Corporate funding and ideological polarization about climate change

Justin Farrell[1]

School of Forestry & Environmental Studies, Yale University, New Haven, CT 06511

Drawing on large-scale computational data and methods, this research demonstrates how polarization efforts are influenced by a patterned network of political and financial actors. These dynamics, which have been notoriously difficult to quantify, are illustrated here with a computational analysis of climate change politics in the United States. The comprehensive data include all individual and organizational actors in the climate change countermovement (164 organizations), as well as all written and verbal texts produced by this network between 1993–2013 (40,785 texts, more than 39 million words). Two main findings emerge. First, that organizations with corporate funding were more likely to have written and disseminated texts meant to polarize the climate change issue. Second, and more importantly, that corporate funding influences the actual thematic content of these polarization efforts, and the discursive prevalence of that thematic content over time. These findings provide new, and comprehensive, confirmation of dynamics long thought to be at the root of climate change politics and discourse. Beyond the specifics of climate change, this paper has important implications for understanding ideological polarization more generally, and the increasing role of private funding in determining why certain polarizing themes are created and amplified. Lastly, the paper suggests that future studies build on the novel approach taken here that integrates large-scale textual analysis with social networks.

funding | polarization | politics | computational social science | climate change

Ideological polarization presents increasingly important challenges for sustainability science and solutions for climate change. Much attention has been given to the outcomes of polarization by demonstrating its effect on individual choices about energy-efficient behavior (1), individual attitudes about climate change (2–6), and variation in individuals' trust in science as a whole, especially among Americans who self-identify as politically conservative (7). Climate change is not the first issue with scientific consensus to become so highly polarized (8, 9), but the magnitude of its ecological and human-health effects—and the socio-political roadblocks for mitigating emissions—have led to a wide body of attitudinal research documenting the current socio-political environment of polarization that exists.

Despite this fruitful body of individual-level research, we have considerably less data and understanding about the underlying organizational and financial factors that made polarization possible in the first place. Qualitative and historical research has suggested that well-funded and well-organized "contrarian" campaigns are especially important for spreading skepticism or denial where scientific consensus exist—such as in the present case of global warming, or in historical contrarian efforts to create doubt about the link between smoking and cancer (8–15). The primary way that contrarian campaigns create skepticism and ideological polarization is through the production of an alternative contrarian discourse (To be clear, the term discourse here refers to communication in an authoritative fashion about a particular topic or debate), which necessarily takes the form of written text and speech from organizations and individuals (11, 14). Indeed, many scholars have examined climate change discourse in media

coverage of climate change (16–21), but because of data constraints and the difficulty of gathering such complex and furtive data, we still lack a comprehensive data-driven understanding about the actual content and source of contrarian messages, as well as the complex organizational and financial networks within which they are produced. This study presents such an approach, and examines how the production of an alternative discourse is embedded within a particular social structure and how the content itself is influenced by particular funding sources.

Important to this approach is the fact that in the United States, there are a growing number of grassroots lobbying firms who work on behalf of corporations, industry groups, and associations (22, 23). Along with this growth in corporate lobbying, other social and political opportunities have opened the door for movements like climate change contrarianism to flourish, such as weakening restrictions on political finance (24) and the concentration of corporate wealth more generally (25, 26). With these factors in mind, and building on prior climate change research, this study asks three specific—and closely related—empirical research questions: (*i*) Of all of the organizations in the climate contrarian movement, which ones produced discourse? (*ii*) What are the specific themes contained in this contrarian discourse? (*iii*) Does the reception of corporate funding influence the thematic content and ideological language of this discourse? And, how do all of these factors change over time?

These important questions have not been adequately addressed because of the difficulty of collecting and analyzing such large amounts of longitudinal textual content and funding data. Given the breadth of actors in this movement and the sheer volume of texts produced, scholars have either focused on smaller samples of text (11–13), or on rigorously identifying the organizations
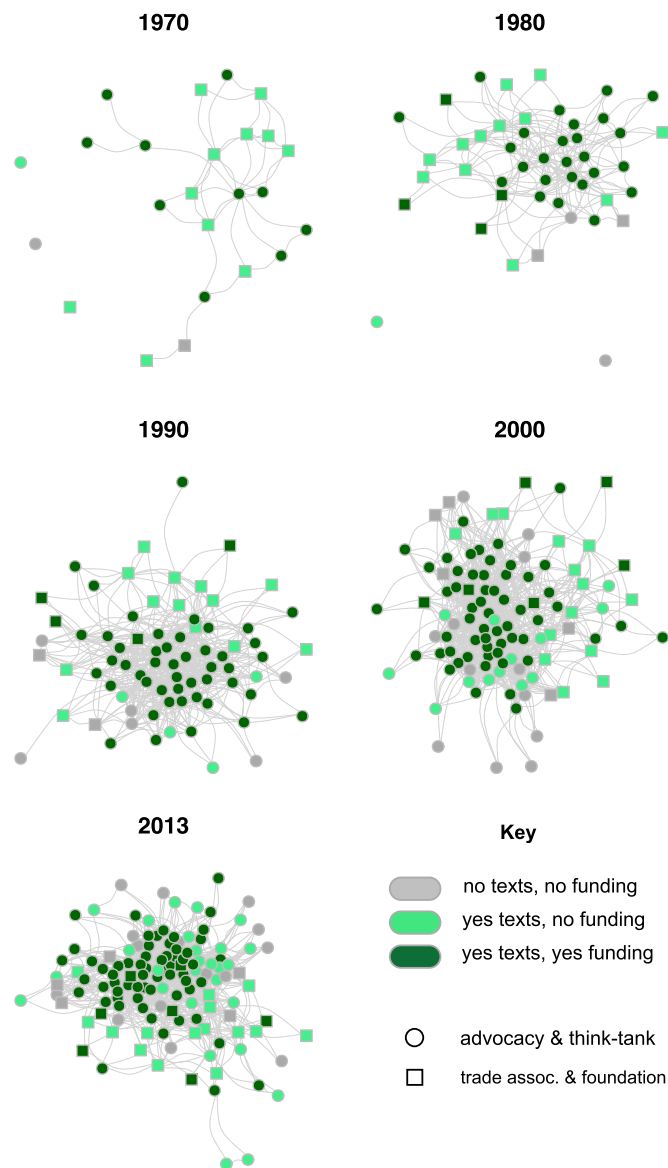

Fig. 1. Which organizations produced discourse? Nodes in this network are organizations. The color of the node indicates whether they produced a text, and whether they received corporate funding. The shape of the node indicates the type of organization. Graphed using the Fruchterman–Reingold algorithm. The data are subsetted by decade based on the year an organization was founded, giving a sense of younger versus older organizations. Attention should be paid to the full model (year = 2013), which includes all nodes and ties, and is the network on which the significant measures were calculated.

who participate in the counter movement (14, 15). This study combines—and significantly extends—these important approaches with the use of novel computational data-collection methods.

The first half of the data include a new social network of all known individuals and organizations who have participated in the climate contrarian movement between 1993 and 2013. As described in much more detail in *SI Appendix*, contrarian organizations are those identified by prior peer-reviewed research as overtly producing and promoting skepticism and doubt about scientific consensus on climate change. This large bipartite network includes 4,556 individuals (e.g., board members, employees, politicians, researchers) with ties to 164 organizations (e.g., think tanks, foundations, trade associations, grassroots

lobby firms). This population of organizations was built-up primarily from a published census of organizations and funding (12, 14, 15), and supplemented with lists from reputable non-profit organizations (27, 28).

The second half of the data include a new dataset of every text about climate change produced by every organization between 1993 and 2013 ($n = 40,785$ texts with more than 39 million words). This massive corpus was constructed with the assistance of automated Python scraping scripts, which gathered, cleaned, digitized, and prepared for analysis the entirety of current and archival press releases, website articles, policy statements, conference transcripts, published papers, and blog articles. Any PDFs were converted to plain text with optical character recognition. (See *SI Appendix* for specific information about how these text and network data were collected and cleaned.)

These data provide a comprehensive analytical framework for examining the influence of corporate funding on the discursive efforts of organizations to create ideological polarization around climate change. The collection of organization ties and their total population of texts is the product of a real-life counter-movement and do not suffer from the methodological drawbacks that hamper survey research or social laboratory experiments. Furthermore, it links the organizations involved in the movement directly with the texts that they produce, enabling analysis about how organizational features might influence the actual thematic content of the texts. With this approach in mind, the data also include several important organizational covariates, such as the year the organization was founded, its total assets, its mission focus, the organization type, and an organization's influence on the transfer of information within the network (i.e., betweenness centrality) (29). Most importantly for the purposes of this study, the data also include a measure of corporate funding from entities that prior literature on contrarian movements have identified as especially influential (8, 9, 14, 15): ExxonMobil (EM) and the Koch family foundations (KFF). Relying on this past research, along with Internal Revenue Service data (see *SI Appendix* for important details), this measure records whether or not an organization in the network received funding from either of these entities between 1993 and 2013. Donations from these corporate benefactors signals entry into a powerful network of influence.

Analytical Approach

To empirically examine these data in relationship to the research questions above regarding the influence of corporate funding on textual content, the study employs a combination of social network analysis and a form of large-scale computational text analysis called latent dirichlet allocation (LDA) (30) topic modeling. Generally stated, topic modeling is a computer-assisted content analysis procedure whereby a set of texts are coded into substantively meaningful themes called "topics." These topics are not given to the machine beforehand, but emerge inductively as algorithms learn the hidden patterns underlying a collection of texts. The model assumes a relational theory of meaning by measuring the patterns of co-occurrence of words in individual texts and across the entire corpus. Thus, in addition to providing a way to conduct reliable content analysis on massive collections of text that are too big to code by hand, topic models allow researchers to use machine learning to discover patterns and relationships that may have otherwise been missed by hand coding methods. Recent research has shown that unsupervised methods like the ones used here have performed as well as human coders on the same set of documents (31).

More specifically, this paper uses a recently developed approach to topic modeling called Structural Topic Modeling (STM) (31, 32), because it enables the discovery of topics and their prevalence based on document metadata, such as the year written, or important organizational attributes examined here,
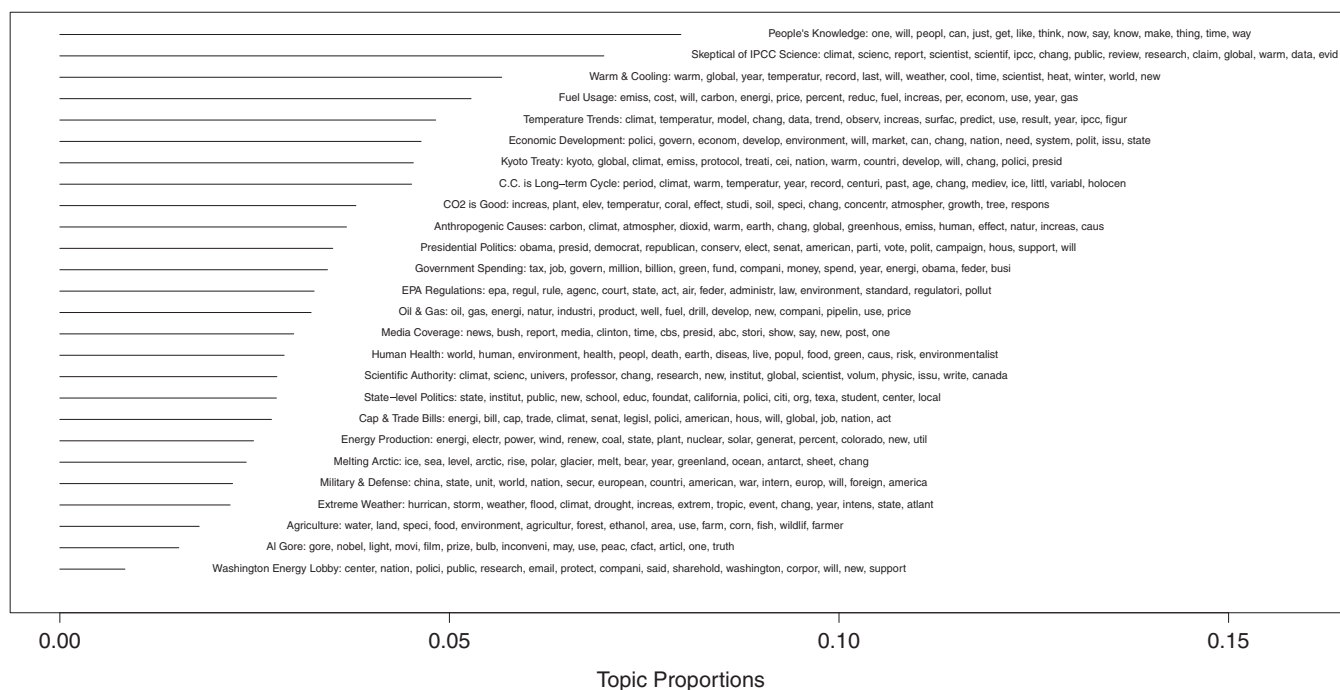
Fig. 2. Structural topic model results from 40,785 documents, including the topic label and the top 15 words associated with each. The topic proportions indicate the proportion of the corpus that belongs to each topic.

such as corporate funding. The inclusion of metadata such as the year in which a document was published is especially useful given that these text data cover such a broad period (1993–2013), and given that the climate change discourse during this period was susceptible to thematic change. Building on the standard LDA model (30), Roberts et al. (31) explain, importantly, that:

> As in LDA, each document arises as a mixture over K topics. In the STM, topic proportions (θ) can be correlated, and the prevalence of those topics can be influenced by some set of covariates X through a standard regression model with covariates $\theta \sim LogisticNormal(X\gamma, \Sigma)$. For each word ($w$) in the response, a topic (z) is drawn from the response-specific distribution, and conditional on that topic, a word is chosen from a multinomial distribution over words parameterized by β, which is formed by deviations from the baseline word frequencies (m) in log space ($\beta_k \alpha \exp(m + \kappa_k)$)...Thus, there are three critical differences in the STM as compared to the LDA model...(1) topics can be correlated; (2) each document has its own prior distribution over topics, defined by covariate X rather than sharing a global mean; and (3) word use within a topic can vary by covariate U. These additional covariates provide a way of "structuring" the prior distributions in the topic model, injecting valuable information into the inference procedure.

The STM method is applied to the totality of text data described above to (*i*) map the entire thematic landscape of content produced by the climate contrarian movement, and (*ii*) plot covariate interactions using text metadata to examine how corporate funding ties influence topic prevalence and change over time. It is important to note that while these are automated computational methods, they require a deep human understanding of the corpus that guides the number of topics that are estimated, and a process of recursive interpretation of the model's results based on prior findings (see *SI Appendix* for important details about topic model estimation and validation).

Taken together, these analyses develop a quantitative account of all public information produced by the climate change countermovement. Most importantly, they assess the substantive content of information within the organizational structure of the contrarian network, providing important analytical leverage to

understand the ways that private corporate funding influences the actual content and prevalence of polarizing discourse.

Results

To begin, Fig. 1 provides a broader bird's-eye-view context with which to interpret the main text analysis findings below. The graph describes which organizations (nodes) produced texts and whether the organization received corporate funding. For example, a dark green circle node is an advocacy or think-tank organization that produced texts about climate change and received corporate funding. It is a one-mode network constructed from the original bipartite social network data, where ties between organizations are a function of a tie they share with individuals (see *SI Appendix* for details). To give a sense of change over time, the graph includes cross-sections from each decade that restricts the nodes and ties to organizations founded before that year. This approach means that 2013 is the final graph that includes all nodes and ties in the data.

There are two related findings that emerge from this network analysis and serve to inform the main text analysis below. First, organizations who produce climate contrarian texts (versus those who did not) are significantly ($P < 0.006$) more likely to have higher betweenness centrality scores, which is a reliable indicator that they exert more organizational influence in the network (29). One can spot this trend in the figure by recognizing that these nodes tend to cluster toward the center of the network. Second, organizations who produced texts and received corporate funds were also significantly more likely ($P < 0.002$) to have higher betweenness centrality scores. Taken together, these results suggest that organizations within the movement who made an effort to produce textual discourse about climate change are the most central to the movement itself, providing them more influence over the transfer of information. This finding aligns with prior knowledge about resource mobilization (33), discursive influence (34), and corporate mobilization (22). The fact that the production of discourse was more likely to be corporately funded organizations at the core of the network provides an important

backdrop for the novel analytic focus of this article—namely, assessing via computational text analysis the influence of corporate funding on the actual thematic content and ideological polarization of the discourse itself.

Turning to the results of the computational text analysis, Fig. 2 summarizes the results of the structural topic model. This corpus level visualization displays all of the topics along with the top 15 words associated with each topic. This figure also includes the proportion of the corpus that belongs to each topic. It is important to note that these topics and the words should be interpreted in light of the purpose of the texts and the larger goals of climate change contrarianism described above (8–15). Broadly, one can see that these organizations focused their climate change discourse on people's knowledge of the issue, Intergovernmental Panel on Climate Change science, debates over temperature trends, international politics such as the Kyoto Treaty, and related energy issues such as fuel, oil, and cap and trade bills.

How are all of these topics related to one another, and do they cluster together in meaningful ways? Fig. 3 plots the layout of all topic correlations using the Fruchterman–Reingold algorithm. A tie between two topics indicates a positive topic correlation, meaning that both topics were more likely to be discussed together within a document. Node sizes correspond to the topic proportions from Fig. 2. The dashed ellipses were overlaid after graphing to draw attention to four thematic clusters: First, toward the bottom left (in green) is a cluster of topics relating to questions about scientific evidence and disputes over the real causes and effects of CO_2, long-term temperature cycles, global cooling, and who has scientific authority. Adjacent (in yellow) is a second cluster pertaining to public knowledge and the role of

news media and Al Gore's film. Just above this (in blue) is a cluster of topics pertaining to federal and state-level bureaucratic politics related to climate change. Lastly, toward the top (in black) is a cluster of topics related to energy industry concerns, and alarm about the economic and political costs of enacting climate change policy. These four broad clusters provide a comprehensive profile of all discourse produced by climate contrarian organizations from 1993 to 2013.

Lastly, Fig. 4 examines how the actual prevalence of these topics might be influenced by corporate funding. It focuses on four topics that have been particularly salient to the recent history of climate contrarianism (for all topic plots, see *SI Appendix*, Fig. S1). These interaction plots allow the prevalence of the topics to vary by time (by year, 1993–2013) and whether an organization received corporate funding during this same timespan (Red Line = Yes, Black line = No). For example, the plot in the top left demonstrates that temperature trends became much more salient between 2007 and 2013 for organizations who received funding, whereas they became less salient for organizations who did not. Similarly, the issue of energy production (top right), the positive benefits of CO_2 (bottom left), and climate change being a long-term cycle (bottom right) all showed distinctly positive trends over time for organizations who received corporate funding, whereas those who did not receive corporate funding remained more constant over the same time period.

On the whole, these findings show that despite being part of the same contrarian network, the organizations who received corporate funding actually produced discourse that was qualitatively different from organizations that did not receive such funding, and these differences tended to revolve around energy

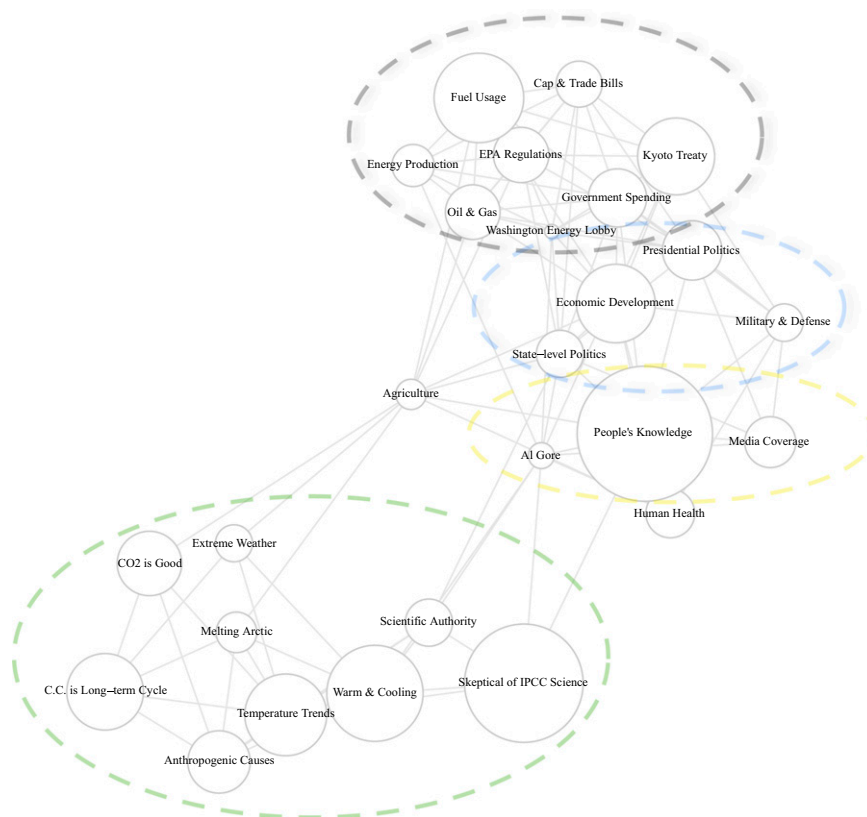**Fig. 3.** Graphing positive correlations between topics from Fig. 2. Topics near each other, and with a tie, indicate that they are more likely to be discussed within a document. Node sizes correspond to the topic proportions from Fig. 2 and are graphed using the Fruchterman–Reingold algorithm. Colored ellipses were added after graphing to point the reader toward the emergence of four distinct clusters.
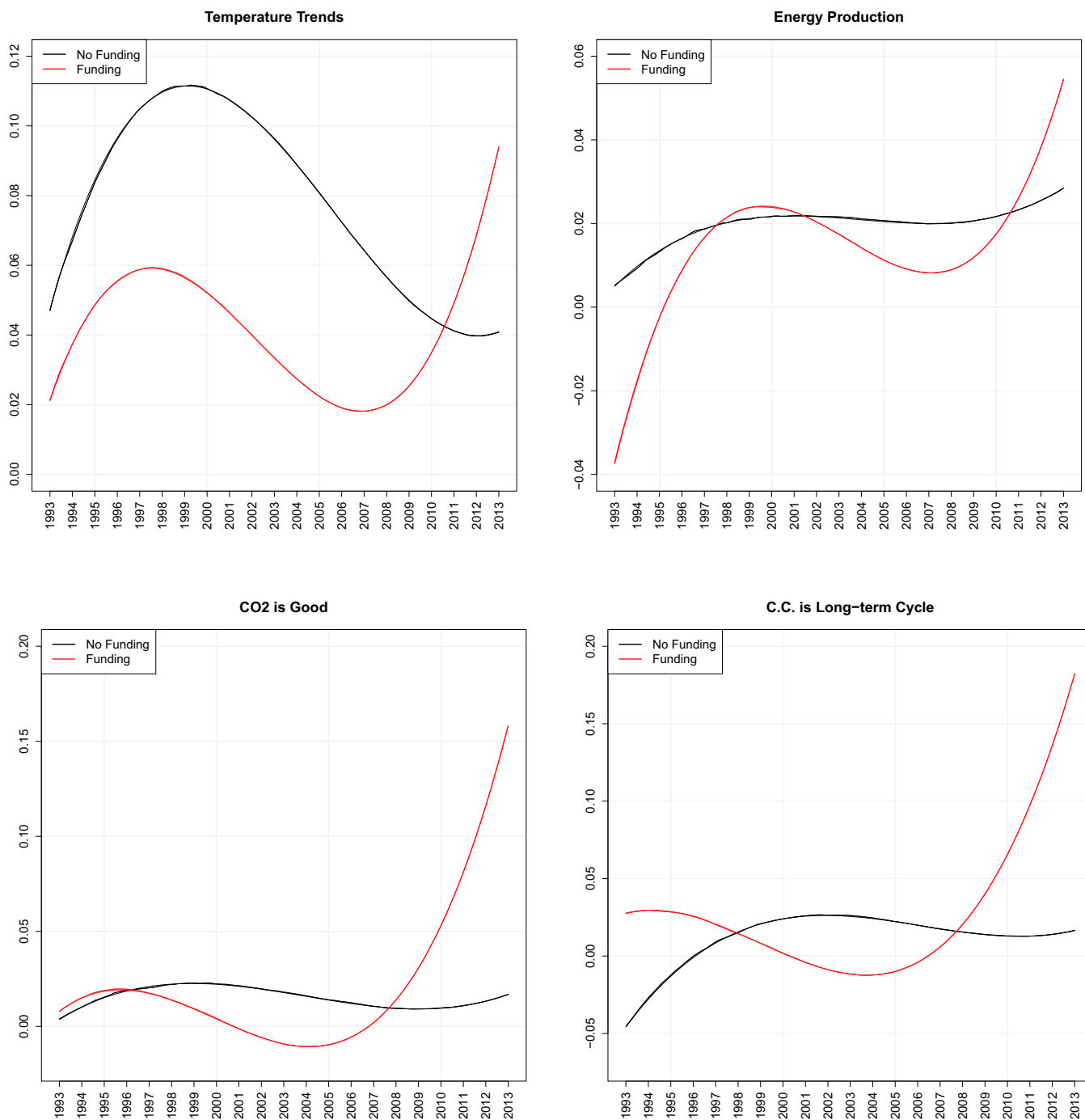
Fig. 4. The influence of corporate funding on themes within the climate change contrarian movement, 1993–2013. The *y* axis indicates how much a topic was written about. The red line represents the prevalence of the topic in the texts of contrarian organizations who received money, and the black line represents the prevalence of the topic for contrarian organizations who did not receive money. Interaction plots of all other topics are provided in *SI Appendix*.

production, debates about the effects of CO_2, and questions about the scientific veracity of long-term climate change.

Discussion

The main empirical and theoretical contribution of this analysis is that corporate funding influences the actual language and thematic content of polarizing discourse. These effects were visible over time, as the prevalence of certain thematic content shifted within the 20-year span of textual data. A secondary finding of this analysis is that organizations that received corporate funding were

more likely to have written and disseminated contrarian texts. Taken together, these findings are especially important because they contribute comprehensive empirical evidence to confirm what has widely been thought to be the case about climate change knowledge and politics, but has heretofore not been demonstrated scientifically with robust data.

While these findings are fundamental to understanding the past and present of climate change policy in the United States, they also have much broader implications for understanding ideological polarization itself. It is well understood that polarization is an

effective strategy for creating controversy and delaying policy progress, especially around environmental issues. However, what is less understood are the complex organizational and financial systems that affect the creation of ideological polarization in the first place. Thus, a broader contribution of this paper is to uncover empirically the actual social arrangements within which large-scale scientific (mis)information is generated, and the important role private funding plays in shaping the actual ideological content of scientific information that is written and amplified.

A notable limitation of these findings concerns the furtive nature of modern foundation funding and political action committees (8, 9, 11, 22), which creates a causal dilemma about whether corporate funding leads to the increased production of discourse, or whether the organizations already creating discourse attracted corporate funding. Of course, the strong correlational evidence provided here cannot untangle with certainty all of the microlevel causal processes, but it can provide traction to reject competing explanations. Most importantly, supplementary analysis revealed few discursive differences between organizations that received money before producing discourse (e.g., "front groups") versus organizations who received it later (e.g., established think tanks). Thus, the observed effects of money on discourse cannot be explained by recourse to one static temporal model or another, but funding influenced organizations who were already writing texts, as well as organizations who were established with the help of such funding. The organizations who were already in operation also created new texts, established and strengthened a relationship to promulgate shared interests, and ultimately amplified the messages that were central to the countermovement. Investigating how this happens at the microlevel is beyond the scope of this paper, and the data cannot address such questions. Future research should continue to develop these lines of inquiry about the multiple ways in which funding operates. Continued social-scientific research is also needed to further develop the scholarly understanding about the sources and influences of polarization on sustainability efforts—both presently with climate change, but also for predicting and preempting future socio-political conflicts from becoming fraught with uncertainty, skepticism, and misinformation.

Lastly, this paper provides a novel analytical model for future research to rigorously capitalize on the rise of the Internet and the unprecedented flood of available social scientific data, much of which is (*i*) textual and (*ii*) networked in some way. Unfortunately, much of the early analyses of large amounts of naturally occurring and digitized text data were first restricted to experts with the technical expertise to collect and organize it (e.g., computer scientists), resulting in "big" data content analyses that lacked theoretical depth or meaningful focus (35). This paper provides a model that focuses simultaneously on the fine-grained discursive content of texts and at the same time the ways in which these texts are embedded and produced within larger, complex, and densely connected socio-political networks. This analytical approach, as modeled here, allows researchers to examine how discursive and ideological content varies within and across organizations in a relational field, with a specific focus on the influence of covariates such as external funding and other organizational attributes (e.g., assets, mission focus, location, board interlocks). Thus, combining cutting-edge machine learning techniques with the long history of network science approaches, results in a robust empirical framework that allows researchers to situate fine-grained discursive results within the larger web of political and economic relationships that will continue to influence science and policy into the future.

Materials and Methods

Comprehensive information about the data collection, cleaning, and statistical analyses are provided in *SI Appendix*.

1. Gromet DM, Kunreuther H, Larrick RP (2013) Political ideology affects energy-efficiency attitudes and choices. *Proc Natl Acad Sci USA* 110(23):9314–9319.
2. McCright AM, Dunlap RE (2011) The politicization of climate change and polarization in the American's public view of global warming, 2001-2010. *Sociol Q* 52(2):155–194.
3. Ding D, Maibach EW, Zhao X, Roser-Renouf C, Leiserowitz A (2011) Support for climate policy and societal action are linked to perceptions about scientific agreement. *Nature. Clim Change* 1:462–466.
4. Leiserowitz A, Maibach EW, Roser-Renouf C, Smith N, Dawson E (2012) Climategate, public opinion, and the loss of trust. *Am Behav Sci* 57(6):818–837.
5. Scruggs L, Benegal S (2012) Declining public concern about climate change: Can we blame the great recession? *Glob Environ Change* 22(2):505–515.
6. Hart PS, Nisbet EC (2011) Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communic Res* 39(6):701–723.
7. Gauchat H (2012) Politicization of science in the public sphere: A study of public trust in the United States, 1974 to 2010. *Am Sociol Rev* 77(2):167–187.
8. Oreskes N, Conway EM (2010) *Merchants of Doubt* (Bloomsbury, London).
9. Coll S (2012) *Private Empire: ExxonMobil and American Power* (Penguin Group US, New York).
10. Brulle R, Antonio R (2011) The unbearable lightness of politics: Climate change denial and ideological polarization. *Sociol Q* 52(2):195–202.
11. Jacques PJ, Dunlap RE, Freeman M (2008) The organisation of denial: Conservative think tanks and environmental scepticism. *Env Polit* 17(3):349–385.
12. McCright AM, Dunlap RE (2000) Challenging global warming as a social problem: An analysis of the conservative movement's counter-claims. *Soc Probl* 47(4):499–522.
13. Elsasser SW, Dunlap RE (2013) Leading voices in the denier choir: Conservative columnists' dismissal of global warming and denigration of climate science. *Am Behav Sci* 57:754–776.
14. McCright AM, Dunlap RE (2003) Defeating Kyoto: The conservative movement's impact on US climate change policy. *Soc Probl* 50(3):348–373.
15. Brulle RJ (2013) Institutionalizing delay: Foundation funding and the creation of U.S. climate change counter-movement organizations. *Clim Change* 122:681–694.
16. Boykoff MT, Boykoff JM (2004) Balance as bias: Global warming and the US prestige press. *Glob. Environ. Chang.* 14:125–136.
17. Akerlof K, Rowan KE, Fitzgerald D, Cedeno AY (2012) Communication of climate projections in US Media amid politicization of model science. *Nat Clim Chang* 2: 648–654.
18. Nature Climate Change (2015) The IPCC news circuit. *Nat Clim Change* 5:281.
19. Boykoff MT (2014) Media discourse on the climate slowdown. *Nat Clim Chang* 4: 156–158.
20. Boykoff MT (2013) Public enemy No. 1?: Understanding media representations of outlier views on climate change. *Am Behav Sci* 57:796–817.
21. Neil GT, Marshall T (2011) Mediated climate change in Britain: Skepticism on the web and on television around Copenhagen. *Glob Environ Change* 21:1035–1044.
22. Walker ET (2009) Privatizing participation: Civic change and the organizational dynamics of grassroots lobbying firms. *Am Sociol Rev* 74(1):83–105.
23. Richter BK, Samphantharak K, Timmons JF (2009) Lobbying and taxes. *Am J Pol Sci* 53(4):893–909.
24. Citizens United v. Federal Election Commission (2010).
25. Kristal T (2010) Good times, bad times: Postwar labor's share of national income in capitalist democracies. *Am Sociol Rev* 75(5):729–763.
26. Morris M, Western B (1999) Inequality in earnings at the close of the twentieth century. *Annu Rev Sociol* 25:623–657.
27. Union of Concerned Scientists (2013) *Global Warming Skeptic Organizations* (Union Concerned Sci, Cambridge, MA).
28. Union of Concerned Scientists (2006) Smoke, Mirrors and Hot Air. How ExxonMobil Uses Big Tobacco's Tactics to Manufacture Uncertainty on Climate Science (Union Concerned Sci, Cambridge, MA).
29. Easley D, Kleinberg J (2010) *Networks, Crowds, and Markets: Reasoning About a Highly Connected World* (Cambridge Univ Press, New York).
30. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet Allocation. *J Mach Learn Res* 3: 994–1022.
31. Roberts ME, et al. (2014) Structural topic models for open-ended survey responses. *Am J Pol Sci* 58(4):1064–1082.
32. Roberts ME, Stewart BM, Tingley D (2014) stm: An R package for the structural topic model. R package 1.0.12.
33. McCarthy JD, Zald MN (1977) Resource mobilization and social movements: A partial theory. *Am Sociol Rev* 82(6):1212–1241.
34. Bail C (2012) The fringe effect: Civil society organizations and the evolution of media discourse about Islam since the September 11th attacks. *Am Sociol Rev* 77(6):188–205.
35. Bail C (2014) The cultural environment: Measuring culture with big data. *Theory Soc* 43:465–482.