center to use machine learning to quantify some of the ways that members of Congress communicate with the modern American public in official outreach efforts.

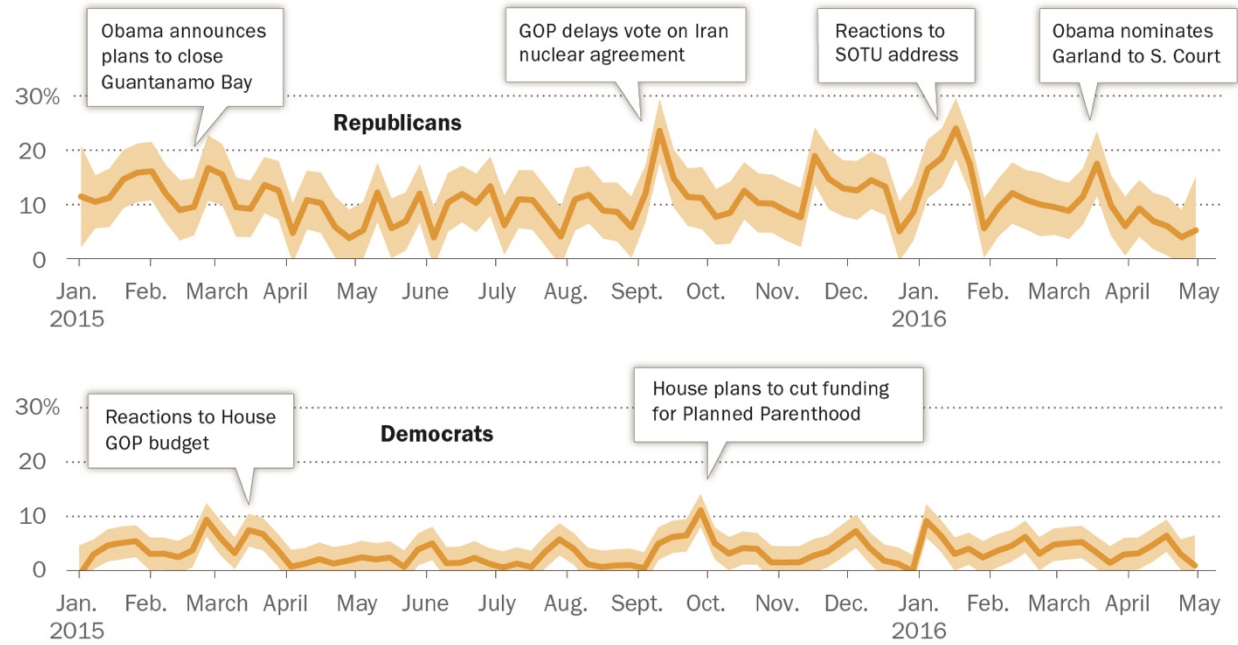
Other important findings

In the 114th Congress, Republicans – as the party that controlled the legislative branch but not the White House – criticized President Barack Obama in their press releases and Facebook posts more often than they criticized other Democrats. Indeed, the average Republican member directed indignant disagreement at the president in 15% of press releases, compared with just 2% that attacked other Democrats. Democratic members, for their part, took an emotionally charged stance against Republicans in 4% of their press releases.

Partisan disagreement often comes in bursts, which appear to follow partisan policy conflicts. In the 114th Congress, strong Democratic opposition to Republicans in press releases followed events such as the GOP's March 2015 budget proposal and September 2015 plans to defund Planned Parenthood. On the other side of the aisle, Republican opposition to the president and his party surfaced after Obama announced initiatives to sign a nuclear nonproliferation treaty with Iran in August and September 2015; prepared to close the Guantanamo Bay detention center in February 2015; and nominated Merrick Garland to the Supreme Court in March 2016.

Criticism coincides with contentious political developments

% of press releases expressing indignation toward the other party or its leader



Note: Over time estimates smoothed using Locally Weighted Scatterplot Smoothing (LOWESS). The shaded region represents a 95% confidence band, an attempt to quantify the uncertainty surrounding the estimated trend lines.

Source: Pew Research Center analysis of data from congressional websites and Lexis-Nexis. See Methodology section for details.

"Partisan Conflict and Congressional Outreach"

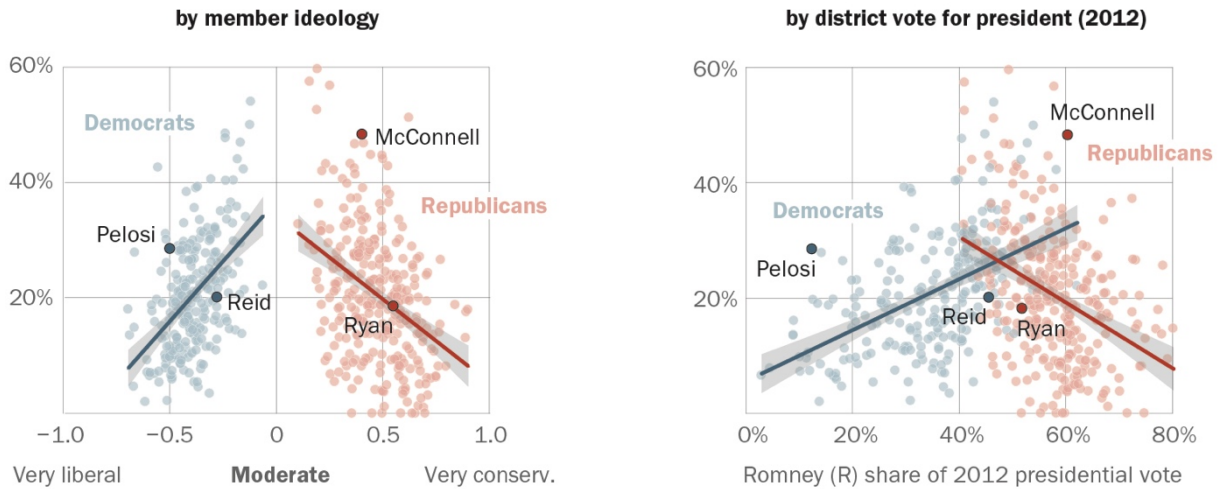
PEW RESEARCH CENTER

On the other hand, Democrats and Republicans issued bipartisan press releases at about the same rate. Bipartisan language appeared in about 21% of press releases for members of both parties. However, the average member of Congress discussed bipartisanship in just 6% of Facebook posts, again with only modest differences between members of the two major parties.

This difference is especially noteworthy in light of the fact that among Americans who pay attention to government and politics on Facebook, those who hold consistently liberal or conservative policy preferences are [more likely to follow politicians](#). Additionally, Americans with consistently liberal or conservative policy preferences are [less willing to support compromise](#) in Washington.

Moderates, those that represent competitive districts more likely to discuss bipartisanship

% of press releases containing bipartisan language



Note: Each point in the figure represents a member of Congress. Party leaders of the 114th Congress are shown as a point of reference. The leaders identified include Nancy Pelosi, Harry Reid, Mitch McConnell, and Paul Ryan. Estimated ideology based on roll call votes (DW-NOMINATE). The fitted line shows the relationship between each legislator's proportion of documents containing bipartisan language and their ideology (left) or district-level Romney vote share (right), for each party (estimated via ordinary least squares regression). The shaded region is that estimate's 95 percent confidence band, an attempt to quantify the range of uncertainty in this relationship.

Source: Pew Research Center analysis of data from congressional websites, Lexis-Nexis, Voteview.org, and Jacobson, Gary (2015). See Methodology section for details.

"Partisan Conflict and Congressional Outreach"

PEW RESEARCH CENTER

1. Who is going negative? Leadership and very liberal or conservative members most critical of other side

While it is easy enough for observers and pundits to declare that modern American politics is “too negative,” it is much harder for social scientists to measure that phenomenon with precision. This report draws on new computational tools and statistical methods that allow researchers to look for patterns in large amounts of text. These tools provide a new way to examine one corner of the political world: the extent to which disagreement and strong language feature in the press releases and Facebook posts of members of Congress.

Not only does this analysis put a rough number on the amount of negativity in these channels of communication, it also reveals patterns in who goes negative and how often. To do so, it classifies every communication collected using a method that combines human judgements and machine learning.

To start, Pew Research Center collected 94,521 press releases from the websites of members of the 114th Congress and from LexisNexis for a 16-month period from Jan. 1, 2015, to April 30, 2016. The Center also collected 108,235 public posts on each representative’s official accounts on Facebook for the same time period (see the methodology section for details).

These two media are vital channels for lawmakers to reach their constituents. Members can use press releases and Facebook posts to reach thousands – sometimes even millions – of people at once. Some alternatives include postal mail, which takes days to arrive and must meet strict “franking” rules, town halls, or direct interactions with news media outlets. Meanwhile, creating social media posts or issuing press releases demands a limited set of internal resources available to all members. And there is evidence that newspapers frequently use these press releases as a source when covering local politics.⁷

⁷ One analysis of coverage in a collection of local newspapers found that from 6% to 32% of members’ press releases were directly quoted in local papers, and the correlation between issues discussed in press releases and in newspaper stories mentioning representatives was 0.52 for low-circulation newspapers and 0.15 for high-circulation newspapers. Furthermore, there is evidence that this coverage increases the type of knowledge the public has about a member. See Grimmer, Justin. 2013. Representational Style in Congress: What Legislators Say and Why It Matters. p. 37, 128, 133–34.

Lawmakers use press releases and Facebook posts to reach the public

Examples focused on a new policy sponsored by Robert Hurt (R-Va.)

Press Releases

Home / Newsroom / Press Releases Print Tweet BOOKMARK

Recent Press Releases

Apr 27 2016

LEGISLATIVE UPDATE: House Approves Robert Hurt Bill to Encourage Investment in Startup Companies

WASHINGTON, D.C. – Congressman Robert Hurt (R-Virginia) today released the below statement following a vote by the House of Representatives to approve H.R. 4498, the Helping Angels Lead our Startups (HALOS) Act, which ensures that startup companies can connect with angel investors without inadvertently violating SEC rules. This bill, which Hurt introduced along with Congressman Steve Chabot (R-Ohio) and Congresswoman Kyrsten Sinema (D-Arizona), which was initially favorably reported from the Financial Services Committee with a bipartisan vote of 44-13, today was passed by the full House of Representatives on a bipartisan 325 to 89 vote. After the House vote, Congressman Hurt released the following statement:

"At a time when Virginia's Fifth District continues to face unacceptably high unemployment rates, we remain committed to removing overregulation of small businesses as a barrier to job creation. The HALOS Act would alleviate the burden placed on startup companies with regard to privacy and compliance concerns, which often require entrepreneurs and startups to take on burdens that they do not have the means to handle. In doing so, it allows them to focus on innovating and expanding, creating the jobs we so desperately need. The HALOS Act is a simple, bipartisan, bicameral solution that will ensure that investors and companies can continue their interaction. I was pleased to see the House of Representatives approve this bill, and urge the Senate to take up this legislation and work with us to allow our small businesses to prosper. I remain committed to removing unnecessary, bureaucratic red tape to make it easier for small businesses to strengthen our economy."

Permalink <http://hurt.house.gov/index.cfm/2016/4/legislative-update-house-approves-robert-hurt-bill-to-encourage-investment-in-startup-companies>

Congressman Robert Hurt ✓
April 28 · 🌐

I was pleased to see the House of Representatives approve the HALOS ACT yesterday evening, and urge the Senate to take up this legislation and work with us to allow our small businesses to prosper. The HALOS Act is a simple, bipartisan, bicameral solution that will ensure that investors and companies can continue their interaction. I remain committed to removing unnecessary, bureaucratic red tape to make it easier for small businesses to strengthen our economy.



684 Views
36 Likes 2 Comments 1 Share

Share

Source: Facebook, house.hurt.gov
"Partisan Conflict and Congressional Outreach"

PEW RESEARCH CENTER

Unfortunately for lawmakers who have made press releases a core part of their communications strategy, there are [fewer daily newspapers now than there were 10 years ago](#). However, [politicians – and Americans generally](#) – have been using social media more and more. Facebook, which [Americans use more often than any other social media site](#), offers politicians an opportunity to reach a very large audience with very few limitations on what they can say. By using Facebook, a lawmaker's staff can also keep track of how often users like a post, comment on it or share it with their friends. Thanks to this information, they can estimate how well or how widely a message was received in that network – and researchers can, too.

After the press releases and Facebook posts were collected, Pew Research Center analyzed the text using a combination of trained human readers and machine learning, with the goal of identifying the share of members' communication that included opposition to the other side, the share that contained indignant remarks, and the share that expressed bipartisan sentiment. The Center

developed criteria by which human content coders could judge whether a given communication included disagreement with the other side or opposition to its policies, whether the lawmaker who sent it was “going negative” and whether the message expressed bipartisan sentiment. Researchers used this sample to “train” machine-learning algorithms which were then applied to the remainder of the texts, yielding estimates of bipartisanship, opposition and negativity for every post and press release the Center obtained.

To better illustrate the differences between categories used in this report, the following section contains example documents for disagreement, indignant disagreement and bipartisanship. In each case, all five analysts who read the document agreed that each message contained “disagreement,” “indignant disagreement,” or “bipartisanship.” These documents have been chosen for illustrative purposes and do not represent a random sample.

Examples of congressional press releases and Facebook posts

Disagreement

“It may be the start of a new Congress, but we're seeing the same old GOP giveaways to Wall Street. Today, Republicans held a vote on a bill, introduced late last night, that combines 11 bills into 1 big gift for special interests and big banks.” – *Democratic Facebook post*

“As Ways and Means Committee chairman, I’m proud to keep blocking President Obama’s tax increases on American energy, which would send Texas jobs overseas, and to help successfully lift the federal ban on exporting crude oil.” – *Republican Facebook post*

“This bill gives the Obama administration the tools it needs to use diplomacy to stop Iran from developing a nuclear weapon. Unlike last month's unfortunate attempt by Senate Republicans to sabotage the diplomatic process, this bill provides a meaningful role for Congress without undermining diplomacy.” – *Democratic press release*

“Over six years have passed since the Keystone XL pipeline application was first submitted to the U.S. Department of State, and all the while, instead of approving its construction, President Obama has supported an energy agenda counterproductive to American success. Throughout the 112th and 113th Congresses, my fellow House Members and I have fought endless battles to overcome these administration-instituted delays.” – *Republican press release*

Indignant disagreement

“This is an outrage. House Republicans have already spent over 20 months and \$4.5 million digging into Hillary Clinton's emails and dragging her in to testify Now, they are not satisfied with the results of their last fishing expedition so they are shifting gears to investigate Hillary's records directly because they're concerned about ‘federal record-keeping.’ Give me a break As Majority Leader McCarthy openly admitted, this is a purely political effort to influence the U.S. presidential election. It should be funded by Republican campaign donors instead of the American people. It's time to end this egregious waste of taxpayer resources and stop this unethical abuse of the powers of Congress.” – *Democratic press release*

“This is a blatant power grab by the Obama Administration The House of Representatives ... has voted multiple times to overturn this shameless seizure of power The EPA claims this will be used to clean up “dirty waters”. Unfortunately, now the dirtiest thing about these waters is the fat, power hungry fingers of federal regulators dipping in to regulate it.” – *Republican Facebook post*

“Republicans have only been in control for a week and already they are picking an unnecessary political fight that risks shutting down the Department of Homeland Security and endangering our security... This is not a game and it is time for Republicans to take their responsibility to govern seriously, instead of playing to the most extreme voices in their party.” – *Democratic press release*

“My colleagues in both chambers of Congress rightfully feel the moral necessity of responding to such unaccountable overreach from this president. His actions were purely political and designed to drive a wedge between burgeoning minority communities and a center right country which prefers its government focus on economic issues and growing a stagnant economy. We must respond to the president's assault on the democratic process by exercising our constitutional powers...”
– *Republican press release*

Bipartisanship

“Don't say bipartisanship is dead. Today, most Republicans joined almost every Democrat to force reauthorization of the Export-Import Bank through the House over the objection of Tea Party hardliners. I joined this bipartisan coalition because the Ex-Im Bank supports good paying jobs across the United States, and we must keep our economic recovery

moving forward. I hope the Senate will follow our example and send reauthorization of the bank to the President as soon as possible.” – *Democratic Facebook post*

“The bipartisan passage of NAHASDA demonstrates that the House of Representatives is capable of working together to get important things accomplished... [t]ogether we found a solution that we all could agree, even if we each do not agree with every provision in the bill... this legislation is the product of a truly bipartisan process.”

– *Democratic press release*

“Republicans and Democrats alike support eliminating the costs of unnecessary and obsolete regulations to help economic recovery.” – *Republican Facebook post*

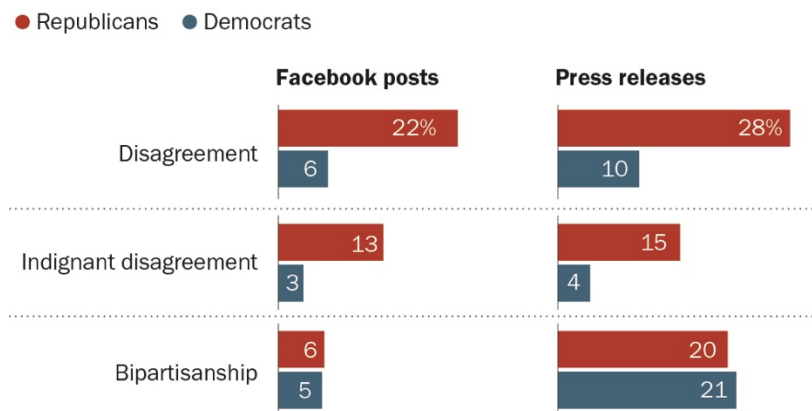
“So much more unites us than divides us. When it comes to helping America get back to work, there is no partisanship. I will continue to work with both Republicans and Democrats to find ways to make life just a little bit easier.” – *Republican Facebook post*

Democrats criticized Republicans, who criticized Obama

Overall, the kind of indignant disagreement that one might think of as “going negative” is relatively rare in official congressional communications. This study found that for an average member, roughly 10% of press releases and 9% of Facebook posts could be characterized as being strongly negative toward the other party or its leadership, a smaller share than were categorized as being bipartisan. On average, 21% of press releases and 15% of Facebook posts contained some form of disagreement.

Under Obama, Democrats less likely than Republicans to criticize other side

Average % of members’ communications containing or mentioning...



Note: Figures are averages of machine-learned data.

Source: Pew Research Center analysis of data from Facebook OpenGraph API, congressional websites, and Lexis-Nexis. See Methodology section for details. “Partisan Conflict and Congressional Outreach”

PEW RESEARCH CENTER

But politics is situational, and different actors are called to negativity at different times. For members of Congress, political strategy differs greatly depending on which party controls the White House, the Senate and the House of Representatives. For the session of Congress under study – the 114th – the executive branch was controlled by the Democrats while Republicans controlled both the House and the Senate.

Reflecting GOP disagreement with President Obama, the analysis finds that in their public communications, Democrats and Republicans did not argue against the other’s policies and trade insults in equal measure. Instead, Democrats criticized Republicans, while for the most part the focus of Republican ire was President Obama.⁸ In this asymmetrical relationship – with congressional Republicans facing off against a Democratic president – Republican legislators came off as more likely to be negative in their press releases and social media posts.

In official press releases, the average Republican was almost three times as likely to disagree with Democrats or Obama than the average Democrat was to disagree with Republicans (28% for Republicans vs. 10% for Democrats). On average, Republicans also expressed indignant disagreement nearly four times more often than Democrats (15% vs. 4%).⁹ This pattern holds for Facebook posts as well.

For example, Republicans were more than three times as likely as Democrats to express disagreement with the other side on Facebook (22% for Republicans vs. 6% for Democrats), and about four times more likely to create Facebook posts containing indignant disagreement (13% vs. 3%), on average.

Party leaders expressed more disagreement and indignation than rank-and-file members

Congressional party leaders, including the majority and minority leaders in each chamber and their party whips, expressed more disagreement and issued more indignant statements than rank-and-file members. Fully 37% of press releases from leaders expressed disagreement with the other side, compared with just 20% for the average non-leadership member. And 21% of leaders’ press releases expressed indignation with the other side, compared with 10% for others, on average.

⁸ Intra-party disagreement was too rare to estimate reliably; only inter-party disagreement is considered. See the methodology section for details.

⁹ The “disagreement” and “indignant disagreement” categories are *not* mutually exclusive – statements that contain indignant disagreement are a subset of those that contain disagreement more broadly.

In the 114th Congress, when Democrats controlled the executive branch, Republican leaders expressed more negativity than their Democratic counterparts – a difference in line with overall between-party trends. On average, nearly one-in-four (24%) of Republican leaders' posts expressed indignant opposition, compared with 15% for rank-and-file members. For Democrats, it was 17% for party leaders and 4% for the rank-and-file. It's worth noting that Democratic Party whips expressed substantially lower levels of indignant opposition, on average: Just 3% of their press releases contained indignation, compared with 30% for Democratic minority leaders in the House and Senate.

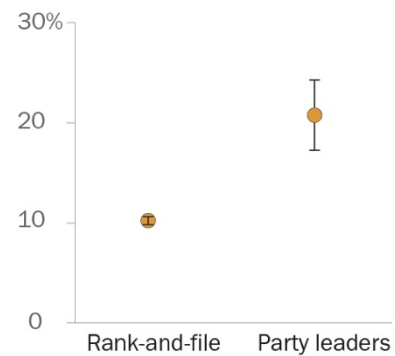
Negativity more likely among most conservative and liberal legislators

There is a consistent relationship between the liberal or conservative leanings of a legislator and the extent to which they express political disagreement, both in press releases and Facebook posts. A given legislator's ideological position is estimated using the DW-NOMINATE scale, which places lawmakers on a spectrum from roughly -1 (very liberal) to 1 (very conservative) based on similarities and differences in their legislative roll-call voting records.

Relative to moderate members of their party, Republicans with the most conservative voting records¹⁰ were more likely to express disagreement with President Obama. While the most moderate Republican members disagreed with Obama in 19% of press releases, among those estimated to be most conservative, fully 34% of press releases opposed the president on average. Similarly, the most conservative Republicans were twice as likely to issue Facebook posts that disagreed with the president, compared with the most moderate members of their party: 24% versus 12%, respectively. The same pattern exists for disagreements with Democrats in general, but at a much lower rate. For example, the most moderate Republicans expressed disagreement with Democrats in 2% of releases, compared with

Leadership of both parties more indignant than rank-and-file

Average % of press releases containing indignant disagreement



Note: Points are averages of machine-learned data. Lines indicate the standard error, an attempt to quantify the uncertainty surrounding each estimate. Congressional party leaders include the majority and minority leaders in each chamber and their party whips. Source: Pew Research Center analysis of data from congressional websites and Lexis-Nexis. See Methodology section for details. "Partisan Conflict and Congressional Outreach"

PEW RESEARCH CENTER

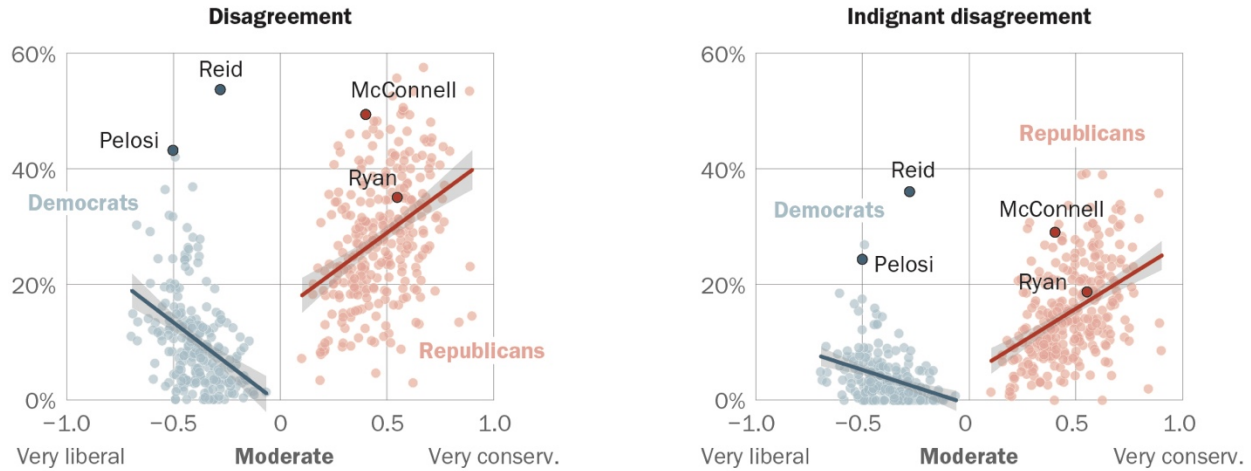
¹⁰ The most conservative or liberal lawmakers in each party are defined here as those with the top and bottom 10% of all DW-NOMINATE scores. Moderates are defined as the 10% of members in each party closest to the midpoint of the DW-NOMINATE scale.

6% for the most conservative Republicans. On Facebook, the most conservative Republicans disagreed with Democrats in 4% of posts, compared with just 1% among moderate Republicans.

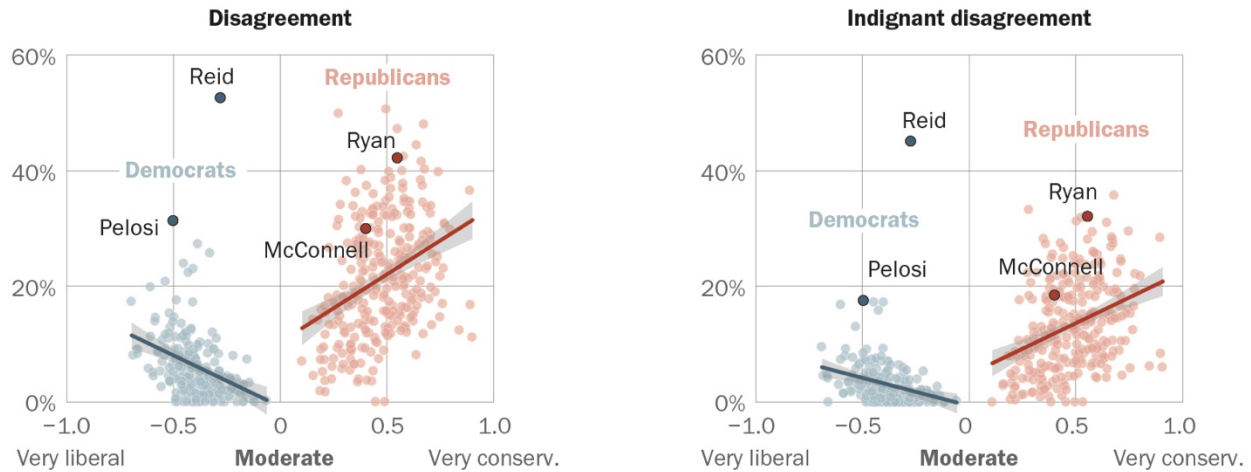
The pattern among Democrats is similar, with the most liberal Democrats expressing the highest rates of disagreement with Republicans. However, this trend is less pronounced, in part because of the relatively low rates of disagreement expressed by members of the Democratic Party. The most liberal Democrats disagreed with Republicans in 15% of press releases on average, compared with 5% among moderate Democrats. On Facebook, 10% of the most liberal members' posts opposed Republicans on average, while the most moderate Democrats took such a stance in 2% of posts.

Moderates less likely to disagree with members of the other political party

% of press releases containing ...



% of Facebook posts containing ...



Note: Each point in the figure represents a member of Congress. Party leaders of the 114th Congress are shown as a point of reference. The leaders identified include Nancy Pelosi, Harry Reid, Mitch McConnell, and Paul Ryan. Estimated ideology based on roll call votes (DW-NOMINATE). The fitted line shows the relationship between each legislator's proportion of expressed disagreement or indignant disagreement and their ideology, for each party (estimated via ordinary least squares regression). The shaded region is that estimate's 95 percent confidence band, an attempt to quantify the range of uncertainty in this relationship.

Source: Pew Research Center analysis of data from congressional websites, Lexis-Nexis, Facebook OpenGraph API, and Voteview.org. See Methodology section for details.

"Partisan Conflict and Congressional Outreach"

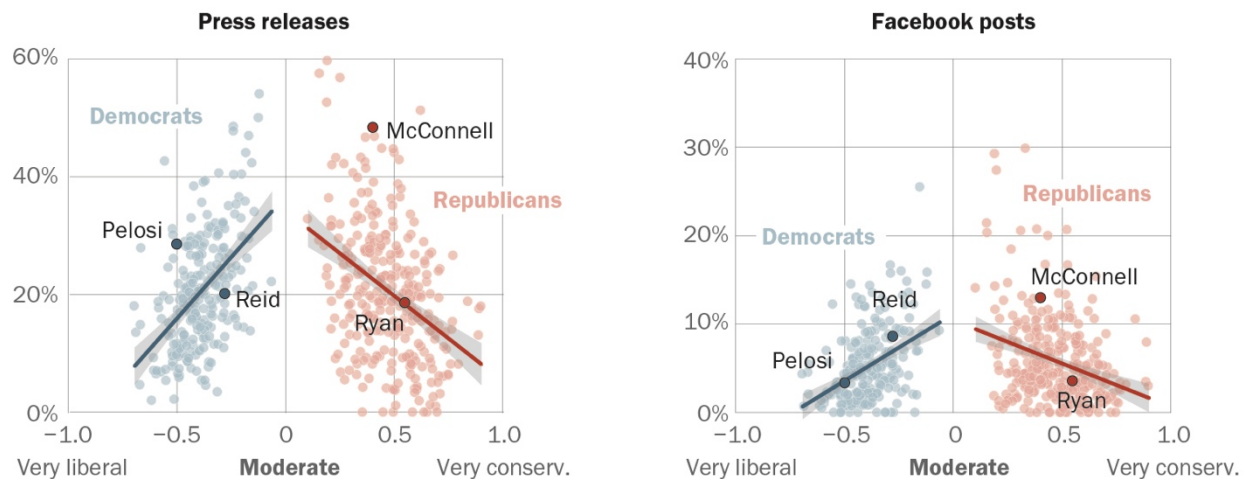
PEW RESEARCH CENTER

Moderates in both parties emphasized bipartisanship and constituent benefits

In contrast to very conservative and liberal members, ideological moderates – members who more often break ranks when it comes to voting on legislation – were more likely to advertise the fact that they were willing to cross the aisle. Among Republicans, moderates discussed bipartisanship in 28% of press releases, while moderate Democrats did so in 30%. But very liberal and conservative legislators only brought up bipartisan action in 14% and 12% of press releases, respectively.

Moderate members of Congress most likely to highlight bipartisanship

% of press releases and Facebook posts containing bipartisan language



Note: Each point in the figure represents a member of Congress. Party leaders of the 114th Congress are shown as a point of reference. The leaders identified include Nancy Pelosi, Harry Reid, Mitch McConnell, and Paul Ryan. Estimated ideology based on roll call votes (DW-NOMINATE). The fitted line shows the relationship between each legislator’s proportion of expressed bipartisanship and their ideology, for each party (estimated via ordinary least squares regression). The shaded region is that estimate’s 95 percent confidence band, an attempt to quantify the range of uncertainty in this relationship.

Source: Pew Research Center analysis of data from congressional websites, Lexis-Nexis, Facebook OpenGraph API, and Voteview.org. See Methodology section for details.

“Partisan Conflict and Congressional Outreach”

PEW RESEARCH CENTER

The Center separately measured how often members of Congress discussed constituent benefits in press releases, defined as favorable legislative outcomes and/or government spending that directly benefit an elected official’s home state or district. When members of Congress advertised their policy accomplishments to constituents, they often emphasized how their efforts were having – or

would have – a direct impact on those back home in their districts. Such communications most frequently took the form of announcements of new government expenditures, programs that provide jobs, tax benefits, and funding for local programs. Past work has shown that such messages most often have a more powerful impact in terms of building support for representatives than other types of communications.¹¹

On average, Republicans were less likely than Democrats to discuss how government policies and programs benefit their constituents: they did so in just 13% of press releases, compared with 23% for Democrats.¹² This finding is notable in the context of lower support for earmarks and other forms of federal spending that have played an increasing role in Republican politics since the House and Senate adopted a [ban on earmarks](#) in 2011, after a [concerted push from newly elected tea party-affiliated legislators](#) who campaigned on the issue.¹³

¹¹ Grimmer, Justin, Messing, Solomon, and Westwood, Sean. J. 2012. [How words and money cultivate a personal vote: The effect of legislator credit claiming on constituent credit allocation](#). American Political Science Review; Grimmer, Justin, Sean J. Westwood, and Solomon Messing. 2015. [The Impression of Influence: Legislator Communication, Representation, and Democratic Accountability](#).

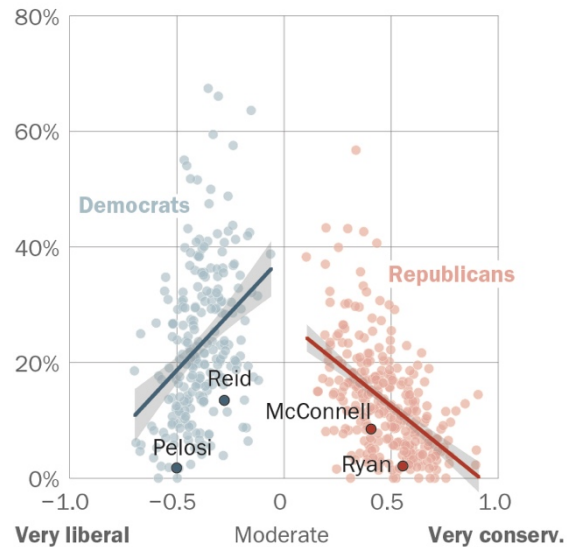
¹² Overall, 20% of all press releases coded by humans in the initial sample discussed constituent benefits. However, the number of Facebook posts focusing on constituent benefits was very low – about 6% of the total number of posts that were manually reviewed. Reliable machine classification across the entire collection could not be achieved for the latter, so this analysis examines only discussion of benefits in press releases. This could be due to some combination of the low incidence rate of cases that discuss constituent benefits, the brevity of Facebook posts, and/or the comparatively wide range of language used to discuss benefits. Some other measures used in this report that occur at similarly low rates – Republican disagreement with Democrats for example – involve a narrower range of language, which can translate to more reliable machine classification.

¹³ See also Grimmer, Justin, Sean J. Westwood, and Solomon Messing. 2015. [The Impression of Influence: Legislator Communication, Representation, and Democratic Accountability](#). This shows a marked decline in the number of Republican press releases claiming credit for bringing constituents benefits back to their districts through 2010.

Overall, ideological moderates in both parties were more than twice as likely to announce constituent benefits in their press releases, compared to very liberal and very conservative members (23% for moderates vs. 9%). But this difference is most pronounced among Democrats: about one-in-four (26%) of the most moderate Democrats' press releases focused on benefits, compared with just 13% for the most liberal legislators, as measured by DW-NOMINATE. Furthermore, the share of members' press releases mentioning constituent benefits has a strong, negative correlation with indignant disagreement.¹⁴ These differences also stand out when considering members of specific caucuses, like the conservative [House Freedom Caucus](#) and the liberal [Progressive Caucus](#). Republicans in the [Liberty Caucus](#) and the Freedom Caucus, who tend to oppose expanding the role of government, were least likely to highlight constituent benefits, raising the topic in less than 8% of press releases, on average.

Very liberal, very conservative members less likely to highlight constituent benefits

% of press releases mentioning constituent benefits



Note: Each point in the figure represents a member of Congress. Party leaders of the 114th Congress are shown as a point of reference. The leaders identified include Nancy Pelosi, Harry Reid, Mitch McConnell, and Paul Ryan. Estimated ideology based on roll call votes (DW-NOMINATE). The fitted line shows the relationship between each legislator's proportion of benefits-focused releases and their ideology, for each party (estimated via ordinary least squares regression). The shaded region is that estimate's 95 percent confidence band, an attempt to quantify the range of uncertainty in this relationship.

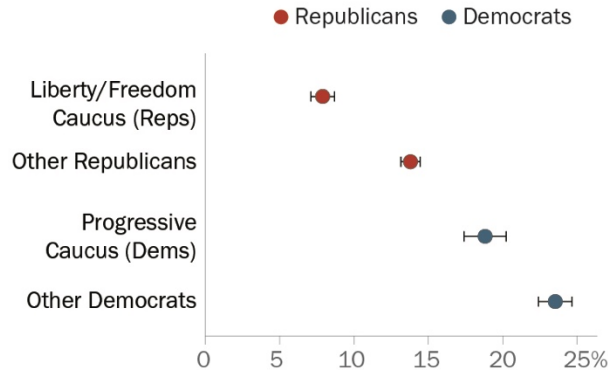
Source: Pew Research Center analysis of data from congressional websites, Lexis-Nexis, Facebook OpenGraph API, and Voteview.org. See Methodology section for details. "Partisan Conflict and Congressional Outreach"

PEW RESEARCH CENTER

¹⁴ The correlation coefficient is 0.53. This association persists in a multivariate OLS regression model predicting percentage of press releases containing indignant disagreement based on DW-NOMINATE, DW-NOMINATE-squared (to capture extremism), and party leadership status.

Members of more liberal, conservative caucuses discuss benefits less

Average % of members' press releases mentioning constituent benefits



Note: Lines indicate the standard error, an attempt to quantify the uncertainty surrounding each estimate.
Source: Pew Research Center analysis of data from congressional websites and Lexis-Nexis. See Methodology section for details.

“Partisan Conflict and Congressional Outreach”

PEW RESEARCH CENTER

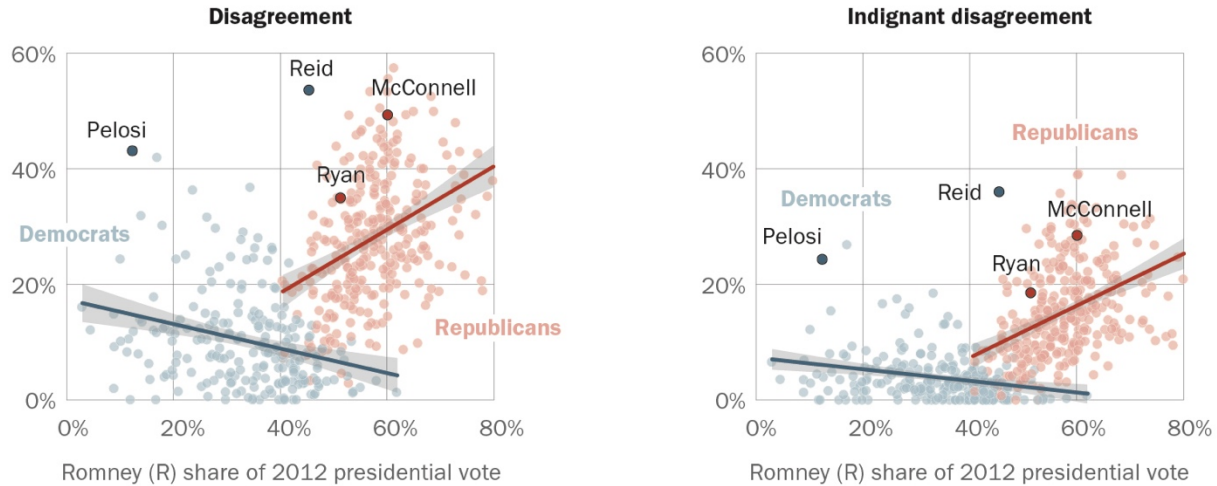
Legislators in less competitive districts more likely to attack the other party

Elected officials who represent more competitive districts and states – where the vote for president in 2012 was closer – tended to avoid criticism in their press releases and posts. But Republicans representing the least competitive districts – where Obama did the worst in his run for re-election – issued a substantially higher proportion of press releases containing disagreement (32%) than did Republicans in the most competitive districts (20%).¹⁵ On average, Republicans in the least competitive districts also issued about twice as many indignant releases (19% vs. 9%). A similar version of this pattern holds for Democrats. Democrats representing the least competitive districts – where GOP presidential candidate Mitt Romney did worst – issued about three times as many press releases containing disagreement (14% vs. 5%) and indignation (6% vs. 2%) than Democrats in the most competitive districts.

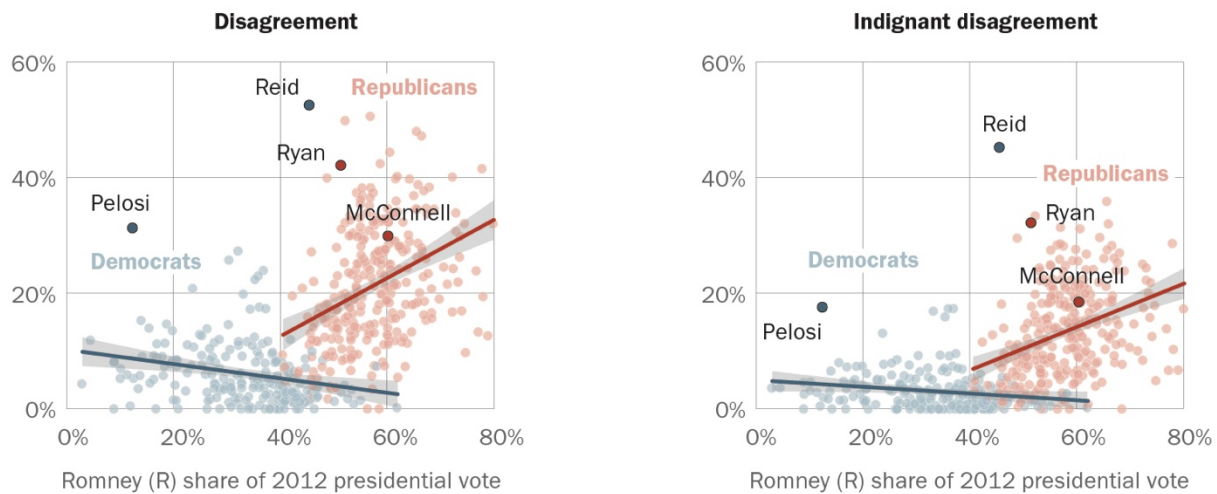
¹⁵ The “most competitive” districts are defined as those where votes for Romney and Obama in 2012 were closer than that in 90% of other districts. The “least competitive” districts are those in which Romney or Obama won by a margin higher than that of 90% of other members’ districts. These two groups of top 10% districts are computed separately for members of each party.

Members representing less competitive districts most likely to criticize other side

% of press releases containing ...



% of Facebook posts containing ...



Note: Each point in the figure represents a member of Congress. Party leaders of the 114th Congress are shown as a point of reference. The leaders identified include Nancy Pelosi, Harry Reid, Mitch McConnell, and Paul Ryan. The fitted line shows the relationship between each legislator's proportion of expressed disagreement or indignant disagreement and district-level Romney vote share, for each party (estimated via ordinary least squares regression). The shaded region is that estimate's 95 percent confidence band, an attempt to quantify the range of uncertainty in this relationship. Romney vote share aggregated at state level for senators, district level for representatives.

Source: Pew Research Center analysis of data from congressional websites, Lexis-Nexis, Facebook OpenGraph API, and Jacobson, Gary (2015). See Methodology section for details.

"Partisan Conflict and Congressional Outreach"

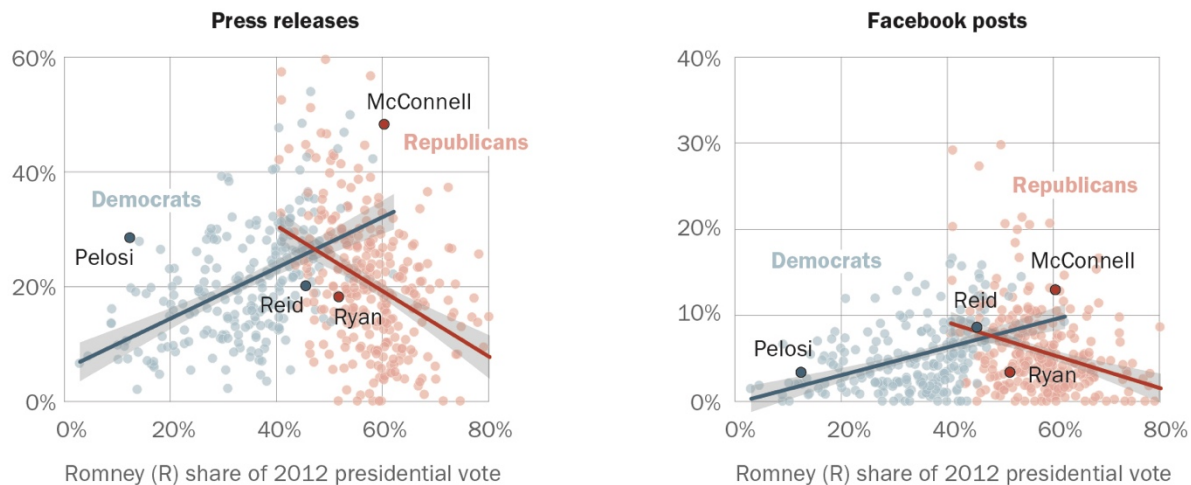
PEW RESEARCH CENTER

Members from competitive districts more likely to discuss bipartisanship

Members representing the most competitive districts were more likely to promote bipartisanship in their outreach efforts. On average, Republicans representing the most strongly Republican states and districts issued bipartisan messages in just 16% of press releases, compared with 29% among Republicans in the most competitive districts. By the same token, 12% of the press releases issued by Democrats in the least competitive districts contained bipartisan language. Those in competitive districts, by contrast, discussed bipartisanship in 30% of press releases, on average. These findings are consistent with existing research suggesting that “going bipartisan” is a strategy meant to increase support among moderates or swing voters.¹⁶

Members representing less competitive districts are less likely to discuss bipartisanship

% of press releases and Facebook posts containing bipartisan language



Note: Each point in the figure represents a member of Congress. Party leaders of the 114th Congress are shown as a point of reference. The leaders identified include Nancy Pelosi, Harry Reid, Mitch McConnell, and Paul Ryan. The fitted line shows the relationship between each legislator’s proportion of documents containing bipartisan language and district-level Romney vote share, for each party (estimated via ordinary least squares regression). The shaded region is that estimate’s 95 percent confidence band, an attempt to quantify the range of uncertainty in this relationship. Romney vote share aggregated at state level for senators, district level for representatives.

Source: Pew Research Center analysis of data from congressional websites, Lexis-Nexis, Facebook OpenGraph API, and Jacobson, Gary (2015). See Methodology section for details. “Partisan Conflict and Congressional Outreach”

PEW RESEARCH CENTER

¹⁶ See Trubowitz, Peter and Nicole Mellow. 2005. “‘Going Bipartisan’: Politics by Other Means,” *Political Research Quarterly*.

2. How the public reacted on Facebook

Politicians are [increasingly turning to social media](#) to engage directly with their constituents and other members of the public. Unlike press releases, where audience feedback is indirect and difficult to measure, Facebook enables people to react directly to what members of Congress say in the form of likes, comments and shares, providing an easily accessible and direct measurement of engagement with members' content.

On Facebook – the social media platform with the [largest user base](#) among online U.S. adults – the audience for posts includes people who follow a legislator's Facebook page, and depending on other factors such as whether the post is promoted, may include followers' friends and others. What's more, a lawmaker's official Facebook page and posts are publicly available.

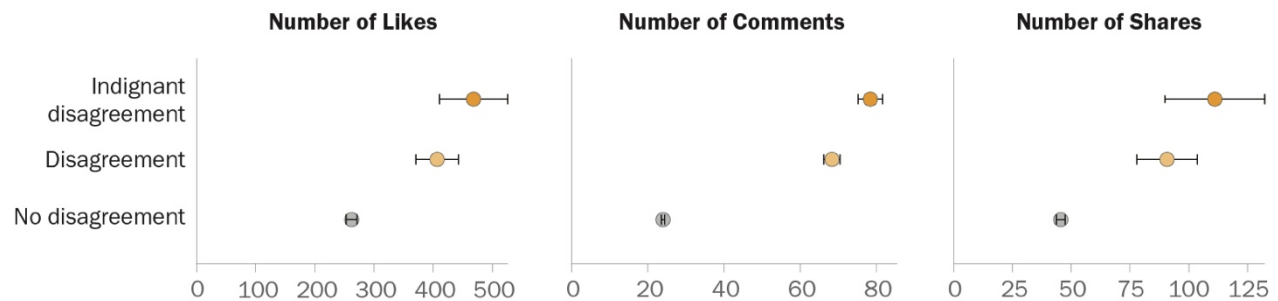
Followers and other users may see posts from members in their News Feed, which relies on a ranking system that [gives prominence to content that receives more likes, comments and shares](#). Thus, posts that generate more engagement are likely to be seen by more people.

Disagreement, indignation saw greater engagement on Facebook

Posts that contained political disagreement and indignant rhetoric were far more likely to elicit user engagement than posts that did not. On average, posts expressing disagreement with the opposing party received one and a half times as many likes as posts without disagreement (406 likes, compared with 262), and almost three times as many comments (68 vs. 24). This pattern holds for the number of shares, which was twice as high for posts with disagreement (91 shares vs. 45).

Critical posts get more likes, comments, and shares than other posts

Average number of likes, comments, and shares per Facebook post containing ...



Note: Lines indicate the standard error, an attempt to quantify the uncertainty surrounding each estimate. The “disagreement” and “indignant disagreement” categories are not mutually exclusive: statements that contain indignant disagreement are a subset of those that contain disagreement more broadly.

Source: Pew Research Center analysis of data from Facebook OpenGraph API. See Methodology section for details. “Partisan Conflict and Congressional Outreach”

PEW RESEARCH CENTER

When disagreement crossed the line to indignation, posts received an even higher boost, with almost two times as many likes (468 vs. 262), about three times as many comments (78 vs. 24) and more than twice as many shares, (111 vs. 45) on average.

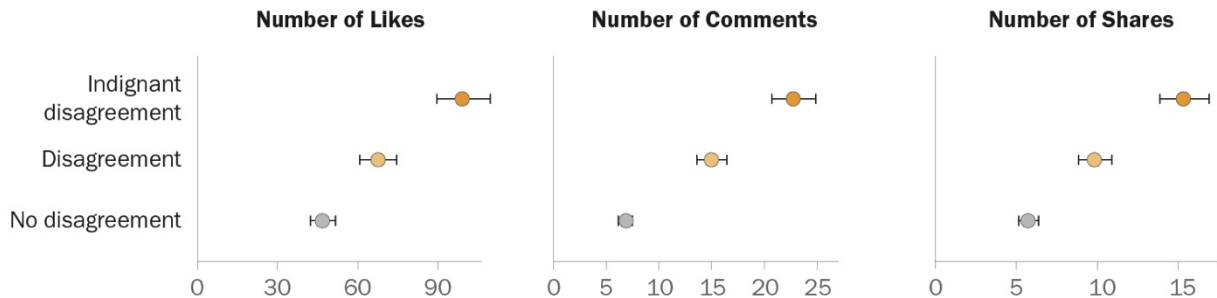
There was far less engagement with posts focused on bipartisan activity. Those posts averaged just 166 likes, 28 comments and 30 shares.

The most liberal and conservative members received more likes, comments and shares than moderates’ posts did. Very conservative Republicans – those in the top 10% of the DW-NOMINATE measure – averaged 1,164 likes, 123 comments, and 187 shares while moderate Republicans earned 78 likes, 14 comments, and 12 shares. Among Democrats, the most liberal elected officials earned 1,266 likes, 63 comments, and 285 shares, while moderates earned 140 likes, 22 comments and 18 shares, on average.

Statistical models can help answer the question of whether negativity is associated with engagement overall, or whether members who are more likely to express negativity happen to have large groups of highly engaged followers for other reasons. These models provide evidence for a direct association between negativity and engagement.

Critical statements earn more engagement, regardless of members' popularity

Estimated Facebook engagement for a post containing ...



Note: The figures show the estimated number of additional likes, comments, and shares for posts containing disagreement or indignation. The estimates are based upon a model that controls for a member's overall popularity, defined as their average engagement on Facebook posts. See methodology for details. Lines indicate the standard error, an attempt to quantify the uncertainty surrounding each estimate. Statements that contain indignant disagreement are a subset of those that contain disagreement more broadly.

Source: Pew Research Center analysis of data from Facebook OpenGraph API. See Methodology section for details. "Partisan Conflict and Congressional Outreach"

PEW RESEARCH CENTER

The relationship between disagreement expressed in posts and engagement with each post persisted when employing statistical models that account for average engagement with each member's posts, suggesting that the additional likes, comments and shares associated with disagreement are not merely artifacts of particularly popular legislators posting critical content.¹⁷

An important caveat is that this analysis did not control for the total number of people who actually saw the post, as Facebook does not make these data publicly available. It is possible, for example, that posts with more controversial content prompted more comments than other posts, which in turn increased their score in the News Feed ranking system, increasing the number of people who saw these posts. There may have been other differences, for example due to the timing of when posts were issued or when audiences were online.

¹⁷ Specifically, the models examine the relationship between the log (base 10) number of likes, comments and shares that a given post receives and the party of the member who created the post. The models include a random intercept for each member, which accounts for each member's average level of online engagement. Fixed effects models produce similar results, and both sets of models are reported in the methodology section. .

3. Partisan language in congressional outreach

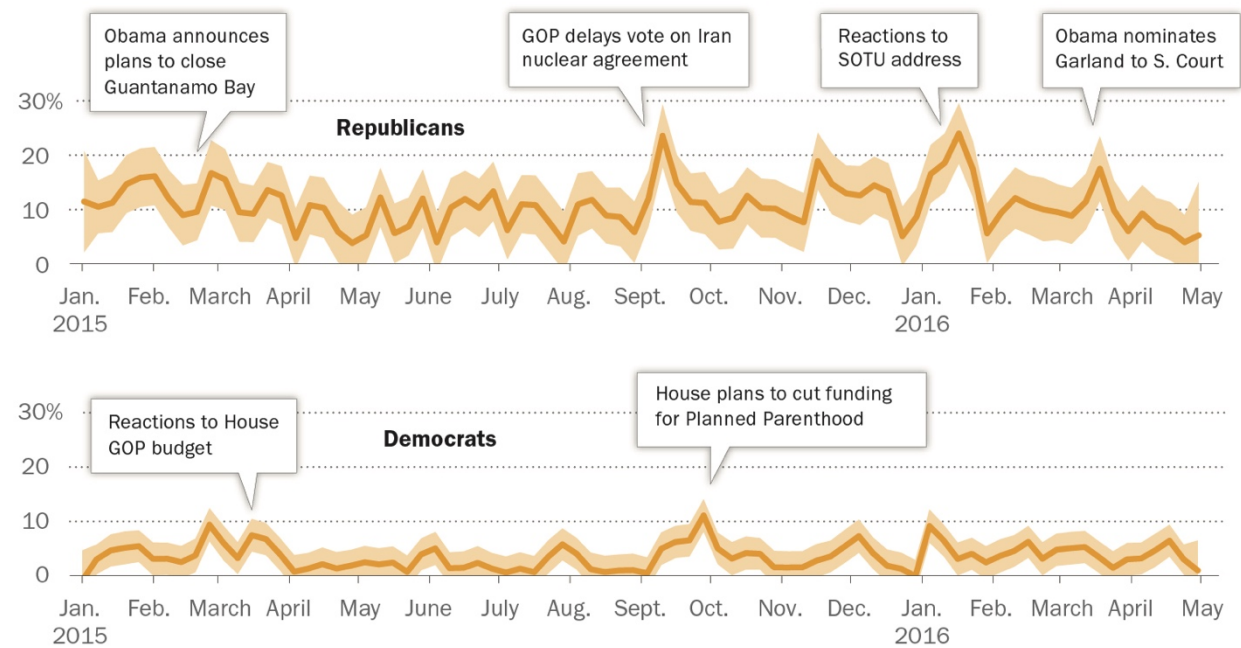
Drilling down into the issues that arise most often when the nation's legislators criticize the other party helps to reveal the arguments at the heart of a contentious Congress. What follows is a more systematic examination of how negative partisan speech fluctuates with the ebb and flow of political developments, as well as the specific words and phrases that best distinguish indignant disagreement from the many other ways legislators communicate with the public.

What disagreement looks like

Partisan disagreements do not merely occur on a constant basis, or at random. Instead, they come in bursts that follow partisan policy fights, often related to political developments.

Criticism coincides with contentious political developments

% of press releases expressing indignation toward the other party or its leader



Note: Over time estimates smoothed using Locally Weighted Scatterplot Smoothing (LOWESS). The shaded region represents a 95% confidence band, an attempt to quantify the uncertainty surrounding the estimated trend lines.

Source: Pew Research Center analysis of data from congressional websites and Lexis-Nexis. See Methodology section for details.

"Partisan Conflict and Congressional Outreach"

PEW RESEARCH CENTER

Democrats criticized Republicans in press releases after events such as the GOP's March 2015 budget proposal and September 2015 plans to defund Planned Parenthood. Republican indignant opposition to the president and his party surfaced after Obama announced initiatives to sign a nuclear nonproliferation treaty with Iran in August and September 2015; prepared to close the Guantanamo Bay detention center in February 2015; and nominated Merrick Garland to the Supreme Court in March 2016.

Mutual Information and Text Classification

Researchers measured how unusual or distinctive particular words were in a specific document category (such as disagreement, indignant disagreement, or bipartisanship). The measure presented in this section is called [normalized pointwise mutual information](#). It captures how often a word actually appeared in a particular type of document – like those that contained indignant disagreement – compared with how often that word might have appeared by chance alone. The plots below rank the 25 most distinctive words in various categories in the y-axis.

For example, consider the word “Republican,” as shown in the first plot illustrating the differences between indignant disagreement and other statements from Democrats below. As the y-axis shows, that word has the highest mutual information and thus most reliably distinguishes indignant disagreement from other statements by Democrats. In other words, the most unusual word to see in the non-indignant statements from Democrats was the word “Republican.” Thus, the word “Republican” can be used to resolve uncertainty about the category to which it belongs; most often, statements from Democrats that contained the term “Republican” expressed indignant disagreement.

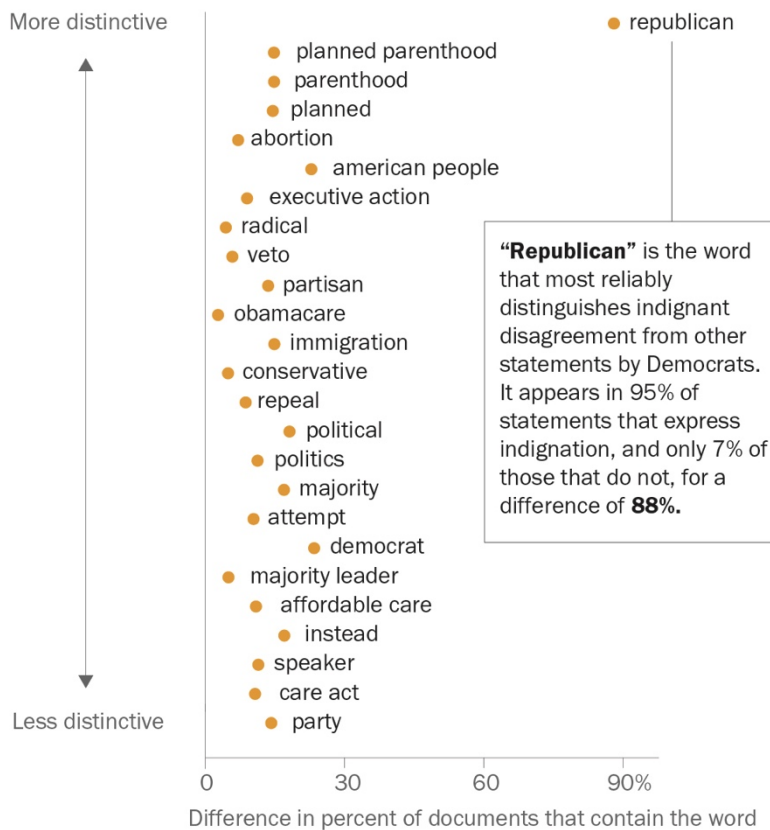
Researchers also compared how often the same words appeared in each category. Specifically, this captures the difference between how often a given word appeared in documents of a particular type and how often it occurred in other types of documents.

Using the same example, the x-axis in the plot below shows how much more frequently the word “Republican” occurred in documents containing indignant disagreement compared with those that did not. Overall, 95% of indignant statements from Democrats contained the word “Republican,” while only 7% of other statements issued by Democrats did. Thus the difference in the percent of documents from each category is 88.

Words and phrases that are plainly associated with immigration, the Affordable Care Act (ACA), efforts to defund Planned Parenthood and replacing the late Supreme Court Justice Antonin Scalia

Words that distinguish indignant disagreement among Democrats

The 25 words most distinctive of **Democratic** press releases or Facebook posts that contain indignant disagreement, in rank order by pointwise mutual information



Source: Facebook OpenGraph API, Pew Research Center analysis of data from congressional websites and Lexis-Nexis. See Methodology section for details. “Partisan Conflict and Congressional Outreach”

PEW RESEARCH CENTER

all distinguished Democratic attacks on Republicans from other content. One example of this kind of attack, which came from Sen. Barbara Boxer (D-Calif.) in the wake of a Republican-supported bill limiting access to abortion, follows:

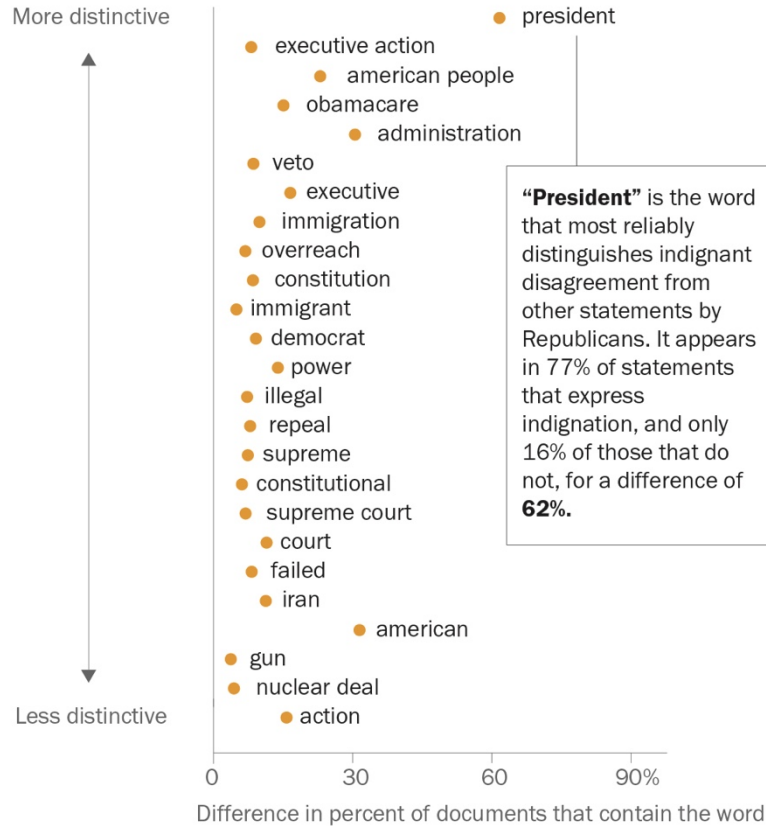
“The women of America won’t forget these outrageous attacks on their health and their lives. We won’t let Republicans turn back the clock. We will fight back.”

Republican criticism of Democrats used words and phrases related to the president, the ACA, a nuclear deal with Iran, immigration, executive power and veto actions. Notably, the word “Democrat” did not appear; since disagreement with congressional Democrats was so much less frequent than disagreement with Obama, the word does not distinguish Republican indignant disagreement.¹⁸

¹⁸ To validate these results, researchers manually examined samples of documents containing indignant disagreement and checked the correspondence between words in those documents and the words included in the figures here.

Words that distinguish indignant disagreement among Republicans

The 25 words most distinctive of **Republican** press releases or Facebook posts that contain indignant disagreement, in rank order by pointwise mutual information



Note: Percentages do not add up due to rounding.

Source: Facebook OpenGraph API, Pew Research Center analysis of data from congressional websites and Lexis-Nexis. See Methodology section for details. “Partisan Conflict and Congressional Outreach”

PEW RESEARCH CENTER

Congressional appeals to bipartisanship

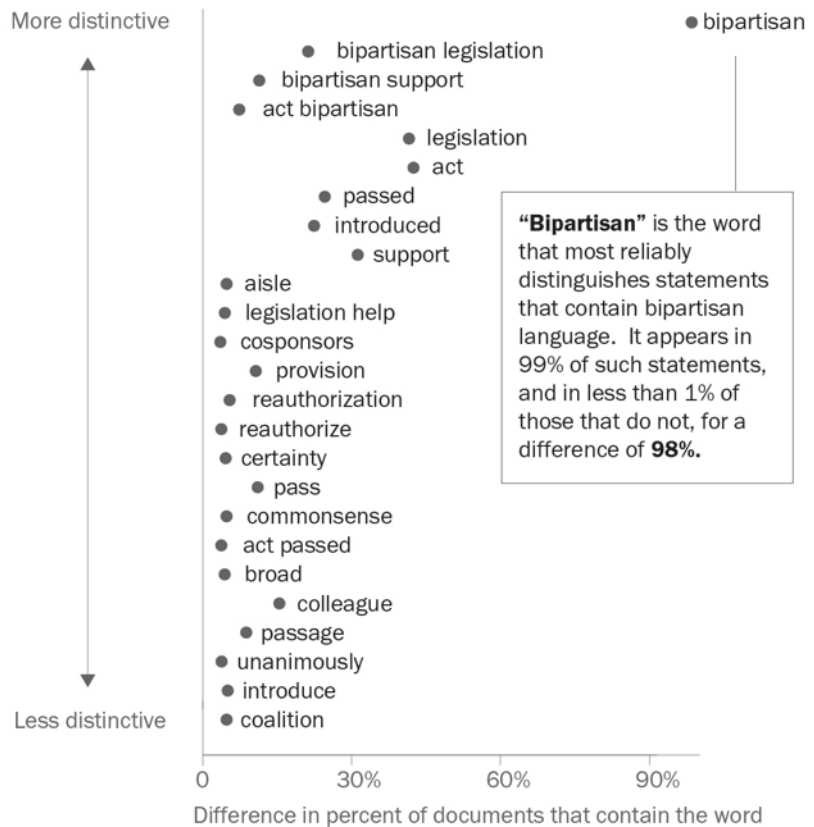
When elected officials highlighted bipartisanship, they emphasized the importance of working across party lines, irrespective of ideological differences. Members often released this messaging following their efforts to progress toward passage of bipartisan legislation.

Members most often explicitly used the term “bipartisan” in such communications.

However, occasionally they omitted that term, so only examining documents containing the term “bipartisan” would miss important cases. For example, Rep. Joseph Crowley (D-N.Y.) issued a press release discussing collaboration across the aisle: “Crowley, Serrano, and Curbelo, along with Reps. Ben Ray Lujan (D-N.M.) and Will Hurd (R-Texas), successfully led an amendment to the America COMPETES Reauthorization Act to require the National Science Foundation (NSF) to establish a grant program to support undergraduate STEM education at HSI’s.” And then-House Speaker John Boehner (R-Ohio) advocated for a bill by stating: “Now the SPACE Act boldly goes to the Senate, and we hope, the president’s desk, where it can begin a long life of helping our kids and grandkids prosper. Republicans and Democrats should come together to make it so.”

Words that distinguish bipartisanship

The 25 words most distinctive of press releases or Facebook posts that contain bipartisan language that contain bipartisanship, in rank order by pointwise mutual information



Source: Facebook OpenGraph API, Pew Research Center analysis of data from congressional websites and Lexis-Nexis. See Methodology section for details. “Partisan Conflict and Congressional Outreach”

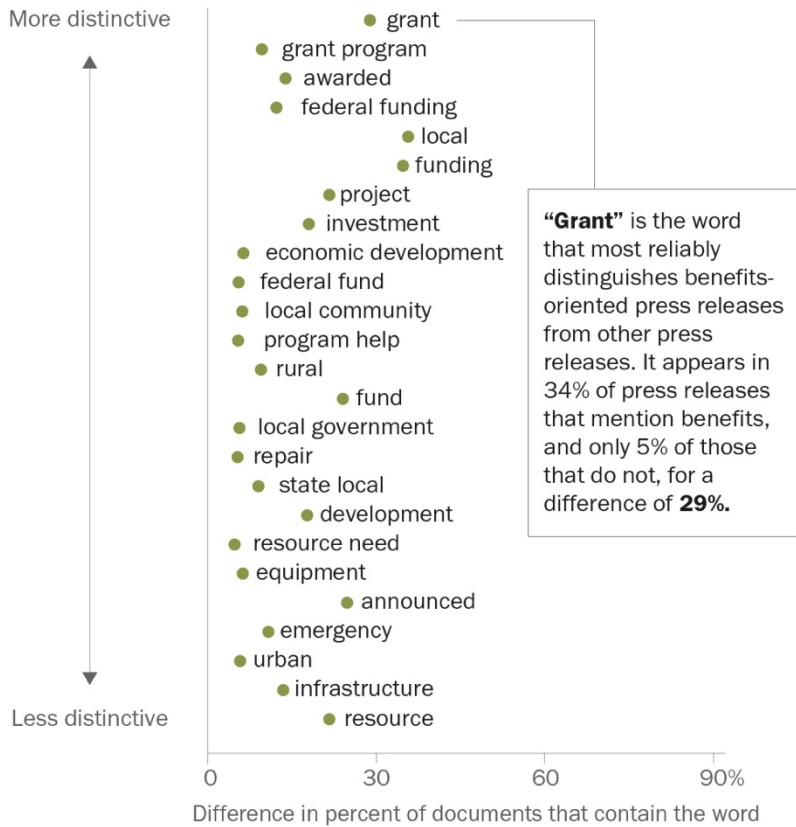
PEW RESEARCH CENTER

Constituent benefits: Grants and earmarks

When lawmakers discussed constituent benefits in press releases, their prose was marked by invocations of resources from the federal government, such as grant funding, local investments and other projects, especially related to economic development and emergency services. An example of one such announcement comes from the office of Rep. Mark Desaulnier (D-Calif.), which announced a federal job training program in a press release: “These federal funds will help train 51 students and place at least 45 graduates in environmental jobs including solar energy installation, hazardous waste removal, and construction.”

Words that distinguish constituent benefits

The 25 words most distinctive of press releases containing mention of constituent benefits, in rank order by pointwise mutual information



Source: Pew Research Center analysis of data from congressional websites and Lexis-Nexis. See Methodology section for details. “Partisan Conflict and Congressional Outreach”

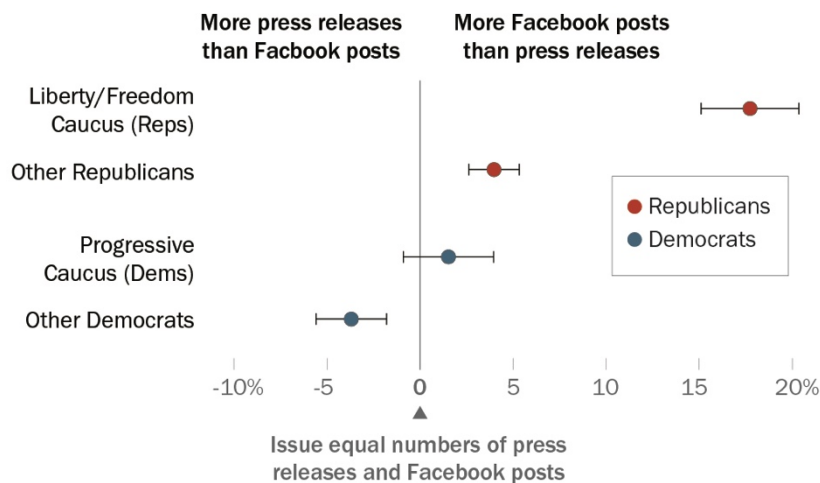
PEW RESEARCH CENTER

Republicans, Liberty and Freedom caucuses more reliant on social media

Congressional members of the Liberty and Freedom caucuses issued far fewer press releases than non-caucus Republicans on average, but posted more often than other members on Facebook. On the other side of the aisle, members of the Democratic Progressive Caucus were slightly more likely to favor Facebook posts to press releases than other Democrats. Overall, Democrats were on average more likely to issue press releases than Facebook posts.

Republicans, ideological caucus members use Facebook more than Democrats

Average share of outreach consisting of Facebook posts compared with press releases



Note: Lines indicate the standard error, an attempt to quantify the uncertainty surrounding each estimate.

Source: Pew Research Center analysis of data from congressional websites, Lexis-Nexis, and Facebook OpenGraph API. See Methodology section for details. "Partisan Conflict and Congressional Outreach"

PEW RESEARCH CENTER

Acknowledgments

This report is a collaborative effort based on the input and analysis of the following individuals:

Primary researchers

Solomon Messing, *Director, Data Labs*
 Patrick van Kessel, *Data Science Associate*
 Adam Hughes, *Research Associate*

Research team

Nick Judd, *Data Labs Fellow*
 Rachel Blum, *Data Labs Fellow*
 Brian Broderick, *DevOps Engineer*

Communications and editorial

Rachel Weisel, *Communications Manager*

Graphic design

Peter Bell, *Information Graphics Designer*

Copy editing and production

Travis Mitchell, *Digital Producer*

In addition, the report benefited from feedback provided by the following Pew Research Center experts:

Claudia Deane, *Vice President, Research*
 Michael Dimock, *President*
 Ruth Igielnik, *Research Associate*
 Galen Stocking, *Research Associate*
 Katerina Matsa, *Senior Researcher*
 Brad Jones, *Research Associate*
 Andrew Mercer, *Senior Methodologist*
 Amy Mitchell, *Director, Journalism Research*
 Jocelyn Kiley, *Associate Director, Research*

The following external scholars provided valuable feedback: Justin Grimmer, Sean J. Westwood, Doug Rice, and Andrew J. Clarke. Participants at the 2016 “New Directions in Analyzing Text as Data” conference also provided useful feedback. While the analysis was guided by our consultations with these advisers, the Pew Research Center is solely responsible for the interpretation and reporting of the data.

Methodology

Data collection

To analyze the content of congressional outreach, researchers gathered a comprehensive set of Facebook posts and press releases issued by offices of members of the U.S. Senate and House of Representatives between Jan. 1, 2015, and April 30, 2016. Facebook posts were obtained from the legislators' official pages (via API) and press releases were collected by combining two sources: official websites and an online database (LexisNexis). The research team analyzed the resulting corpus of posts and releases using a combination of traditional content analysis and machine-learning techniques.

Facebook posts from official accounts

Collecting Facebook page posts via API

In order to identify official Facebook accounts associated with members of the 114th Congress, researchers first recruited workers from [Mechanical Turk](#), an [online labor market](#). Researchers provided a list of U.S. senators and representatives in the 114th Congress to workers, who then searched for their Facebook accounts.¹⁹ The assignments were issued and completed on Feb. 1, 2016. A total of 59 Mechanical Turk workers assisted in the account identification process.

Researchers manually verified that accounts were associated with the correct Congress member for cases in which workers reported only a single Facebook account, or reported a single account for more than one politician. During this process, 17 accounts were found to be unofficial “fan pages” or parody accounts, and were removed from the list. Researchers also cross-referenced and supplemented account information with data from the open-source [@unitedstates project](#).

Most members of Congress maintain multiple social media accounts, one of which is often used as the member's “official” account. These official accounts are used to communicate information as part of the member's representational or legislative capacity, and House members may use official staff resources appropriated by the U.S. House or Senate. Personal and campaign accounts may not draw on these government resources under official House and Senate guidelines.²⁰

¹⁹ For more on the use of Mechanical Turk and other online labor markets for coding tasks, see Benoit, Kenneth, Drew Conway, Benjamin E. Lauderdale, Michael Laver, and Slava Mikhaylov. 2016. “Crowd-Sourced Text Analysis: Reproducible and Agile Production of Political Data” *American Political Science Review*.

²⁰ Straus, Jacob R. and Glassman, Matthew E. “Social Media in Congress: The Impact of Electronic Media on Member Communications.” Congressional Research Service, 2016.

Researchers classified Facebook accounts as official or unofficial based on the links to and from their official “.gov” pages. Accounts were classified as official if they met any of the following criteria: if the politician’s official house.gov or senate.gov homepage linked to the Facebook account, if the Facebook account’s profile contained a link to a house.gov or senate.gov homepage, or if it referenced or was referenced by a Twitter profile that met either of the two previous homepage linking conditions. Rules prohibit linking between official (.gov) and campaign websites or accounts, as well as linking from an official site or account to a personal site or account. But linking from a personal site or account to an official website or account is allowed. Hence, it is possible that some Facebook accounts classified as official were in fact personal accounts. However, researchers examined accounts manually and deemed this highly unlikely based on their content.

Accounts not classified as official according to this scheme were deemed unofficial accounts. Five Congress members had two pages that met these conditions. In these cases, posts from both pages were included in the analysis. Additionally, the 29 members that were initially determined to have no official accounts using this method were reviewed manually. During this process, six of them were identified as being official despite failing the above criteria, and were accordingly corrected.

After generating a list of official and unofficial congressional Facebook accounts, researchers used Facebook’s Open Graph Pages API to access all publicly available posts for each member’s page(s). These included plain-text status updates as well as updates with shared links, photos, and videos, and event announcements and invitations. The text from each post was saved as a distinct document, including the URL, title, preview excerpt and/or caption associated with any shared content (such as a link or video), as well as any comments made by the member. All posts used in this report were re-synced and updated via the API between June 14 and June 16, 2016, in order to ensure that the posts were as up-to-date as possible, and to ensure that likes and comments for posts created in late April were not artificially constrained by our sample timeframe. The share counts for each post were synced between Nov. 3 and Nov. 5, 2016.

The resulting dataset contains 162,609 Facebook posts from 978 different accounts belonging to 541 different members of Congress.²¹ A total of 523 of these accounts are official and belong to 518 different members, while 455 are unofficial, belonging to 449 different members. According to the official-unofficial classification strategy used here, 518 members of Congress (96%) who use Facebook for outreach had an official account, and 426 (79%) had at least one unofficial account in addition to their official ones. Furthermore, 92 members (17%) had only official accounts, and 23

²¹ Researchers included nonvoting members of Congress in the initial sample.

(4%) had unofficial campaign or personal accounts but no official accounts. Overall, 109,411 (67.3%) of the 162,609 posts were from official accounts.

The vast majority of the posts in the dataset are link shares (130,850, or 80.5%) or plain-text status updates (29,479, or 18.1%), with the remaining 1.4% comprised of miscellaneous other forms of posts. Photo and video posts were excluded from this analysis. Researchers also did not include any content from shared webpage links except for the caption and preview made available on Facebook. The findings presented in this report exclude posts by nonvoting representatives as well as members with fewer than 10 official Facebook posts, bringing the final number of Facebook posts to 108,235 posts from 509 members.

Press releases

Researchers obtained press releases issued by members of Congress in two ways: by downloading documents directly from the members' official websites, and by searching for and retrieving documents from LexisNexis.

Researchers collected a total of 96,064 press releases for 542 different members of Congress. All analyses here exclude members with fewer than 10 published press releases, bringing the sample to 538 legislators. We further excluded nonvoting members, bringing the total to 532 members and 94,521 individual releases.

Collecting press releases from official websites

To collect press releases from members' official websites, a list of homepage links was collected from the house.gov and senate.gov websites. In-house researchers and contracted assistants from Upwork manually examined each member's website, identified the "News Release" or "Press Release" section for each, and assessed the source code of the website to determine how the releases were organized, and where various attributes of each release such as the title, text and publication date could be found on each page. This information was then formalized into a set of Python scripts that used custom logic and parameters (i.e., XPath queries) to identify, parse, and store data for each representative's website.

Collecting press releases from LexisNexis

The press releases collected from members' official websites were supplemented with documents from LexisNexis. Using the LexisNexis Web Services Kit API, five different aggregate news collections were included in the search: US Fed News, Targeted News Service, US Official News, CQ Congressional Press Releases, and Government Publications and Documents. This query relied on a Boolean search designed to capture combinations of indicators of press releases.²² These parameters were the result of iterative testing that minimized false positives.

Researchers used a Python script to search for and download the raw text of documents published between January 2015 and April 2016. Researchers then used regular expressions to extract members' first and last names from key sections in each document that contained authorship information (see sidebar). Press releases issued by multiple members were attributed to all members listed.

Researchers used a fuzzy matching algorithm²³ to check each identified name against a list of members of Congress during the document's year of publication. Nine possible combinations of members' first, middle and last names, nicknames and suffixes were tested on each identified name, and a similarity ratio was computed for every pairing. If the top-matching name was found to have a similarity ratio higher than 70%, the corresponding member was

Name Matching

Politicians were identified using a set of four sections of text, combined with three regular expressions into a set of seven different search conditions. To compile the list of text sections to search for each document, the algorithm first scanned the raw document XML for metadata, which occasionally contained the names of members of Congress (1). This was then added to a list along with the article's title (2), the first 250 characters of the document's text (3), and the last 150 characters of the text (4). Three regular expressions were developed to search through these sections, available in Appendix A, section A2.

The four search sections and three regular expressions were then combined into search criteria representing patterns where politicians' names had been reliably identified during a process of manual review. In that process, none of the following combinations produced a false positive: 1a, 1b, 2a, 2b, 3a, 3b, 4c. If any names were extracted using these seven searches, the names were then used to search a list of Congress members for the document's year of publication.

²² The complete Boolean search is available in Appendix A, section A1.

²³ Text comparisons were made by computing the maximum value of the full and partial Levenshtein ratios, the latter of which removes the penalty for matches that vary greatly in length, such as when a word or phrase is a subset of a longer phrase that contains it along with additional information. Since names may take many forms depending on the context (a middle name may or may not be included, or a nickname may be used), computing partial ratios can often improve the fuzzy matching success rate.

then associated with the document and the press release was saved to the database.²⁴

To correct mismatches that may have occurred between politicians with similar names, random samples of 20 LexisNexis press releases were extracted for each politician and reviewed to confirm that the associated legislator was among the authoring politicians mentioned in the text of the release. A total of 40 sets of politicians were identified between whom press releases had been misattributed – many of which shared a common last name – and custom scripts were written and applied to correct the errors. Misattribution patterns were also created to filter out releases by state senators, congressional committees and state political parties.

While largely automated, this process was closely monitored and deletions were manually verified for every single member of Congress whenever more than 25% of their LexisNexis releases failed to meet the criteria.

Cleaning and deduplication

Since this dataset was derived from multiple sources, there was a high probability of duplicate releases. Deduplication required the removal of boilerplate content, which often existed in one source but not another, as well as the correction of errant dates.

Removing boilerplate content

The raw text of the press releases often contained boilerplate text, defined as sentences that did not contain any information related to the substantive topic(s) of the press release. For example, the websites of members of Congress often append contact information at the end of each release, and LexisNexis articles often contain publisher and copyright information at the top of each release. In order to accurately represent the substantive content of these documents and make comparisons between them, this extra information had to be removed.

For a subset of LexisNexis publishing outlets, there were several boilerplate patterns consistent enough to remove automatically using regular expressions when the documents were first collected. These patterns occurred consistently at the beginning (r'issued the following news release\:') or end (r'Copyright Targeted News Services', r'Read this original document at: http[\S]+', r'In case of any query regarding this article') of a press release, the text before or after which consisted of additional content added by the various wire services.

²⁴ An additional constraint also required that the absolute difference between the partial and full Levenshtein distance ratios was no greater than 0.4, in order to prevent matching on short names that may be subsets of longer names belonging to a different politician. While this preventative measure was likely unnecessary for matching LexisNexis names to our list of politicians, it has been useful in other text matching applications.

Every press release from LexisNexis was scanned for these patterns, and if any of these markers were found, boilerplate text was removed. During this process, unusual cases – where the text to be removed consisted of 250 or more characters or constituted greater than 40% of the entire document’s length – were reviewed manually to ensure that deviations from the expected boilerplate patterns did not result in erroneous text removal. A few such cases were encountered, all of which were short press releases that fit the expected pattern. Similarly, while website-based press release boilerplate was less consistent, one footer pattern that indicated the end of an official release was frequent enough across multiple websites to remove in the same manner as those above (r`###`).

However, most instances of bad content were too inconsistent and diverse to be removed with a small set of fixed rules.

To implement a more flexible approach, researchers developed a parsing algorithm to split press release content into sentences, based on the [NLTK Punkt English tokenizer](#). The tokenizer was expanded to use the following regular expressions to mark the beginning of sentences, which helped identify boilerplate sentences.²⁵

Next, a machine-learning approach was employed to classify sentences as boilerplate or not. First, a sample of 4,516 documents across 634 different members of Congress (and the sentences therein) was extracted for manual review and verification.²⁶ Examples of boilerplate sentences include:

- Read more here: <http://www.newsobserver.com/2014/02/04/3590534/rep-mcintyre-of-nc-laments-too.html>
- District Office Washington Office District Office 1717 Langhorne Newtown Rd.
- Your browser does not support iframes.
- Welcome to the on-line office for Congressman Donald Payne, Jr.
- ">'); document.write(addy71707); document.write('<\a>'); //-->kdcr@dordt.edu');//-->
- You can reach Rep. Beyer on the web at www.beyer.house.gov , on Facebook at [facebook.com/RepDonBeyer](https://www.facebook.com/RepDonBeyer) and on Twitter @RepDonBeyer.

²⁵ See Appendix A, section A3.

²⁶ One to twenty press releases were extracted for each member of Congress, proportional to their total press release count (seven documents per politician, on average), resulting in a set of 4,516 documents across 634 different members. After each document was split into sentences, each sentence was assigned an ordered percentile ranking based on its distance from the top and bottom of the document. Sentences were then assessed manually, and flagged if they were boilerplate and should be removed. Typically this consisted of contact information, erroneous pieces of source code, welcome statements and generic descriptions of the member that were included at the end of each of their releases.

- Washington, DC Office 2417 Rayburn HOB Washington, DC 20515 Phone: (202) 225-2331 Fax: (202) 225-6475 Hours: M-F 9AM-5PM EST
- He represents California's 29th Congressional District, which includes the communities of Alhambra, Altadena, Burbank, East Pasadena, East San Gabriel, Glendale, Monterey Park, Pasadena, San Gabriel, South Pasadena and Temple City.

Of the 98,057 sentences that were assessed, 6% (6,089) were identified as boilerplate. A classification algorithm²⁷ was then trained on this dataset, using both each sentence's location in its originating document, as well as character²⁸ and word²⁹ features to make the determination. Finally, a custom set of binary regular expression pattern flags were added as additional features, based on a list of indicators that researchers identified to be commonly associated with boilerplate sentences.³⁰

The resulting model correctly identified known boilerplate sentences 88% of the time (recall), and the sentences it flagged agreed with researcher labels 86% of the time (precision). This level of accuracy was sufficient to facilitate deduplication.

Of the final set of documents used in this analysis, 65.8% had some of their text removed during this process, across 542 different members of Congress.³¹ Among those that were modified, 6.6% of the text was removed on average,³² with a median of 1.2%.

Fixing incorrect dates

As a first pass in correcting errant dates, researchers reviewed outliers. A number of dates stood out on which a highly disproportionate number of press releases appeared to have been published. Researchers iterated over all unique politician, source (LexisNexis or website), and date combinations, and examined any that had more than ten press releases issued by a single politician on a single day, all from the same source. Researchers found 41 such combinations for the timeframe covered by this report. For each combination, researchers then randomly selected three documents for manual review. When an incorrect date was found, researchers attempted to determine the nature of the error and codify a rule for making a correction.

²⁷ Linear support vector classifier, using 5-fold cross-validation, scored on precision, and evaluated on a 25% hold-out

²⁸ TF-IDF n-grams of 2-6 characters, minimum document frequency of 5, maximum document frequency of 95%

²⁹ TF-IDF n-grams of 1-4 words, minimum document frequency of 5, maximum document frequency of 95%

³⁰ Reported in Appendix A, section A4.

³¹ 63,239 of the 96,064 final, de-duplicated press releases had text different than that of the original source document; overall numbers prior to de-duplication are not available.

³² Based on total characters.

Correcting dates in press releases collected from member websites

Several systematic inaccuracies found in press releases collected directly from members' websites resulted in errant dates being associated with each release. The process used to identify and remedy these issues is detailed below.

Releases for four politicians had systematic time-stamp errors, but all four included the correct dates in the first sentence of each press release. Researchers extracted the correct dates using regular expressions.³³

The most common cause of errant dates was “archive dumping.” Archive dumping occurs when a large collection of older press releases from a prior version of a politician’s website was transferred to their latest website, and dated by the transfer date rather than the original publication date. This affected 16 politicians, spanning 25 different dates. For 21 of those dates, press release dates were cleared entirely. For the remaining four dates, all press releases occurring on or before that date were cleared. There were systematic errors for prior dates as well. However, the vast majority of these cases occurred prior to the timeframe in this report – only 162 of the press releases used in this study were impacted by these issues (151 dates were cleared, and 11 were corrected). As a precautionary measure, dates were cleared out for all of the press releases that occurred on the first available date for each politician, in order to avoid any remaining archive dumps that may have been missed.

LexisNexis dates did not appear to have significant errors. In some cases the LexisNexis dates lagged behind those on the website, however, this did not hamper the deduplication process.

Identifying and removing duplicates

In order to compare documents against each other and consolidate copies of press releases, documents were first grouped by their associated politician. Each politician’s set of press releases was then parsed, tokenized and represented as a term frequency-inverse document frequency (TF-IDF) matrix. Documents were compared against all other documents, and any pair with a cosine similarity of at least 0.70 were grouped together as potential duplicates. For each such group, each press release was compared against the others once again and the set was then grouped into even smaller subsets based on whether or not a pair of documents had a Levenshtein distance ratio of at least 0.70.

³³ For example, one of these politicians happened to have different webpage versions of the same press releases, one format of which did not include the dates; to resolve this issue, duplicates were identified using a 99% cosine similarity match and dates were then consolidated. Another politician’s website mistakenly recorded all 2010 press releases as being issued in 2001.

A number of issues necessitate additional processing to detect duplicate press releases. First, politicians occasionally reuse content across their press releases – using boilerplate press releases for announcements related to a particular topic area, for example, where only a few words might change between a press release announcing the introduction of a bill and a release announcing its passage several months later. Announcements of grant funding are another common pattern, in which an entire press release will essentially be copied-and-pasted multiple times with the dollar amount and recipient being the only substantive variations.

Hence, two nearly identical press releases may be similar, but still distinct. Dates can often be helpful in making this determination, but a pair of duplicates collected from different sources might not have exact date matches. A release may have taken a day or two to make its way from a politician’s press office, to a wire service, and then to LexisNexis.

After a manual examination of potential remaining duplicates, it was clear that a machine-learning approach would be necessary to identify remaining duplicates for removal. A random forest classifier was trained on the coded sample, scored on recall, and assessed on a 25% hold-out from the sample.³⁴ Features included: full and partial Levenshtein ratios for a pair of documents, and for each metric, the difference between the two; indicators of whether the two documents had the same day, month, and day/month combination; the total difference between the dates in days; whether or not the documents were from the same source; for each source, whether at least one of the two documents was from that source; whether the documents’ text matched completely; and the length of the shortest document, the longest document and the character difference in text length between the two. Additional regular expressions that might indicate false positives were also included, representing keywords that tend to vary in otherwise identical releases.³⁵ The resulting model achieved 93% precision and 92% recall in identifying false positives (non-duplicates), and 96% precision and 97% recall in identifying true positives (correct duplicates).

This approach was used with every set of politicians’ press releases and press release pairs that were identified as duplicates were consolidated as a single document. If one press release had more complete metadata than another, the missing fields were filled in. If the two press releases had a date conflict, the earlier date was chosen as the correct one. This occurred frequently when website press releases were matched to LexisNexis duplicates, the latter of which often had dates trailing the former by several days.³⁶

³⁴ The random forest classifier outperformed approaches based on support vector machines.

³⁵ Regular expressions appear in Appendix A, section A5.

³⁶ This is presumably because releases can be posted online immediately, but may take time to be picked up by news wires.

Prior to deduplication, 87,362 press releases were collected from congressional webpages, and an additional 234,534 were collected from LexisNexis. This resulted in a total of 321,896 press releases from both sources. During deduplication, 5,119 duplicates were found among the webpage-based press releases themselves, an additional 145,485 duplicates were found within the LexisNexis press releases themselves, and 75,248 LexisNexis press releases were dropped in favor of duplicates found amongst the webpage versions. A total of 225,852 press releases were removed, resulting in a final collection of 96,064 remaining press releases. In all, 85% of the press releases collected from webpages were also found in LexisNexis, and 92% of the press releases collected from LexisNexis were also found on congressional webpages. A random sample of 50 remaining LexisNexis articles were examined by researchers to confirm that they did not represent missed duplicates. All articles reviewed were determined to be press release content that was missed during the website data collection process, most commonly because a representative had posted the release in a different section of their website, separate from other press releases.

Other data sources

Researchers also used two datasets containing contextual information about individual legislators: DW-NOMINATE ideology estimates and 2012 presidential election results. Ideology estimates are publicly available at Voteview.org. Researchers obtained election results at the level of states and congressional districts in the replication materials for Gary Jacobson's "It's Nothing Personal: The Decline of the Incumbency Advantage in U.S. House Elections," published in the *Journal of Politics*. The replication materials are also [available online](#).

Content analysis

Codebook

After finalizing the collected data, researchers developed a codebook or classification scheme for human analysts to classify documents. This codebook differs from most in that all variables are binary, indicating the presence or absence of the content in question.

To assess political disagreement, the following measure was coded: “Does the document express opposition toward or disagreement with any of the following?” followed by a list of political targets. The first, President Obama or his administration, was defined as: “The president himself or one of his own decisions and actions, or those of his administration.” Coders were instructed to look for the terms “Obama,” “The President,” “the or this Administration,” “The White House,” and “executive,” in the context of actions or decisions. Coders were instructed to avoid policy terms like “Obamacare” and mentions of federal agencies that were not explicitly linked with the president.

The second political target was “Democrats (other than Obama), or ‘liberals’ in general.” For that item, coders were told to look for the words “Democrats,” “DNC,” “liberals,” “left-wing,” “progressives” and any elected official identified as a Democrat, but to avoid references to the president, unions, activist groups like Occupy Wall Street, and other organizations that are not formally linked with the party. Third, coders identified disagreements with “Republicans, or ‘conservatives’ in general” by looking for the terms “Republican,” “RNC,” “conservative,” “right-wing,” “Tea Party” and any elected official described as a Republican. Coders were instructed to avoid coding religious groups, advocacy groups or general ideologies such as “conservatism” as Republican targets, unless the document explicitly linked them to Republican politicians.

To assess political indignation, researchers instructed coders to identify whether “the document expresses any indignation or anger,” by evaluating strong adjectives, negative emotional language and resentment. Coders were instructed to look for language that “expresses a degree of anger, resentment, or indignation,” alongside conflictual terms like “fight” and “attack,” as well as moral imperative statements. Coders were told to avoid coding disagreement that lacked emotional rhetoric as indignation. Researchers provided a list of examples of indignation, such as “President Obama’s foreign policy is irresponsible, shortsighted and DANGEROUS” and “Instead of being honest and upfront about their goals, the Republicans have used a number of budgetary gimmicks to cover-up the devastating impact that their budget will have on the lives of ordinary Americans ...” Examples of non-indignant text included: “discrimination has no place in our federal government” and “I have serious concerns those requirements were not met.”

Bipartisanship references were identified as cases in which “the document mentions one or more members of both parties agreeing on something or working together in a bipartisan manner.” Specifically, researchers described how the text must “mention some way in which Republicans or Democrats have agreed, worked together, or compromised, are currently doing so, or will be in the future,” and that statements opposed to bipartisanship did not qualify. Examples of bipartisanship included: “Both Republicans and Democrats agree: the President must work with Congress on any Iran deal,” “The only reason that Mr. Boehner was able to get his two-year deal in the first place was with overwhelming Democratic support. It would be unwise to wander too far from that bipartisan framework,” and “President Obama will veto a bipartisan defense bill to coerce more domestic spending.”

To capture statements about government benefits, coders were instructed to determine whether “the document mentions a specific benefit for people in the district.” These references could include “current or future, real or hypothetical benefits for a local constituency, often regional or local investment,” but they could not be announcements related to local meetings or town halls unless they were related specifically to a benefit. As examples of references to benefits, coders read the following: “I am thrilled to join with College Possible and College Possible Philadelphia in celebrating their \$3 million innovation grant award from the U.S. Department of Education” and “My vote to protect Joint Base McGuire-Dix-Lakehurst and our national defense programs... Sources have said the loss of the tankers would be devastating to the joint base, which is Burlington County’s largest employer.” Coders were instructed to avoid vague references to benefits, such as “Thanks to the Northmoreland Township Volunteer Fire Company for the opportunity to tour its headquarters and discuss my legislation to protect local fire companies from Obamacare with its members.”

The complete classification list follows:

- Does the document mention a specific benefit for people in the district?
 - Does the document discuss a foreign policy or international issue?
 - Does the document mention one or more members of both parties agreeing on something or working together in a bipartisan manner?
 - Does the document describe the U.S. political system/process or government as broken, corrupt, wasteful, or dysfunctional?
 - Does the document express any indignation or anger?
 - Does the document express opposition toward or disagreement with any of the following?
(Select all that apply)
- President Obama or his administration

- Democrats (other than Obama), or “liberals” in general
- Republicans, or “conservatives” in general
- A specific federal agency or regulatory group
- A court or court decision
- An advocacy group, union, industry, or private company
- The media or a news outlet

Document samples for hand coding

Because exploratory random sampling showed that disagreement and indignation were relatively rare, a set of keywords representing various political targets were used to over-sample documents more likely to contain attacks or criticism to ensure that enough positive cases would be present for machine learning. This was accomplished using four different regular expressions:

Democrats	[L]iberal* [D]emocrat* [Pp]rogressiv*
Republicans	[C]onservative* [Rr]epublican* \bGOP\b \bgop\b Tea Party* Freedom Caucus
President Obama	[Oo]bama ([Tt]he [Tt]his [Hh]is)([a-z\s\W]{0,20})[Aa]dministration White House President\'s desk ([Tt]he [Oo]ur [Tt]his) [Pp]resident*
Misc. other	[Aa]genc* [Bb]ureau* [Dd]epartment* bias media court* justice* judge* decision special interest* corrupt* [A-Z]{3,5} district* industr*

Random samples of documents were then collected, which were stratified across pattern-matched content categories, members of Congress, and party. Oversampling low-incidence strata increased the number of positive cases available for training, which resulted in substantially higher model accuracy (particularly for recall).

	Sample size	Purpose	Oversampling
Facebook	3,000	Training	20% Democrats, 40% Republicans, 20% Obama, 20% Random
	1,000	Training	60% Democrats, 15% Republicans, 25% Misc. Other
	200	Validation	20% Democrats, 40% Republicans, 20% Obama, 20% Random
Press releases	2,000	Training	20% Democrats, 40% Republicans, 20% Obama, 20% Random
	1,000	Training	60% Democrats, 15% Republicans, 25% Misc. Other
	200	Validation	20% Democrats, 40% Republicans, 20% Obama, 20% Random

Crowd-sourced document classification

To classify the content of Facebook posts and press releases, researchers employed coders from Amazon’s Mechanical Turk. Only Mechanical Turk workers who earned Amazon’s “Masters” certification and had an account associated with an address in the United States were eligible to classify the documents. In all, 3,200 press releases and 4,200 Facebook posts were each coded by five unique coders between July 8 and Aug. 22, 2016. Coders were paid \$0.30 per press release and \$0.20 per Facebook post. The median worker submission took 51 seconds to classify a Facebook post, and 79 seconds to classify a press release. The median hourly rate for online workers was \$13.67 for press releases and \$14.12 for Facebook posts.

Each Mechanical Turk worker who agreed to code a set of documents was also asked to complete a brief survey of their political views and basic demographic information in order to qualify for the work. After completing the qualification survey, coders then used a custom online coding interface embedded within the Amazon Mechanical Turk assignments to classify each Facebook post and press release. The creation of coding assignments, as well as the process of downloading and compiling the results, was accomplished using the Amazon API directly.

Validating the crowd-sourced data

In order to assess the validity of the crowd-sourced document classifications, two researchers manually coded the same 200 press releases and 200 Facebook posts that a set of Mechanical Turk coders also classified for comparison.³⁷ Internal agreement between Pew Research Center’s two experts was then assessed, after which the coders reviewed cases where they disagreed to arrive at a consensus. Accordingly, a single consolidated expert code for each document was generated and compared to the Mechanical Turk workers’ codes (aggregated as described below) and the final machine learned codes.

Researchers found little evidence that coding decisions were related to coder partisanship or ideological identification. Both attributes of coders had no substantively or statistically significant relationship with the classification decisions we examine here.

³⁷ During subsequent number checking and data verification steps, additional date corrections and more aggressive de-duplication steps were found to be necessary (described earlier). As a result, some training and validation cases were consolidated or removed from the final samples prior to model training and evaluation. A handful of documents were also sampled into both the training and validation sets, and these cases were excluded when evaluating the models against the Mechanical Turk and expert coders. This resulted in a final total of 2,448 training and 159 validation press releases, and 3,999 training and 183 validation Facebook posts.

In order to aggregate all five Mechanical Turk workers' responses into a single set of codes, for each variable, researchers computed a threshold for the number necessary to infer that the document contained the content in question. This was accomplished in via the following steps.

First, using the internal codes arrived at by the Center's experts as a comparative baseline, researchers selected candidate thresholds for aggregating the crowd-sourced codes (i.e. the proportion of Turk coders required to consider a case "positive") that either maximized agreement for each indicator variable (Cohen's Kappa), or produced agreement that was within the standard error of the maximum. This was done for each indicator variable for Facebook posts and press releases separately.

Second, to maximize the comparability of the key measures across both Facebook posts and press releases, for each variable, any candidate thresholds that were not shared between both document types were discarded from consideration. This resulted in a single threshold for each key variable, with the exception of the three disagreement items. Since these disagreement items were similar in nature and were to be compared and grouped together during analysis, researchers selected the single remaining threshold shared in common across these three measures.

After finalizing the consolidation thresholds for each classification variable, the selected cutoffs were used to aggregate the Mechanical Turk codes for each document in the (full) training sample for both press releases and Facebook posts.

To produce the variables used in the actual report, researchers first created a "disagreement" variable that was positive if any of the three disagreement categories (disagreement with the president, disagreement with Democrats, disagreement with Republicans) was positive (indicating its presence in the document). Researchers then created an "indignant disagreement" variable that was positive if both indignation and disagreement were positive.

To produce a single measure of inter-rater reliability for all measures used in the report, researchers concatenated all codes for each document using the consolidated categories for internal coders and comparing to the consolidated categories for Mechanical Turk workers. For Facebook posts, the combined Cohen's Kappa was 0.78; that number was 0.61 for press releases. The higher level of accuracy for Facebook is a result of excluding the "constituent benefit" category, whose incidence on Facebook was too infrequent to include in the analysis. Cohen's Kappa between the Center's expert two coders was 0.59 for Facebook and 0.57 for press releases. Levels of agreement for specific categories are provided in the table below.

Document	Category	MTurk threshold	Expert-Turk Kappa	Expert Kappa
Press releases	Bipartisanship	2/5	0.79	0.76
	Benefits	3/5	0.50	0.65
	Disagreement	2/5	0.92	0.83
	Indig. disagreement	2/5	0.71	0.46
Facebook posts	Bipartisanship	2/5	0.91	0.76
	Benefits	3/5	0.58	0.66
	Disagreement	2/5	0.89	0.87
	Indig. disagreement	2/5	0.80	0.45

Cleaning the text

To produce a dataset useful for machine learning, the text of each document was converted into a set of features, representing words and phrases. To accomplish this, each document was passed through a series of pre-processing functions (on top of the cleaning and deduplication steps outlined above). First, to avoid including words that could bias the machine-learning models towards particular politicians or districts, a set of custom stopword lists were used to filter out names and other proper nouns, comprised of the following:

- A list of 318 common English stopwords, taken from the [Glasgow Information Retrieval Group](#)
- A list of 9770 first and last names, taken from a Pew Research Center database of 13,942 current and historical politicians, and filtered using WordNet
- A list of 896 state names and state identifiers (e.g. “West Virginian”, “Texan”)
- A list of 18,128 city and county names, taken from a Pew Research Center database of geocoded campaign contributions, and filtered using WordNet
- A list of 24 month names and abbreviations
- 331 additional stopwords, manually identified through a process of iterative review by Center researchers

Some people and locations have names that are also common English words, some of which are used far more frequently as the latter. To avoid unnecessarily excluding these words from our training data, potential stopwords were assessed using WordNet, which provides information on a word’s alternative definitions and where they fall on a spectrum of generality to specificity (using a hyponymy taxonomy). If a word met two or more of the following criteria, it was flagged as being

common and/or versatile enough to be included in the training data, and was removed from the stopword list:

- The word has more than two different definitions (synsets)
- One or more of the word's definitions (synsets) had a variation (lemma) with a depth of less than five (indicating generality)
- One or more of the word's definitions (synsets) had at least two variations (lemmas)

After combining all of these lists into a single set of stopwords, an additional 66 words were removed, based on a list compiled by Center researchers during a process of manual iterative review. This ultimately resulted in a list of 27,579 stopwords that were removed from the text of all documents. After removing stopwords, the text of each document was lowercased, and URLs and links were removed using a regular expression.³⁸ Common contractions were expanded into their constituent words, punctuation was removed and each sentence was tokenized using the resulting whitespace. Finally, words were lemmatized (reduced to their semantic root form), and filtered to those containing three or more characters.

Extracting features

Machine-learning models were trained on term-frequency inverse-document frequency matrices, containing 1 to 4 grams with a minimum document frequency of 10 and maximum document proportion of 90%. In addition to a matrix constructed from each document in its entirety, a subset of classifiers also utilized matrices produced from a subset of sentences that contained keywords relevant to the classification variable, identified using regular expressions:

- **Bipartisanship:** sentences that contained a match on `[Bb]ipartisan*`
- **Disagreement with Obama:** sentences that matched on the Obama regular expression used for oversampling
- **Disagreement with Democrats:** sentences that matched on the Democrat regular expression used for oversampling
- **Disagreement with Republicans:** sentences that matched on the Republican regular expression used for oversampling
- **Constituent benefits:** sentence that contained the name of a state, or matched on `[Ll]ocal|[Dd]istrict*|\$([0-9]{1,3}(:?(:\.,[0-9]{3})+)?(?:\.[0-9]{1,2})?)\s`

³⁸ See Appendix A, section A6.

Additionally, the constituent benefit classifier also utilized counts extracted from the text using the following three regular expressions, representing common patterns related to monetary amounts:

- `r'\$([0-9]{1,3}(?:\,[0-9]{3})+)?(?:\.[0-9]{1,2})?\s'`
- `r'\$[0-9]{1,3}(\,[0-9]{3})+\s'`
- `"thousand*|million*|billion*|hundred*"`

Model training

Machine-learning algorithms were then used to classify the entire scope of documents used in this report based on these training results from human coders.³⁹ The relevant classification models, called Support Vector Machines (SVMs), use information about the known relationships between particular words and classification decisions in human-coded texts (training data) in order to classify texts that are not already coded (test data). The use of machine-learning models mitigates the cost and time that would be required to use human-based methods to classify the entire range of Facebook posts and press releases.

SVMs were trained on these features for each classification variable and document type, hyper-parameterized over different penalty levels⁴⁰ and kernels⁴¹ with 5-fold [cross-validation](#). The best parameters were selected using Matthew's correlation coefficient as the scoring function. Training data was weighted and evaluated using traditional sample weighting based on the population proportions of over-sampled keywords (their true value across the range of documents to be studied in this report). Additional weights were used only during the model training process (not during evaluation), weighting cases using the inverse proportion of their class and balanced across party lines using weighted proportions (artificially adjusted to 50%) so that each political party was given equal consideration by the algorithm.⁴²

After training, the models were evaluated by comparing them to the 200 validation cases for their respective document type, as well as to their 5-fold cross-validation averages.⁴³ The following table shows the performance for each model:

³⁹ See Grimmer, Justin and Brandon Stewart. 2013. "Text as Data: The Promises and Pitfalls of Automatic Text Analysis," *Political Analysis*; Hopkins, Daniel J. and Gary King. 2010. "A Method of Automated Nonparametric Content Analysis for Social Science," *American Journal of Political Science*.

⁴⁰ C=1, 10, 100, 1000

⁴¹ Linear and radial basis function kernels

⁴² This does not apply to the three disagreement variables, which were trained on the subset of documents associated with the opposing party.

⁴³ As described in Footnote 38, there were 159 validation cases for press releases and 183 for Facebook posts; for party-specific measures (such as disagreement), researchers evaluated only documents for the relevant political party.

Document	Category	Accuracy	Precision	Recall
Press releases	Bipartisanship	0.91	0.87	0.77
	Benefits	0.82	0.56	0.55
	Disagreement	0.93	0.80	0.82
	Indig. disagreement	0.92	0.57	0.59
Facebook posts	Bipartisanship	0.98	0.87	0.76
	Benefits	0.94	0.36	0.25
	Disagreement	0.94	0.79	0.81
	Indig. disagreement	0.92	0.59	0.55

Reliability statistics comparing the hand-coded data to these machine-learned measures follows:

Document	Category	Model-Expert Kappa	Model-Turk Kappa
Press releases	Bipartisanship	0.77	0.65
	Benefits	0.56	0.51
	Disagreement	0.84	0.78
	Indig. disagreement	0.69	0.57
Facebook posts	Bipartisanship	0.72	0.64
	Benefits	0.13	0.24
	Disagreement	0.75	0.76
	Indig. disagreement	0.54	0.50

The final resulting population means for the model-based measures used in this report are reported below, compared against the sample means from Mechanical Turk.

Party	Variable	Facebook		Press Releases	
		Machine estimate	Coder estimate	Machine estimate	Coder estimate
Democrats	Disagreement	6.0%	6.7%	9.8%	9.7%
Democrats	Indignant disagreement	3.0%	4.2%	3.9%	5.0%
Democrats	Bipartisanship	5.5%	5.1%	22.1%	26.8%
Democrats	Constituent benefits	*	*	26.8%	23.7%
Republicans	Disagreement	20.5%	19.5%	27.9%	27.9%
Republicans	Indignant disagreement	12.6%	11.4%	14.7%	14.0%
Republicans	Bipartisanship	5.2%	5.1%	22.5%	26.7%
Republicans	Constituent benefits	*	*	14.7%	16.0%

* Since constituent benefits-related Facebook posts were rare and difficult to classify, we excluded them from this and all other analyses.

Overall, the model estimates of disagreement, indignant disagreement, bipartisanship and constituent benefits are comparable to those observed in the sample of documents classified by Mechanical Turk workers.

Multivariate analysis of Facebook engagement metrics

Random effects analysis of Facebook engagement

This Facebook-post level analysis shows that disagreement and indignant disagreement systematically predict additional Facebook likes, comments and shares. The number of likes, comments and shares used in this analysis is logged (base 10) due to the skewed distribution of each measure. The model includes an intercept for each member of Congress, which helps account for the underlying popularity of a given member's posts.

	Facebook likes	Facebook comments	Facebook shares
Disagreement	0.155** (0.006)	0.310** (0.006)	0.207** (0.006)
Indignant disagreement	0.166** (0.007)	0.171** (0.007)	0.178** (0.008)
Republican	-0.155** (0.043)	0.071 (0.038)	-0.075 (0.040)
Member of Congress random effects	√	√	√
Observations	107,977	107,977	104,948
Log likelihood	-66,541	-66,550	-70,759
Akaike inf. criterion	133,093	133,112	141,530
Bayesian inf. criterion	133,150	133,170	141,587

Fixed effects analysis of Facebook engagement

This Facebook-post level analysis verifies that disagreement and indignant disagreement systematically predict additional Facebook likes, comments and shares. The number of likes, comments and shares used in this analysis is logged (base 10) due to the skewed distribution of each measure. The model includes a dummy variable, or fixed effect, for each member of Congress. This is another way to account for the underlying popularity of a member's posts.

	Facebook likes	Facebook comments	Facebook shares
Disagreement	0.155** (0.006)	0.310** (0.006)	0.206** (0.006)
Indignant disagreement	0.166** (0.007)	0.171** (0.007)	0.178** (0.008)
Republican	-0.579** (0.098)	-0.453** (0.098)	-0.382** (0.104)
Member of Congress fixed effects	√	√	√
Observations	107,977	107,977	104,948
R²	0.55	0.52	0.49
Adj. R²	0.54	0.52	0.49

Multivariate analysis of communications content

Ordinary least squares: Predicting the proportion of disagreement

This member of Congress-level analysis shows the attributes of individual elected officials that are associated with a higher proportion of expressed disagreement in press releases and Facebook posts. Ideological distance from the midpoint of DW-NOMINATE and district competitiveness are strongly associated with a higher proportion of disagreement.

	Disagreement in:			
	Press releases	Facebook posts	Press releases	Facebook posts
DW-NOMINATE	-0.005 (0.033)	0.033 (0.030)		
DW-NOMINATE²	0.286** (0.035)	0.195** (0.031)		
2012 Romney vote			-0.209** (0.059)	-0.119** (0.052)
2012 Romney vote × Republican			0.747** (0.095)	0.600** (0.082)
Republican	0.162** (0.028)	0.112** (0.025)	-0.201** (0.048)	-0.162** (0.041)
House member	0.015 (0.011)	0.008 (0.010)	-0.002 (0.012)	-0.007 (0.010)
Years in Congress	0.017 (0.011)	0.025** (0.009)	0.004 (0.011)	0.015 (0.009)
Party leader	0.157** (0.030)	0.176** (0.026)	0.148** (0.030)	0.168** (0.026)
Ranking member or chair	0.019 (0.012)	0.006 (0.011)	0.010 (0.013)	-0.003 (0.011)
Constant	0.019 (0.020)	0.007 (0.018)	0.166** (0.030)	0.090** (0.026)
Observations	532	509	532	509
R²	0.55	0.55	0.53	0.55
Adj. R²	0.54	0.54	0.53	0.54

Ordinary least squares: Predicting the proportion of indignant disagreement

This member of Congress-level analysis shows the attributes of individual elected officials that are associated with a higher proportion of expressed indignant disagreement in press releases and Facebook posts. The patterns largely match those for expressions of disagreement without indignation.

	Indignant disagreement in:			
	Press releases	Facebook posts	Press releases	Facebook posts
DW-NOMINATE	0.056* (0.023)	0.039 (0.023)		
DW-NOMINATE²	0.182** (0.024)	0.131** (0.024)		
2012 Romney vote			-0.100* (0.040)	-0.065 (0.039)
2012 Romney vote × Republican			0.558** (0.065)	0.432** (0.063)
Republican	0.052** (0.020)	0.059** (0.019)	-0.184** (0.033)	-0.131** (0.031)
House member	0.011 (0.008)	0.001 (0.008)	-0.003 (0.008)	-0.010 (0.008)
Years in congress	0.013 (0.007)	0.013 (0.007)	0.002 (0.008)	0.005 (0.007)
Party leader	0.097** (0.021)	0.117** (0.020)	0.091** (0.021)	0.111** (0.020)
Ranking member or chair	-0.001 (0.009)	-0.002 (0.009)	-0.008 (0.009)	-0.008 (0.009)
Constant	0.008 (0.014)	0.009 (0.014)	0.073** (0.020)	0.055** (0.020)
Observations	532	509	532	509
R²	0.52	0.48	0.51	0.48
Adj. R²	0.51	0.47	0.50	0.48

Ordinary least squares: Predicting the proportion of bipartisanship

This member of Congress-level analysis shows the attributes of individual elected officials that are associated with a higher proportion of bipartisan references in press releases and Facebook posts. Ideological moderation is strongly associated with bipartisanship, and House members appear less likely to focus on bipartisanship than their Senate colleagues.

	Bipartisanship in:			
	Press releases	Facebook posts	Press releases	Facebook posts
DW-NOMINATE	0.056 (0.035)	0.023 (0.016)		
DW-NOMINATE²	-0.347** (0.037)	-0.117** (0.017)		
2012 Romney vote			0.394** (0.061)	0.145** (0.028)
2012 Romney vote × Republican			-0.950** (0.099)	-0.335** (0.045)
Republican	-0.020 (0.030)	-0.005 (0.014)	0.455** (0.050)	0.164** (0.022)
House member	-0.064** (0.012)	-0.025** (0.006)	-0.040** (0.012)	-0.016* (0.006)
Years in Congress	-0.014 (0.011)	-0.005 (0.005)	0.003 (0.012)	0.001 (0.005)
Party leader	0.034 (0.032)	0.012 (0.014)	0.045 (0.032)	0.016 (0.014)
Ranking member or chair	-0.014 (0.013)	0.007 (0.006)	-0.006 (0.013)	0.010 (0.006)
Constant	0.350** (0.022)	0.105** (0.010)	0.101** (0.031)	0.015 (0.014)
Observations	532	509	532	509
R²	0.21	0.15	0.20	0.15
Adj. R²	0.20	0.14	0.19	0.14

Appendix A

A1

(“office of” W/s (senator OR sen OR rep OR representative) W/s (released OR issued) W/s following W/s (statement OR release)) OR “U.S. SENATE DOCUMENTS” OR “U.S. HOUSE OF REPRESENTATIVES DOCUMENTS” OR (“PRESS RELEASE” AND “Congressional Press Releases”)

A2

a: (?:Sen\.?|Rep\.?|Senator(?:s?)|Representative(?:s?)|Congressman|Congresswoman|Congress member|Member|MP|Chairman|Chairwoman(?:s?))\s(?:[A-Z][a-z]+\s(?:[A-Za-z\.\.]+\s)?[A-Z][A-Za-z\.-]+)\[\,\s\(\]

b: ([A-Z][a-z]+\s(?:[A-Za-z\.\.]+\s)?[A-Z][A-Za-z\.-]+)\s\([RD]\-?[A-Za-z\.\.]{2,3}\)

c: Speaker ([A-Za-z\.\.]+\s)\s0,2?s Press Office|Office of Speaker ([A-Za-z\.\.]+)

A3

(\<|\-end\-\|Sign up|share\:\s+f\s+t\s+|Press Release\:\s|\#\s?\#char' s?\#\|(?:\s+[Aa]1,|[Aa]1,\s+)2,)

A4

r '[\V#!\$%^ &*;:;{}=\-_'~0]', r '[0-9]', r '[A-Z]', r '(((Phone\?:?|Fax\?:?)\s+?)?\(?[0-9]{3}\)\?[\s\-\.\.])?\[0-9]{3}[\s\-\.\.]?[0-9]{4}', r '(Hours)?\?:?\s?(?(\s+?)\(?M\F\)?(\s+?)?)?[0-9]{1,2}(\:[0-9]{2})?\s?(PM|pm|AM|am|P\M\.|A\M\.|a\.m\.|p\.m\.)\(\s?\-\ \s?[0-9]{1,2}(\:[0-9]{2})?\s?(PM|pm|AM|am|P\M\.|A\M\.|a\.m\.|p\.m\.)\)?(\s+?(EST|CST)?\s+?)?((\s+?)\(?M\F\)?(\s+?)?)?', r 'Monday\~Friday', r '\w+\[w\.\.]+\.(gov|edu|com|net)', r '\s((20|19)[0-9]{2})\s', r '\((([DR]\-)?[A-Z][A-Za-z](\-\ ?[0-9]{1,2})?)\)', r '\((HR|H\R\.|H\Res\.|S\|S\R\.)\(\s)?[0-9]{1,4}\)', r 'http(s)?\:\V\/(www\.)?+\.(gov|edu|com|net)', r '\\$[0-9]{1,3}(\,[0-9]{3})+\s' r '\[|\]', r '\\', r '\;', r '\#', r '\s', r '[a-z]+([A-Z][a-z]+){2,}', r '\<\>', r '_', "function(", "", "style", "font", "*", "mso-", "-30-", "http", "www", "#", "javascript", "###"

A5

r '\\$([0-9]{1,3}(?:\:\, [0-9]{3})+)?(?:\.[0-9]{1,2})?\s'; r '\W([A-Z][a-z]+\W'; r '\W([0-9]+\W'; 'house', 'senate', 'announced', 'voted', 'passed', 'passage', 'introduced', 'approved', 'committee', 'subcommittee', 'will', 'final', 'signed', 'Following'

A6

http://((https?:\V\/(www\.)?)?[-a-zA-Z0-9@:%_\+~#=#]{2,256}\.[a-z]{2,6})b([-a-zA-Z0-9@:%_\+~#?&//=*])