SCIENCE POLICY

What does Congress want from the National Science Foundation? A content analysis of remarks from 1995 to 2018

A. Lupia[1]*, S. Soroka[1,2], A. Beatty[3]

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

The U.S. Congress writes the legislation that funds the National Science Foundation (NSF). Researchers who seek NSF support may benefit by understanding how Congress views the agency. To this end, we use text analysis to examine every statement in the Congressional Record made by any member of Congress about the NSF over a 22-year period. While we find broad bipartisan support for the NSF, there are notable changes over time. Republicans have become more likely to express concerns about accountability in how the NSF spends its funds. Democrats are more likely to focus on how NSF-funded activities affect education, technology, and students. We use these findings to articulate how researchers and scientific organizations can more effectively conduct transformative science that corresponds to long-term and broadly held Congressional priorities.

INTRODUCTION

Scientific discoveries transform society. They teach us about ourselves, our relationships, the environments in which we live, and the universes beyond. All over the world, people use science's insights to more efficiently achieve their aspirations and more effectively manage important societal challenges (*1*).

Science has improved how governments deliver many critical public services to improve outcomes in health care, law enforcement, education, and many more domains (*2*). Science strengthens national security by improving instrumentation, informing military strategies, and strengthening efforts to protect vulnerable populations (*3*).

Science also energizes the private sector. It makes factories, offices, and farms more efficient. It does so by allowing rigorous comparisons of old and new ideas. Over time, these competitions produce innovations that entrepreneurs can use to improve production and serve customers better.

Advancing science requires varying combinations of infrastructure, materials, intellect, and teamwork. In many countries, governments provide financial support to these endeavors. This support takes many forms including grants, contracts, and funding for colleges and universities.

Why do governments fund science? One reason is that scientific discoveries produce public goods. Scientific discoveries yield public goods when they produce nonexclusive benefits. A discovery's benefits are nonexclusive when they provide benefits to some people without forcing others to pay for reusing the same idea. Consider, for example, a scientific discovery that helps a farmer improve crop yields. If other farmers or people in related industries can freely use the same discovery to improve their production, the discovery's value multiplies.

However, a known problem with public goods is that potential beneficiaries have incentives to let others pay for the good (also known as the "free rider" problem). For example, private sector actors have strong incentives to invest in research that only they can use and weak incentives to invest in research that potential competitors can use. Economists cite free-rider problems when explaining why the private sector provides fewer public goods than is socially optimal (*4*).

An implication is that if governments do not invest in public good science, then we cannot count on the private sector to pick up the slack. For this and other reasons, many governments support research that can be widely used.

In 1950, the U.S. Congress established the National Science Foundation (NSF) "[t]o promote the progress of science; to advance the national health, prosperity, and welfare; and to secure the national defense…" Today, organizations of all kinds use NSF-funded scientific research to improve knowledge, increase productivity, and build capacity (*5*).

The NSF is responsible for roughly 25% of all federally supported fundamental research. This includes roughly 83% of the federal funding for computer science research at U.S. academic institutions and two-thirds of basic research in the social and behavioral sciences at the same institutions. While there is widespread appreciation of NSF's past support, there are questions about what it should fund now. Members of Congress ask some of these questions. Their views matter for several reasons.

To fund any portfolio of scientific activities at the NSF, majority support on the floor of the House of Representatives and in key House committees is required. The Senate's assent is also needed, which can be harder to achieve due to its super-majoritarian rules and norms. Through these processes, Congress makes decisions about NSF's funding levels and about its discretion in using those funds. For people who care about the availability of federal support for basic research in the United States, a relevant question becomes: What do members of Congress want from the NSF?

To address this question, we have pulled from the Congressional Record a dataset that includes everything in it that any member has said about the NSF from 1995 to early 2018. We use these data to identify patterns in congresspersons goals and concerns. We find broad and persistent bipartisan support for the NSF. We also identify partisan differences that expand over time. In particular, Republicans become more likely to express concerns about accountability, while Democrats focus on how NSF funding affects education, technology, and students.

[1]Department of Political Science, University of Michigan, Ann Arbor, MI, USA. [2]Department of Communication and Media, University of Michigan, Ann Arbor, MI, USA. [3]Gerald R. Ford School of Public Policy, University of Michigan, Ann Arbor, MI, USA.
*Corresponding author. Email: lupia@umich.edu

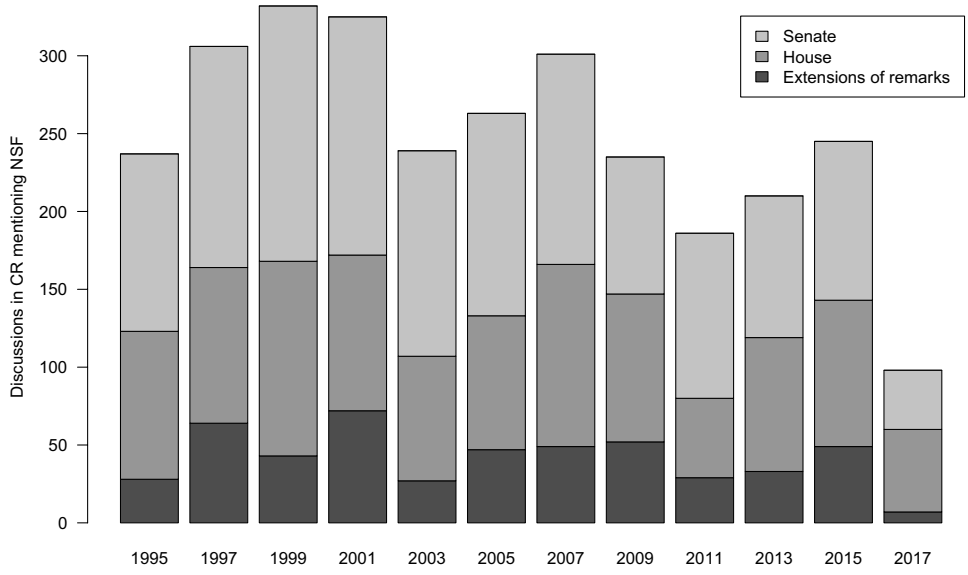

Fig. 1. Number of discussions, by House/Congress. CR, Congressional Record.

Table 1. Most frequent words co-occurring with NSF, 1995–2017.

Top 20 words	Count
Program	2061
Fund	1644
Year	1394
Education	1177
Agency	970
Technology	940
Bill	896
Act	879
Grant	814
Support	765
Department	749
Math	651
Federal	636
Million	615
Fiscal	588
Budget	570
Director	546
Engineering	545
Amend	527
Health	498

RESULTS

Congress discussed the NSF many times between 1995 and the early months of 2018. As Fig. 1 shows, in most Congresses, there were between 200 and 300 discussions about the NSF.

It is from this discussion-level database that we derive a sentence-level dataset, which contains 8206 sentences (distributed over time in roughly the same way as discussions, in Fig. 1). Our first analysis examines the words that most frequently occur in sentences that include the terms "NSF" or "National Science Foundation." Table 1 shows the 20 most frequently used words over the entire time period. Words are in the second column; the actual frequencies are in the third. To facilitate comparisons in later tables, we put the top five terms in bold font.

Table 2 shows how discussions change over time. In it, we compare the words that most distinguish the earlier years of our data (1995–2006) from the later years (2007–2018). The distinction between these two periods is arbitrary—it cuts the time period for which we have data in half but otherwise does not coincide with any fundamental change, either in government or in NSF discussions. Most of the change over time that we examine happens incrementally, however, and the two-period analysis nicely captures that change. In many instances, we include Congress-to-Congress changes in tables S1 to S3 to confirm that the two-period distinction is not masking other meaningful variation.

Table 2 shows the words that most uniquely identify the two time periods. Note that these need not be the most frequently used words in any period: Words that are equally common (or uncommon) in these two time periods will not appear in Table 2. Table 2 provides a sense of how discussions are changing over time. It suggests an increasing focus on education and innovation in the more recent time period. The 1995–2006 data are dominated by relatively standard procedural and budgetary language; the 2007–2018 data highlight words just as education (including stem, teacher, and scholarship), as well as innovation and technology.

Are similar patterns evident within each political party, or are there important differences between parties? We next examine similarities and differences in the words that appear when members of the two major parties discuss the NSF. Table 3 shows the 20 most common words used by each party. The table shows a great deal of similarity. The overlap likely reflects the fact that members of Congress tend to adopt common narrative structures when discussing this issue. They tend to introduce their comments in similar ways, and they

Table 2. Most distinguishing words co-occurring with NSF, 1995–2006 to 2007–2018. Number in parentheses denotes ranking over the entire period 1995–2018. Findings based on sentence-level analysis, in (nontitle) sentences that include a direct mention of the NSF (see Materials and Methods for details on pre-whitening of text). Words in bold are the top five words mentioned over the entire time period. "Most distinguishing words" for each period are the top 20 *diff* scores, where for each word *w*, the following is calculated for the first time period (P) and the second time period (Q), respectively: $diff_w^P = \frac{count_w^P}{\#sentences^P} - \frac{count_w^Q}{sentences^Q}$, and $diff_w^Q = \frac{count_w^Q}{\#sentences^Q} - \frac{count_w}{sentences^P}$.

The 20 words that most distinguish the time periods:

1995–2006	2007–2018
Year	**Program**
Fiscal	Bill
Budget	Teacher
Agency	**Education**
Percent	Grant
Expenses	**Fund**
Necessary	Department
Million	Stem
Federal	Teaching
Secretary	Standards
Carrying	Innovation
Provided	Nist
Committee	Interest
Basic	Energy
Activity	Partnership
Environmental	Office
Epa	Institute
Salary	Scholarship
May	Technology
Director	Billion

Table 3. Most common words co-occurring with NSF, by party. Based on sentence-level analysis, in (nontitle) sentences that include a direct mention of the NSF (see Materials and Methods for details on pre-whitening of text). Words in bold are the top five words mentioned over the entire time period.

Twenty most common words for…

Democrats	Republicans
Program	**Program**
Fund	**Fund**
Year	**Year**
Education	**Agency**
Bill	**Education**
Agency	Bill
Technology	Grant
Department	Million
Support	Technology
Grant	Department
Million	Act
Math	Support
Act	Federal
Budget	Amend
Federal	Fiscal
Nasa	Budget
Engineering	Math
Health	Health
Development	Director
Basic	Administration

tend to use the same focal points such as "program" and "year" when evaluating past actions or proposing new ones. Once we get beyond these structural similarities, however, differences appear.

Table 4 offers an initial picture of those differences, focused on within-party change over time. Table 4D presents the words that most distinguish 1995–2006 Democrats from 2007–2018 Democrats. Table 4R presents the equivalent comparison for Republican members of Congress. (Again, using two time periods does not hide other features of the data; see tables S2 and S3 for top words by party and by Congress.)

On the Democratic side, there is a shift from specific programmatic words early in the time series to more general and technology-related terms in the latter part of the series. Democrats in the first period are likely to use terms such as "EPA," "housing," "HUD," and "coastal," where Democrats in the latter period are likely to use the words highlighted in Table 2, such as program, "education," and "technology." On the Republican side, we see a different shift. Republicans in the early part of the series use many of the words that Democrats used in the later years—words like program, "fiscal," and "support." In the later years, Republicans shift to accountability-related topics. Distinguishing Republican words of the later era include "taxpayer," "accountable," "dollar," and "spending."

Table 4 highlights some telling differences within the two parties over time. What about differences across parties? Table 5 shows results for the words that distinguish the two parties over the entire time period. Many of these words reflect party-level results for the latter time period—a signal that party language is diverging in the latter period. The table highlights education- and innovation-related words for Democrats and fiscal- and administration-oriented words for Republicans.

Figure 2 offers a visual representation of the relationship between the words that the parties used when discussing the NSF. Shown in the figure are a series of dots, each of which represents a word. A dot on the diagonal line that extends from the southwest corner to the northeast corner of the figure is a word that is used in equal proportions by both parties. Words to the right of the diagonal are words that Democrats use more than Republicans. Words to the left of the diagonal represent words that Republicans use more than Democrats. The figure highlights the most distinguishing words for Democrats in blue and the most distinguishing words for Republicans in red.

The figure nicely illustrates patterns seen in Tables 1 to 5. First and foremost, many words are very close to the diagonal, reflecting an extensive shared core of words. This core includes substantive words like technology and "health" and procedural words like "budget."

Table 4. Most distinguishing words co-occurring with NSF, 1995–2006 to 2007–2018. Based on sentence-level analysis, in (nontitle) sentences that include a direct mention of the NSF (see Materials and Methods for details on pre-whitening of text). Words in bold are the top five words mentioned over the entire time period. Most distinguishing words for each period are the top 20 *diff* scores, where for each word *w*, the following is calculated for the first time period (*P*) and the second time period (*Q*), respectively: $diff_w^P = \frac{count_w^P}{\#sentences^P} - \frac{count_w^Q}{sentences^Q}$, and $diff_w^Q = \frac{count_w^Q}{\#sentences^Q} - \frac{count_w^P}{sentences^P}$

Twenty words that distinguish between time periods:

4D. Democrats only:		4R. Republicans only:	
1995–2006	2007–2018	1995–2006	2007–2018
Budget	**Program**	**Year**	Grant
Epa	**Education**	**Program**	Interest
Housing	**Fund**	Fiscal	Taxpayer
Plant	Technology	**Agency**	Accountable
Provided	Bill	Million	Dollars
Hud	Act	Support	Spending
Resolution	Department	Act	Money
Committees	Grant	Federal	American
Five	Support	Shall	Change
Almost	Math	Administration	Climate
Coastal	Standards	Budget	Innovation
Particularly	Teacher	**Education**	Political
Percent	Engineering	Authorize	People
Nsfs	Institute	Activity	Stem
Internet	Energy	Basic	Funds
Nations	Shall	Section	Spent
Genome	Section	Health	Mission
Half	Year	Engineering	Awarded
Increased	Billion	Fund	Sure
Time	Nist	Environmental	Earlier

Table 5. Most distinguishing words co-occurring with NSF, 1995–2018, by party. Based on sentence-level analysis, in (nontitle) sentences that include a direct mention of the NSF. Text is pre-whitened to remove numbers and standard stopwords. Words in bold are the top five words mentioned over the entire time period (by both parties). Most distinguishing words for each party are the top 20 *diff* scores, where for each word *w*, the following is calculated for Democrats (D) and Republicans (R), respectively: $diff_w^D = \frac{count_w^D}{\#sentences^D} - \frac{count_w^R}{sentences^R}$, and $diff_w^R = \frac{count_w^R}{\#sentences^R} - \frac{count_w^D}{sentences^D}$

Twenty most distinguishing words for…

Democrats:	Republicans:
Program	Money
Fund	Amend
Bill	Project
Education	Expenses
Technology	Necessary
Math	Human
Nasa	Board
Students	Projects
Teacher	Study
Institute	Report
Department	Administration
Budget	Fiscal
Support	Director
Innovation	Taxpayer
Nist	Activity
Engineering	Commission
Institutions	Ensure
Development	Authorize
Doubling	Account
Universities	Committee

Words that fall farther from the diagonal distinguish the parties. Again, we see a Republican emphasis on accountability and taxpayers and a Democratic emphasis on education and innovation.

One potential weakness of our approach thus far is that it makes use of a very limited number of words—either the most frequent or most distinguishing words. We can also adopt a method that better leverages the full content of our data and use structural topic modeling (STM) to learn from the entire corpus. A principle advantage to STM (much like other automated approaches) is that it makes no assumptions about the structure of the data—there are no predefined dictionaries, only a predefined number of dimensions. The *stm* package in R is useful for our purposes, because it allows for the identification of dimensions while taking into account metadata such as year or party. Full information about the *stm* package is available at www.structuraltopicmodel.com [and in (*6*, *7*)]. Because this prior work outlines the methods and advantages of STM in some detail, we do not do so here. Here, we use the approach to explore the possibility that our simple analysis misses important trends in the data.

We estimate a structural topic model using all the sentences that mention the NSF and are linked to a speaker from either the Democratic or Republican Party. (This is of course the same corpus as is used in Tables 3 to 5 and Fig. 2.) We include both party and Congress number (each 2-year session has a distinct number) as metadata in the estimation. Because we have no a priori expectations about the correction number of dimensions to use in the analysis, we pretest models with anywhere between 3 and 15 dimensions. On the basis of the diagnostic properties of those models, we present a 10-dimension model here; although the basic structure of the text is relatively similar when we use between 7 and 15 dimensions. (The extent to which Democrats score higher on education- and innovation-related dimensions does not change with different numbers of topics. Increasing the number of topics does, however, tend to remove topics that are clearly about spending, which changes the signal-to-noise ratio of the Republican connection to fiscal issues).

Figure 3 presents STM results, presenting the words that are most common (Highest Prof) and most distinguishing [FREX (Frequency and Exclusivity)] for each dimension and showing the connection between each dimension and the two parties. As suggested in the results presented above, dimensions that appear to capture themes related to funding and administration (topics 3, 5, and 6; shown in red) are more frequent for Republican speakers, while dimensions related to technology and education (topics 1, 7, and 10; shown in

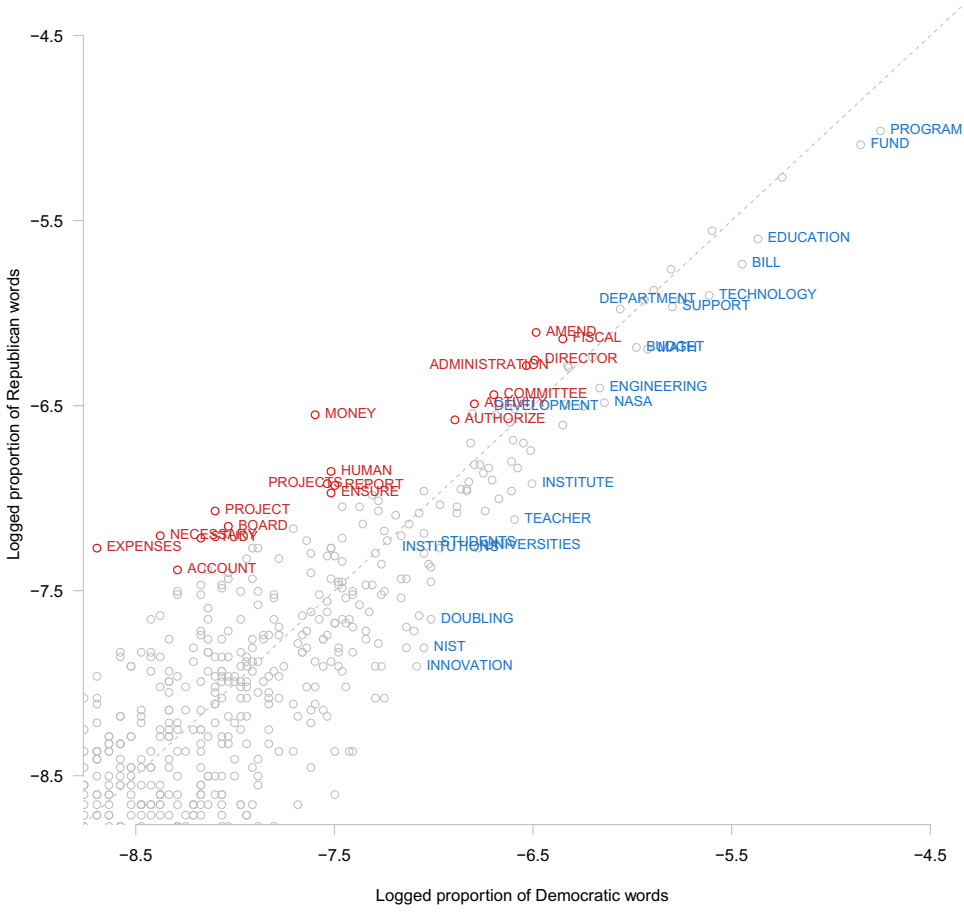

Fig. 2. Democratic and Republican use of words, compared.

blue) are more frequent for Democratic speakers. We take the results of the STM as confirmation of party-level differences evident in earlier tables and figures, this time using a more systematic analysis of the entire corpus.

Given these and preceding results, we can infer that Democratic speech has been more positive about the NSF than has Republican speech. We can of course examine this possibility more directly. There are a good number of useful tools for assigned party positions to legislative text (*8*–*10*). Recent work suggests the advantages of a simple sentiment analysis (*11*), which tends to capture parties' support or opposition to legislation. We are not focused on specific legislation here, but rather on views of the NSF generally. We nevertheless expect that sentiment analysis will work in a similar way with this target.

To estimate sentiment, we leverage a count of positive and negative words from the Lexicoder Sentiment Dictionary, as implemented in the quanteda package in R (*12*). This dictionary has been tested in detail in previous work (*13*); it is also the dictionary used to capture party support for legislation (*11*), and we use the same estimate of "net support" as is used in that work: log[(positive words + 0.5)/(negative words + 0.5)], which is an empirical logit, slightly smoothed toward zero.

One weakness of sentiment analysis is that it does not distinguish between "this is bad for the NSF" and "the NSF is bad." Hence, we must treat these sentiment estimates with some caution, recognizing that they capture the general sentiment of speech surrounding the NSF, not necessarily evaluations of the NSF itself. Keeping this in mind, Fig. 4 shows estimated sentiment for each party in each Congress. The figure is based on a regression model predicting the tone of sentences as a function of party and year (both included as categorical variables). The lines denote estimates for each year, with shading to convey 95% confidence intervals.

One notable element of Fig. 4 is the average positive sentiment of Congressional remarks. The mean estimate in every year is above zero, indicating a consistently greater number of positive versus negative remarks. We take this as signal that there is bipartisan support for the NSF. There also is no statistical significance difference between the sentiment expressed by each party in most years. That said, we see a difference emerge since 2004, and for most of the latter time period, Democratic sentiment has been more positive than Republican sentiment. This is in line with what we expect given the word counts examined above.

Note that the Congress where the difference is greatest is the first Congress of the Obama presidency, the only Congress in our entire time series that features unified Democratic control of Congress. It is possible that the combination of increasing Republican emphasis on accountability combined with a relatively weak bargaining position in Congressional negotiations (as a result of having a minority of members in each Congressional chamber) to create a circumstance where Republicans expressed relatively more frustration with their

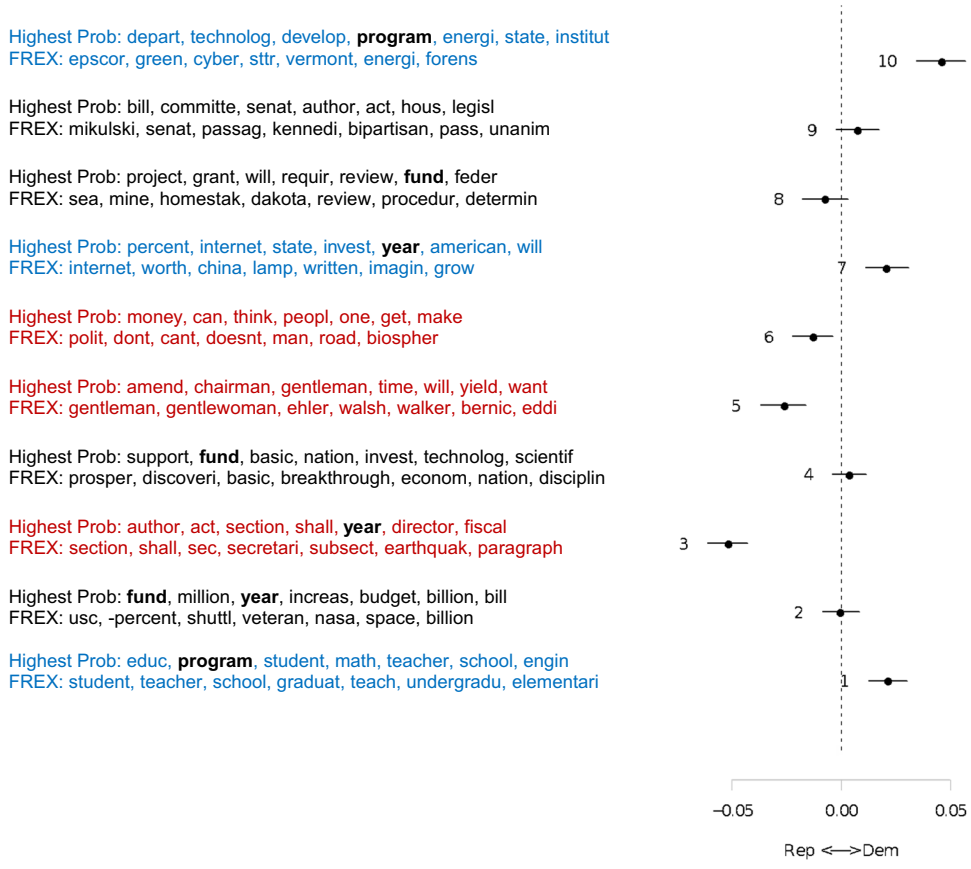

Fig. 3. Structural topic model. Based on sentence-level analysis, in (nontitle) sentences that include a direct mention of the NSF. Text is pre-whitened to remove numbers and standard stopwords, and words are lemmatized for the estimation of the STM. The words used to label each topic are shown on the left. "Highest Prob:" indicates words that occur most frequently in each topic. "FREX:" indicates words that are frequent and exclusive to each topic (*16*). Topics that are significantly more likely to be evident in speeches by Democrats or Republicans are shown in blue or red, respectively, where topic probabilities are shown on the right, based on coefficients and SEs for a 10-topic structural topic model with Democrat/Republican as a predictor. This visualization of results draws directly on (*17*).

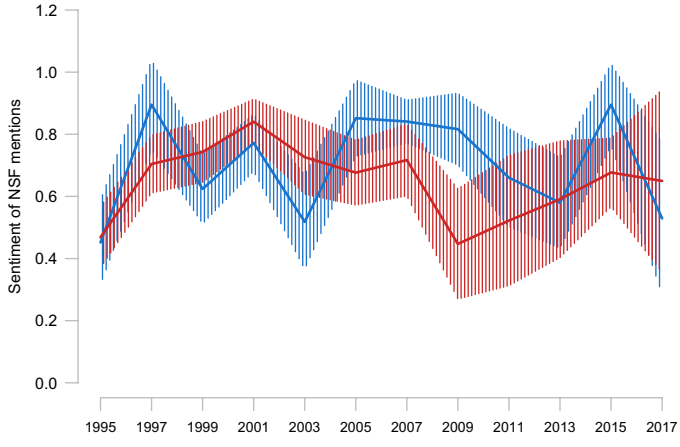

Fig. 4. Sentiment in NSF sentences, by party and Congress.

counterparts' funding priorities. Our data are not sufficient to test this conjecture. However, it may be instructive to remember that the first 45 years of the NSF's existence corresponded to an era where the Democratic Party typically controlled both houses of Congress. Since 1995, Republicans have held the Senate most of the time, and control of the House has vacillated. To the extent that Congressional priorities during NSF's first 45 years influenced the agency's portfolio, we would expect those influences to be challenged after 1995—when different long-term expectations of Congressional control emerged.

DISCUSSION

Our analysis reveals that both political parties tend to express positive sentiment when discussing the NSF. We also observe changes over time where increasing questions about accountability are arising among Republicans. How should scientists and scientific organizations that desire government support adapt to these changing circumstances?

One implication of increased competition in information marketplaces is that saying that science can be beneficial is increasingly likely to be necessary but not sufficient to generate political support for science funding. In competitive funding environments, where many other social interests appeal for greater public funding, prospective supporters of a funding plan need to persuade pivotal stakeholders that proposed activities generate significant and distinctive net benefits to key constituents. If competitors can argue that different programs, or reduced taxes, also create social value, then science's prospective supporters must be willing and able

to make arguments about the science's benefits relative to other alternatives (*14*, *15*). The Executive Office of the President makes this point directly in its Fiscal Year 2019 Research and Development Budget Priorities:

"When considering new research programs, agencies should ensure that the proposed programs are based on sound science, do not duplicate existing R&D efforts, and have the potential to contribute to the public good. Agencies should also identify existing R&D programs that could progress more efficiently through private sector R&D, and consider their modification or elimination where Federal involvement is no longer needed or appropriate. To the extent possible, quantitative metrics to evaluate R&D outcomes should be developed and utilized for all Federal R&D programs."

Similarly, if a scientist wants to be heard and taken seriously by members of Congress or their staffs, then they are likely to achieve greater success by linking the value of their proposed work to members' priorities. To think about how to make these linkages, we can use our findings to reverse-engineer the sentiment identification process in the most challenging time periods in our data. In other words, we can examine, in texts identified in the automated content analysis, similarities and differences in what the two parties were asking for in the 111th and 115th Congresses, when they were most divided on NSF-related questions. Table S4 offers some illustrative examples.

It is important to note that the sentences in table S4 came from focusing on the two Congresses where the parties' estimated sentiments were most different. It is also important to note that even here—at the most divisive moments—each comment refers to an aspirational view of the NSF. Each member wants the NSF to serve the public as effectively as possible. When disagreements emerge, it is about how to do so. This shared foundation reflects the larger finding that NSF has significant bipartisan support and that the parties often use similar language to express this sentiment.

Going forward, many researchers, science advocates, and people who benefit from research ask not only for support for science today but also for Congress to make commitments to types of scientific inquiry that may take years to deliver truly transformative results. To sustain support for science funding over long periods of time requires the support of Congress over long periods of time. Control of Congress, over long periods of time, tends to switch from one party to the other. As a result, appeals for science funding that are responsive and accountable to the core values that unite the two major parties are likely to provide greater leverage to long-lasting bipartisan coalitions than will partisan appeals that divide the parties. In the current era, this means being responsive to both programmatic preferences and calls to rigorously document careful stewardship of resources and tangible benefits to the taxpayer. Given the frequency with which legislators are asked to discuss and defend their values, preferences, and decisions, scientific portfolios that clearly serve broad public interests are important for government accountability and the legislative process. We hope that the data that we have made available and this initial set of findings help members of Congress, scholars, and the public better understand the ways in which science can best serve the nation.

MATERIALS AND METHODS

We collected all speeches in the Congressional Record, including published extensions of remarks, that mentioned "NSF" or the "National Science Foundation" from 1 January 1995 to 28 February 2018. These data cover 12 Congresses and 4 presidential administrations (Clinton, Bush 43, Obama, and Trump). We collected data from 1995 to May of 2017 in May of 2017. We collected subsequent data during February of 2018. We collected content from the "text only" option that appears when conducting a text-based search on the following page (www.congress.gov/congressional-record/browse-by-date). For readers who want to check the validity of our claims or pursue analyses of their own, we are making our working database, alongside the code used to produce all the analyses that follow, available at Harvard Dataverse (https://doi.org/10.7910/DVN/UWNAAO).

Our data analysis focuses only on the sentences in which "NSF" or in which NSF appear—these form the empirical corpus in the analyses that follow. The majority of metadata relevant to our analysis is already included in the downloaded data, including the date on which the comment was made, the identity of the person speaking, and whether the text came from the proceedings of either the House or the Senate, or whether the text came from Extensions of Remarks that were subsequently entered into the Congressional Record. To these data, we add each legislator's partisanship, obtained from the Congress Collection at Congressional Quarterly Press.

We take several different approaches to analyzing the database. We first focus on simple word frequencies, identifying the most frequent words used by members of Congress, together and by party, for every year in our time series. We also identify the words that most distinguish parties from one another and time periods from one another. To be clear, we examine the words that Republicans use when discussing the NSF and that Democrats do not use, and vice versa. These simple analyses clarify important similarities and evolving differences across parties. We then turn to more sophisticated approaches to textual data. We use STM, which better takes advantage of our corpus. Results in this case largely confirm the story told by simple word frequencies. In line with recent work (*11*), we also use a measure of sentiment to examine party positions on the NSF over time.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/33/eaaz6300/DC1

REFERENCES AND NOTES

1. National Science Board, *Science and Engineering Indicators* (2018); www.nsf.gov/statistics/2018/nsb20181/.
2. National Academy of Sciences, From Research to Reward (2016); www.nasonline.org/publications/from-research-to-reward.
3. National Research Council, *Intelligence Analysis for Tomorrow: Advances from the Behavioral and Social Sciences* (The National Academies Press, Washington DC, 2011).
4. M. Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Harvard Univ. Press, Cambridge, MA, 1965).
5. National Science Foundation, NSF and the Birth of the Internet (2017); www.nsf.gov/news/special_reports/nsf-net/textonly/90s.jsp.
6. M. E. Roberts, B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. Kushner Gadarian, B. Albertson, D. G. Rand, Structural topic models for open-ended survey responses. *Am. J. Polit. Sci.* **58**, 1064–1082 (2014).
7. C. Lucas, R. A. Nielsen, M. E. Roberts, B. M. Stewart, A. Storer, D. Tingley, Computer assisted text analysis for comparative politics. *Polit. Anal.* **23**, 257–277 (2015).
8. K. Benoit, M. Laver, Estimating Irish party policy positions using computer wordscoring: The 2002 election—A research note. *Irish Polit. Studies* **18**, 97–107 (2003).
9. W. Lowe, Understanding wordscores. *Polit. Anal.* **16**, 356–371 (2008).
10. J. B. Slapin, S.-O. Proksch, A scaling model for estimating time-series party positions from texts. *Am. J. Polit. Sci.* **52**, 705–722 (2008).

11. S.-O. Proksch, W. Lowe, J. Wäckerle, S. Soroka, Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches. *Legis. Stud. Q.* **44**, 97–131 (2019).
12. K. Benoit, K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller, A. Matsuo, quanteda: An R package for the quantitative analysis of textual data. *J. Open Source Softw.* **3**, 774 (2018).
13. L. Young, S. Soroka, Affective news: The automated coding of sentiment in political texts. *Polit. Commun.* **29**, 205–231 (2012).
14. A. Lupia, Now is the time: How to increase the value of social science. *Soc. Res. Int. Q.* **84**, 669–694 (2017).
15. D. B. Magleby, A. Lupia, Why Do Governments Pay for Your Research? The Political Economy of Science Funding Decisions, Prepared for the 2018 Annual Meeting of the Midwest Political Science Association, 2018.
16. J. M. Bischof, E. M. Airoldi, Summarizing topical content with word frequency and exclusivity, in *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, Scotland, UK, 2012.
17. C. Lucas, R. A. Nielsen, M. E. Roberts, B. M. Stewart, A. Storer, D. Tingley, Computer-assisted text analysis for comparative politics. *Polit. Anal.* **23**, 254–277 (2015).

Disclaimer: Arthur Lupia presently serves as Assistant Director for the Social, Behavior, and Economic Sciences Directorate of the National Science Foundation. Dr. Lupia participated in this project as a Professor at the University of Michigan. All data was collected from public sources and before he assumed the role of Assistant Director at NSF. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Acknowledgments: We are grateful to J. H. Aldrich, B. Baird, C. Boudreau, J. N. Druckman, A. Furnas, Y. Krupnikov, B. Lee, A. S. Levine, S. Morell, M. Oceno, S. Patel, J. Rezaee, and E. Suhay for comments on previous drafts of this paper. We thank C. Hosman, J. Milton, K. Prewitt, and W. Naus for assistance on appropriate data sources for this project. **Author contributions:** A.B. gathered the data, conducted preliminary analyses, and helped draft portions of the final paper. A.L. and S.S. produced the final analyses and paper. **Competing interests:** S.S. and A.B. declare that they have no competing interests. **Data and materials availability:** The data and script used in this paper are archived at the Harvard Dataverse (https://doi.org/10.7910/DVN/UWNAAO). Additional data related to this paper may be requested from the authors.

Submitted 25 September 2019
Accepted 2 July 2020
Published 14 August 2020
10.1126/sciadv.aaz6300

Citation: A. Lupia, S. Soroka, A. Beatty, What does Congress want from the National Science Foundation? A content analysis of remarks from 1995 to 2018. *Sci. Adv.* **6**, eaaz6300 (2020).