Hall of Mirrors: Corporate Philanthropy and Strategic Advocacy

Marianne Bertrand, Matilde Bombardini,
Raymond Fisman, Brad Hackinen, Francesco Trebbi*

June 2021

Abstract

Information is central to designing effective policy, and policymakers often rely on competing interests to separate useful from biased information. We show how this logic of virtuous competition can break down, using a new and comprehensive dataset on U.S. federal regulatory rule-making for 2003-2016. For-profit corporations and non-profit entities are active in the rule-making process and are arguably expected to provide independent viewpoints. Policymakers, however, may not be fully aware of the financial ties between some firms and non-profits – grants that are legal and tax-exempt, but hard to trace. We document three patterns which suggest that these grants may distort policy. First, we show that, shortly after a firm donates to a non-profit, that non-profit is more likely to comment on rules on which the firm has also commented. Second, when a firm comments on a rule, the comments by non-profits that recently received grants from the firm's foundation are systematically closer in content to the firm's own comments, relative to comments submitted by other non-profits. Third, the final rule's discussion by a regulator is more similar to the firm's comments on that rule when the firm's recent grantees also commented on it.

* Bertrand: University of Chicago Booth School of Business, NBER and CEPR; Bombardini: University of California Berkeley Haas School of Business, NBER and CEPR; Fisman: Boston University and NBER; Hackinen: Western University Ivey School of Business; Trebbi: University of California Berkeley Haas School of Business, NBER and CEPR. We would like to thank Ernesto Dal Bo, Kevin Milligan, Matt Gentzkow, Larry Rothemberg and seminar participants at BI Norwegian Business School, University of California Berkeley Haas School of Business, Duke University, Harvard Kennedy School, CIFAR IOG, 2021 CEPR Conference The Politics of Regulation and Central Banking, Western University, University of Rochester, Stanford GSB, and UBC for discussion. Bombardini and Trebbi acknowledge financial support from CIFAR and SSHRC. Kelly Crabtree, Ugonna Eze, Adi Jahic, Daniel Sosa, Jack Vincent, and Andrew Zeller provided excellent research assistance.

1 Introduction

Economists and political scientists have long studied – both theoretically and empirically – the role interest groups play in the formation of laws and regulations (Olson, 1965; Grossman and Helpman, 2001). In the U.S., as in many democracies, there are well-established channels through which interest groups try to influence the laws and rules that may impact their communities, their businesses, and society at large. Through means such as lobbying, grassroots campaigns, testimonies, and public advocacy, interested parties inform politicians and bureaucrats of the costs and benefits of government action.

While interest groups may have expertise on topics of direct relevance to them, they may also be tempted to present information that is tainted by self-interest. This logic is at the core of the literature on informational lobbying.[1] Government officials must therefore weigh both the quality of information and its impartiality, based in part on its source. As such, policymakers may view information provided by for-profit corporations as less credible if that information is not corroborated by other groups with non-aligned interests. Non-profit organizations often represent interests that are unaligned with business.[2] Some non-profits – such as research groups and think tanks – are providers of nonpartisan, technical expertise and are commonly expected to offer a more neutral perspective. Other non-profits – such as environmental groups, social welfare organizations, and advocacy groups – may have opposing interests to business, to the extent that laws or regulations that benefit their members constrain business profits. Overall, non-profit organizations may therefore play an important balancing role in the informational lobbying process.

This role can be affected, or even subverted, however, by the financial ties between corporations and non-profits, when unbeknownst to government regulators and lawmakers.

There exists extensive anecdotal evidence that such concerns are well-founded, as journalists and researchers have uncovered numerous instances of firms using charitable contributions to co-opt ostensibly neutral and even non-aligned non-profits across a range of issues and regulatory agencies. Many of these examples involve persuasion-via-donation in public health debates. Jacobson (2005) describes a ("no-strings attached") $1 million donation from Coca-Cola Foundation

[1] By informational lobbying, we refer to the broad literature on strategic information transmission, which encompasses cheap talk and costly signaling models in the context of lobbying. For a complete discussion, see chapters 4-6 in Grossman and Helpman (2001). Early examples include Potters and Van Winden (1992), Austen-Smith (1993), Austen-Smith (1995) and Lohmann (1995).

[2] As Rose-Ackerman (1996) suggests for interactions with consumers, a rationale is that they "may favor non-profits because they believe that they have less incentive to dissemble because the lack of a profit motive may reduce the benefits of misrepresentation." Easley and O'Hara (1983) also emphasize the role of informational asymmetries. However, ameliorating informational problems is only one of the benefits of not-for-profit status. Other organizational rationales are explored in Glaeser and Shleifer (2001) and Glaeser (2002).

to the American Association of Pediatric Dentistry (AAPD). The gift was accompanied by a shift in the tone of AAPD statements on sugary beverages, from describing soft drinks as "a significant factor" in tooth decay, to describing the scientific evidence of the relationship as "unclear."[3] Similar concerns have been raised with respect to the role of donations from corporations to university research hospitals.[4] A second set of examples comes from oil, chemical, and utility companies' opposition to more stringent environmental regulations. A noteworthy set of cases involved utility companies' provision of financial support to local chapters of the NAACP, then soliciting their support in pushing for fossil-fuel-friendly regulations (EPI, 2019). The practice was sufficiently widespread that the NAACP national office issued a white paper describing – and denouncing – such practices. We provide further detail on these and other case studies in Section 7.

The context of U.S. Federal Regulation, with its far-reaching economic implications and its carefully documented record of communication between private organizations and government agencies, offers an ideal setting to establish evidence pertinent to the interactions of for-profit and not-for-profit entities vis-à-vis the government. U.S. federal agencies are legally required to publish proposed rules in the Federal Register, accept public comments on those proposed rules, and consider these comments before rules are finalized.[5][6] While there is no legal requirement for agencies to act on feedback received in the comments, the agencies themselves often attribute changes between proposed and final rules to arguments made via rule-making (Yackee, 2019).[7] As emphasized by Sunstein (2012), public commentary is also a valuable source of feedback to preempt regulatory mistakes "*when the stakes are high and the issues novel.*" We focus on this environment for our analysis.

---

[3]A more direct link to policy can be found in the soda industry's efforts against New York City's ban on large sugary drinks in the 2010s. In his decision to strike down the Bloomberg administration policy, the presiding judge cited amicus briefs filed by two New York non-profits (the local chapter of the NAACP and the Hispanic Federation), which argued that the ban would disproportionately affect ethnic and racial minority groups. Both non-profits were recipients of funds from Coca-Cola and PepsiCo. See "Minority Groups and Bottlers Team Up in Battles Over Soda," *The New York Times,* March 12, 2013. Aaron and Siegel (2017) show that 95 national public health organizations received funding from Coca-Cola and PepsiCo during 2011-2015; the study does not, however, look at the effect on organizations' publicly stated positions.

[4]See, for example, Gardiner Harris, "Top Psychiatrist Failed to Report Drug Income," *The New York Times,* October 3, 2008; Charles Piller and Jia You, "Hidden conflicts? Pharma payments to FDA advisers after drug approvals spark ethical concerns," *Science News,* July 5, 2018. See also Ross et al. (2000).

[5]The Administrative Procedures Act of 1946, 5 U.S.C. 553(c) states: "*. . . the agency shall give interested persons an opportunity to participate in the rule making through submission of written data, views, or arguments with or without opportunity for oral presentation. After consideration of the relevant matter presented, the agency shall incorporate in the rules adopted a concise general statement of their basis and purpose.*" https://www.law.cornell.edu/uscode/text/5/553. Last accessed 5/1/2021.

[6]There are some exceptions for urgent actions or cases in which the change is so trivial that the agency does not expect comments, but in general, agencies which fail to publish a sufficiently informative proposal or fail to follow the commenting procedure can have their regulations vacated in court.

[7]For instance, the U.S. Food and Drug Administration states on their website: "*these suggestions can, and do, influence the agency's actions*". See https://www.fda.gov/drugs/drug-information-consumers/importance-public-comment-fda. Last accessed 5/1/2021.

The government repository, regulations.gov, provides the largest source for comment information on proposed rules. Our comprehensive dataset includes the vast majority of the comments submitted in the rule-making process since 2003 and all related regulatory material. For each comment, we observe the proposed rule pertinent to that document, the identity of the commenter, as well as the content of the comment itself. We use natural language processing and machine learning tools (most of them customized to our environment) to standardize, clean, and analyze the corpus of all the comments and rules in our sample.

We complement the commentary data with information on corporate foundations and their beneficiaries, using data on charitable donations by foundations linked to corporations in the S&P 500 and Fortune 500 between 1995 and 2016 through detailed tax forms filed with the Internal Revenue Service (IRS).

We document three robust patterns. First, we show that non-profits are more likely to comment on the same regulation as their donors, and that this "co-commentary" is most strongly associated with donations in the year immediately preceding the comments. This result survives the inclusion of firm-grantee fixed effects and hence controls for the general tendency of some firm-non-profit pairs to be both financially connected and active on similar regulatory issues. The effect is large: a donation in the preceding year is associated with a 76% increase in the likelihood of co-commentary. It should be noted that co-commentary is not a rare event: about 10% of the average firm's comments have a co-comment by grantees they recently supported.

In our second set of results, using natural language processing tools, we show that the content of comment pairs from firms and non-profits linked via charitable donations tend to be more similar relative to any other pairs of comments on the same proposed rule. Importantly, the timing of this relationship parallels that of our first set of findings: co-comments in the year immediately following a donation are the most similar, even controlling for the average tendency of a given grantee-firm pair to share similar language. We also investigate the semantic orientation of the comments and show that the comment similarity for firm-grantee pairs does not result from comparably worded comments that express opposing sentiment.

Our third main empirical finding is that co-commenting relationships matter for the rules eventually finalized in the U.S. Code of Federal Regulations. Focusing on all comments made by firms in our dataset, we show that, if the recipient of a recent donation commented on the same proposed regulation as its donor firm, the language of the agency discussion of the final rule is more closely aligned with the firm's comment relative to the comments of other firms. This result is also confirmed when we focus on whether the regulator cites that specific firm in its discussion of the final rule. At the very least, it appears that the firm is able to obtain more attention from the regulator in finalizing the rule.

The welfare consequences of the patterns we document depend crucially on the theoretical mechanism that produces them. We believe there are two primary theoretical interpretations of our findings that warrant discussion:

(i) A "comments-for-sale" view offers the least benign interpretation (in social welfare terms) of our results. Grantees may be simply be "for sale" and willing to change the content of their comments to regulators in exchange for corporations' financial support. Under this interpretation, donations buy comments of certain non-profits. Some of the examples discussed above and in our case study analysis in Section 7 underscore this mechanism.

(ii) A "comments facilitation" view is more benign. Donations may serve to relax the budget constraints of selected grantees. As new regulations are proposed, a firm precisely targets donations toward non-profits that happen to be aligned with its interests at that particular point in time. This funding does not result from an expectation that grantees will change the content of their comments in a quid-pro-quo sense, but because the firm wishes to financially buttress non-profits presenting an independently similar viewpoint to regulators.

We make two observations on this second, more benign mechanism. First, in Section 4 we observe a greater similarity in co-comments between a firm and its grantees following a donation, even relative to the average co-comments made by the same pair when not immediately preceded by a donation. This pattern is also observed when looking within a relatively narrow set of regulatory issues. We acknowledge that these findings admit the possibility that, even within a narrow category of issues, a firm may support non-profits only when a topic of particular alignment suddenly arises. However, the likelihood of such precise targeting needs to be taken into account in evaluating its plausibility. Second, there still may be negative welfare consequences under this more benign interpretation if the donation affects the *probability of commenting*. Even absent a change in the content of comments, when regulatory agencies are not aware of the financial ties between firms and grantees, they misread the signal from a grantee's decision to comment. One can show that, as long as the regulator has a less than perfect knowledge of these financial ties (a realistic assumption given the complexity of the data), welfare losses are to be expected under theoretically plausible circumstances.[8] Firms appear aware of this mechanism. In leaked documents describing Monsanto's funding of grantees that would advocate against the banning of its controversial pesticide, Roundup, a Monsanto executive states that *"the key will be keeping Monsanto in the background so as not to harm the credibility of the information."*[9]

[8]A parsimonious theoretical framework in the working paper version of our article (Bertrand et al., 2018) illustrates this point. These results do not hinge on the outright distortion of the stance of beneficiary non-profits, but result from the selective subsidy of communications only offered to a favorable subset of third-party advocates. This simple framework also shows conditions under which welfare losses from subsidizing non-profit commentary may be less of a concern and when they can be ameliorated by disclosure.

[9]Monsanto. Email 11/30/2010, re: Questions. Available at: https://usrtk.org/wp-content/uploads/2016/01/SachsAR.pdf. See also Gillam (2017) for a discussion.

Our findings, first and foremost, provide a contribution to the literature on the mechanisms by which interest groups seek to influence government policy (for canonical early contributions see, for example, Grossman and Helpman (1994, 2001) and for a more recent discussion Baumgartner et al., 2009; Bertrand et al., 2014; Drutman, 2015). We differ from much of this prior work in our focus on influence via expert commentary, rather than through financial contributions and, much more importantly, in documenting one mechanism by which private interests may cloak biased advice by inducing its provision by a non-obviously aligned party. This finding has implications for how we model the process of governmental information acquisition (Austen-Smith, 1993; Laffont and Tirole, 1993), and is also of direct policy relevance for corporate disclosure requirements (Bebchuk and Jackson, 2013; Peng, 2016).

Our work is also related to prior research that has shown the value of coalitions of diverse interest groups in the adoption of government policy. The benefits of counteracting advocacy have an established rationale within information economics and political economy. Early theoretical explorations include Becker (1983), Austen-Smith and Wright (1994), Dewatripont and Tirole (1999), and Krishna and Morgan (2001). Empirical applications include work focused on the rule-making phase of Title IX of the Dodd-Frank Act of 2010 (Gordon and Rosenthal, 2018). In another study on legislation introduced in Congress between 2005 and 2014, Lorenz (2020) shows that bills supported by interest-diverse coalitions are more likely to receive committee consideration; in contrast, Lorenz (2020) finds no association between committee consideration and lobbying coalitions' size or their interests' PAC contributions. Generalizing beyond the lawmaking process, this prior work complements our findings, in that it suggests that corporations can expect some return for the type of charitable "investments" we uncover in this paper.[10]

From a welfare perspective, we wish to understand potential subversion of the regulatory and rule making process due to distortions in information and beliefs. These are concerns that add to issues of pure regulatory capture (Stigler, 1971; Peltzman, 1976) and are complementary to issues of enforcement vis-à-vis the courts (Glaeser and Shleifer, 2003). Our analysis may also contribute to understanding the complex problem of cognitive or cultural capture of regulators, highlighted by Johnson and Kwak (2010) and Kwak (2014), in providing a mechanism through which regulators' and special interests' beliefs become more strongly aligned.

Finally, our paper expands on earlier work highlighting how corporations may strategically use their corporate philanthropy as an undisclosed tool of political influence. Bertrand et al. (2020) show that corporations allocate more of their charitable giving to congressional districts that are more relevant to the corporations as a result of the committee assignments of their elected representatives. We identify in this paper another, independent, mechanism for "strategic"

[10]Other papers that focus on the returns to lobbying include Bombardini and Trebbi (2011, 2012); Kang (2016); Kang and You (2016).

corporate philanthropy (Baron, 2001) in the government arena.[11]

2 Institutional context and the data

2.1 Rulemaking process

The rule-making process of U.S. federal agencies provides a context in which we may observe both the presence and the content of communication by different entities with interests in influencing the policymaker. While lobbying at the federal or local level does not come with statutory requirements of disclosure of the content or even the exact target of communication, the rule-making process consists of a series of codified procedures that regulate the activity of federal agencies in the production of "rules" under the Administrative Procedure Act (APA) of 1946.[12]

The subject of policy deliberation is a rule "*designed to implement, interpret, or prescribe law or policy,*" according to the APA. The rule-making process may be set in motion by the passage of a new law in Congress, which then requires implementation, or by an agency itself, upon surveying its area of legal responsibility and identifying areas that need new regulations.[13] The rule-making process starts with a Notice of Proposed Rulemaking (NPRM), which includes the objective of the rule and how it would modify the current Code of Federal Regulations. The NPRM is published in the Federal Register, at which point the agency specifies a period of 30 to 60 days during which the public can submit comments on the proposed rule.[14]

This notice and comment process is designed to alleviate the informational problem in federal regulatory agencies. These provisions, explicitly delineated in the APA, are fundamental to U.S. public administration rule-making (Strauss, 1996), and provide an opportunity for protection of consumer and private interests in an environment in which regulators are typically non-elected and not directly accountable to voters (Besley and Coate, 2003).

After comments have been received and additional information collected, the agency may proceed to publish a final rule in the Federal Register or issue a Supplemental Notice of Proposed

[11] See Kitzmueller and Shimshack (2012) for a broader review of corporate philanthropy and corporate social responsibility.

[12] Under the Lobbying Disclosure Act of 1995, lobbying registration and reporting forms only require lobbyists to list the topic and the agency lobbied (e.g., Trade, the Senate of the United States), in addition to clients and payments. See Vidal et al. (2012); Bertrand et al. (2014).

[13] Agencies may decide to engage in rule-making under the recommendation of congressional committees, other agencies, or following a petition from the general public. Only about a third of rules originate via legislation; see West and Raso (2013).

[14] Some complicated rules may have much longer comment periods, as a result of multiple stages of the rule-making process. A rule may start, for example, with an Advance Notice of Proposed Rulemaking document, followed by an initial proposal, then perhaps an updated proposal, and then finally a rule. Each stage might have its own comment period, and the stages could extend over multiple years.

Rulemaking if the initial rule was modified substantially, in which case further comments are invited. This notice-and-comment procedure aims to include the general public and all interested parties in the crafting of the new rule. Importantly, accompanying the final rule, the agency also provides a discussion of the goals and rationale of the policy, and how the comments were incorporated into the final rule; this discussion is published in the rule's Supplementary Information section. Upon finalization of the rule, comments represent part of the official record, and rules can be challenged judicially on procedural or substantive grounds based on comments filed by entities that participated in the rule-making process. Judicial review is an important constraint to rule-making activity in the United States in that it effectively forces regulators to attend to opinions expressed via commentary.

2.2 Institutional context for firm–non-profit interaction

In accordance with the APA, regulators are required to weigh public interest in their rule-making decisions. Consequently, broad coalitions of multiple stakeholders may provide particularly relevant input into a regulatory agency's deliberations. Firms thus have an incentive to mobilize such coalitions to support their positions on specific rules. In the literature on lobbying, such coalitions have empirically shown a degree of success beyond the individual organization,[15] with a particular advantage accruing to more heterogeneous coalitions (Lorenz, 2020).[16]

A firm may plausibly enlist the support of a non-profit in the context of these public policy campaigns. As discussed in Bertrand et al. (2014), large corporations, such as the ones we study here, retain in Washington both in-house government relations specialists and lobbyists, which monitor government agencies on a daily basis. In anticipation of relevant regulatory or legislative activity, specialists and a host of firms' allies are activated (Baumgartner et al., 2009), including non-profits, in order to organize public policy campaigns. As discussed in the introduction, activating arms-length non-profits may be particularly beneficial to a firm, due to the tax exemption from charitable grants and lower disclosure requirements, which are both distinctive advantages relative to federal lobbying expenditures, for instance.

In the analysis that follows we consider the relationship between a firm and a given non-profit as captured by charitable grants, which may be used in the context of these campaigns (though we do not suggest that all corporate philanthropy is politically motivated). We focus on changes around regulatory actions within a firm-grantee pair, rather than on the composition of a firm's broad coalition of allies, because of the more precise identification that this within-pair variation affords to us (see Sections 4 and 5). In fact, such coalitions change issue by issue and are frequently covert (Mahoney and Baumgartner, 2015). Section 6 also investigates whether firms engaging a

[15]See, for example, Nelson and Yackee (2012); Bombardini and Trebbi (2012).

[16]See also DeGregorio (2010); Mahoney and Baumgartner (2015); Phinney (2017).

grantee through a donation receive more attention in an agency's discussion of a final rule than other firms commenting on the same rule.

2.3 Data

Charitable giving by foundations The starting point for our sample is the set of corporations that have appeared at any point during the period 1995 to 2016 in the Fortune 500 or S&P 500 lists, which collectively include 1,397 firms.[17] Data on charitable donations by corporate foundations come from FoundationSearch, which digitizes publicly available Internal Revenue Service (IRS) data on the 120,000 largest active foundations in the U.S. We find 645 active foundations that can be matched by name to 532 of the initial list of 1,397 firms, some of which have more than one foundation.[18]

Each charitable foundation must submit Form 990/990 P-F "Return of Organization Exempt From Income Tax" to the IRS annually, and this form is open to public inspection. Form 990 includes contact information for the foundation, as well as yearly total assets and total grants paid to other organizations. Schedule I of Form 990, entitled "Grants and Other Assistance to Organizations, Governments, and Individuals in the United States," specifically requires the foundation to report all grants greater than $5,000. For each grant, FoundationSearch reports the amount, the recipient's name, city and state, and a giving category created by the database.[19]

While the IRS assigns a unique identifier (Employer Identification Number, EIN) to each non-profit organization, Schedule I does not include this code, so we rely on the name, city and state information to match a grantee to a master list of all non-profits. This list, called the Business Master File (BMF) of Exempt Organizations, is put together by the National Center for Charitable Statistics (NCCS) primarily from IRS Forms 1023 and 1024 (the applications for IRS recognition of tax-exempt status). The BMF file reports many other characteristics of the recipient organization, including address, assets and non-profit sector code called the National Taxonomy of Exempt Entities (NTEE). The results of the matching between all public charities, private foundations or private operating foundations (designated as 501(c)(3) organizations for tax purposes) in the BMF and the recipients of charitable giving by Fortune 500 and S&P 500

[17] The initial number of firms is 1,434, but we combine firms that merge during the sample period.

[18] As noted in Brown et al. (2006), larger and older companies are more likely to have corporate foundations, which may partly result from the fixed cost of establishing a foundation. Thus, the channel of influence we uncover in our paper may be more readily available to larger firms, and further hamper the ability of smaller firms to compete on a level playing field. Brown et al. (2006) also find that state-level statutes – in particular laws relating to shareholder primacy and the ability of firms to consider broader interests in business decisions – predict establishment of a foundation. Various endogenous financial variables are also predictive of foundation establishment. The analysis in Brown et al. (2006) is cross-sectional, so their variables are absorbed by the various fixed effects in many of our analyses.

[19] The ten broad categories are: Arts & Culture, Community Development, Education, Environment, Health, International Giving, Religion, Social & Human Services, Sports & Recreation, and Misc Philanthropy.

companies is described in detail in Bertrand et al. (2020).

Finally, note that direct charitable giving by firms (that is, not through their charitable foundations) or large charitable grants by executives of the firms are unfortunately not traceable and are thus excluded from the analysis. As we emphasize in Bertrand et al. (2020), while influence via corporate foundation giving is hard to trace, direct giving is even more difficult to observe. We thus might expect that attempts at influence that the firm feels even more compelling to hide from view would occur via these other channels, and thus not show up in our analysis.

Comments and rules The source of data on regulatory comments is regulations.gov, a website through which the majority of U.S. federal agencies collect public comments in the notice-and-comment phase of rule-making.[20] The regulations.gov API provides a search function for document metadata, which allows us to identify all comments submitted and stored on the site. Our initial comment sample consists of all comments posted to regulations.gov in the years 2003-2016. We use a custom machine learning tool to extract organization names from the comment metadata. The algorithm identified 981,232 comments that appear to be authored by organizations (as opposed to private individuals) and we downloaded the full text of these organization comments. We are particularly interested in comments submitted by non-profits and by corporations that we observe in our FoundationSearch sample. The comments are linked to corporations' and grantees' names through a custom name matching tool that implements multiple types of fuzzy matching and manual corrections.[21]

Comments on regulations.gov are organized into folders called "dockets" created by agencies to hold documents related to a narrow topic, usually a single proposed rule or a sequence of rule-making documents that culminate in a final rule. For example, docket FNS-2006-0044 from the Food and Nutrition Service (FNS) contains only proposed rule 06-09136, "Fluid Milk Substitutions in the School Nutrition Programs." and the comments submitted regarding that proposal.[22] We rely on the agencies' classification and refer to each of these dockets on a homogeneous topic as a *rule*.[23]

In the last section of the paper, we examine the wording of the discussion of final rules as a function of corporate and non-profit comments. Rulemaking documents such as proposed rules,

[20] A detailed description of our data set construction is offered in Appendix A.

[21] Available at https://github.com/bradhackinen/nama.

[22] There are also complex dockets that contains multiple proposed rules and notices, but these are rare and still constitute a homogeneous topic. See, for example, docket EPA-HQ-OAR-2008-0699, the Environmental Protection Agency's review of the National Ambient Air Quality Standards for Ozone.

[23] Organizations sometimes submit multiple documents to a single docket in the same comment period. For example, when organizations spearhead mass letter writing campaigns, the number of unique documents associated with that organization can number in the thousands. To avoid giving excess weight to multiple submissions from the same organization, we count the entire set of documents submitted by one organization in response to a single rule in the same calendar year as a single "comment".

final rules, and notices are published in the Federal Register. We collect these documents in bulk XML format from the Government Print Office website, and obtain additional identifiers and metadata from the federalregister.gov website API.

Linking comments to specific rules requires additional steps, which we describe in more detail in section 5 and Appendix A in the online supplemental material. Appendix B describes the tools we deploy in our text analysis of the comments.

Basic data facts Recall that our sample starts with the set of companies that appeared at least once in the Fortune 500 or S&P 500 lists between 1995 and 2016. Of the 1,397 firms in that sample, we find 892 that have commented at least once in the period 2003-2016.[24] This is the sample of firms that forms the basis of our regressions. We have a total of 16,008 firm comments over 5,438 rules. Of these 872 firms, 532 have a foundation. To generate the set of non-profits for our analysis, we start from the 212,797 entities that received at least one grant from any foundation in our sample over the period 2001-2016. Our sample consists of the 11,002 of these grantees that comment at least once at any point during the period 2003-20016. This restriction excludes the set of non-profits that never receive any grants *and* never comment. We make this restriction in order to make the combinatorics for firm-grantee pairs tractable in terms of total number of observations, without losing any non-profits that are active in notice-and-comment rulemaking. For our sample of grantees that do comment during our sample period, we have a total of 52,488 comments on 8,018 rules.

There is vast heterogeneity among firms in their activity in the commenting phase. The most actively commenting firm, Boeing, provided comments on 1,174 rules. On average each firm comments on almost 17 rules, but the distribution is skewed: the median firm comments on 5 rules, while the firms at the first and third quartile comment on 2 and 16 rules, respectively. The distribution of comments among grantees is even more skewed. On average each grantee comments on almost 5 rules, but the median is 1 and the third quartile is 3 rules. The most active grantee (The Center for Biological Diversity) comments on 816 rules.

Appendix Table C.3 lists the agencies that receive the highest number of comments from grantees and firms.[25] At the top of the list for firms are the EPA (Environmental Protection Agency), the FAA (Federal Aviation Administration) and the FDA (Food and Drug Administration). The top three agencies as recipients of grantees' comments are the EPA, CMS (Centers for Medicare & Medicaid Services) and FDA.

Tables 1 and 2 provide summary statistics for 2008-2014 (the period during which our data are most complete) on firm and grantee commenting and what we define as "co-comments," which

[24]We only consider comments starting in 2003 because this is when the comments database is complete.
[25]Agency acronyms are listed in Appendix tables A.2 and A.3.

are instances in which firms and grantees comment on the same rule. Table 1 summarizes the firm side: there are 1,457.8 comments by firms in an average year (made by an average number of firm commenters of 384.4 per year, a figure not reported in the table). On average, a firm comments on 1.9 rules per year. Of these rules, 1.3 received comments from non-profits. Of particular interest is the further subset of 0.3 rules that received comments from the firm's grantees (the number is 0.2 if we consider grantees that received recent donations[26]). Overall, about 10% of the average firm's comments have a co-comment by grantees they recently supported.

Table 2 presents the analogous breakdown of commenting for grantees. We note that, of the average annual number of comments (5,073 from 2,516.7 annual grantees, the latter figure unreported in the table), 1,255.6 (almost 25%) come from grantees that have received at least one donation from our sample of firms, and 645.6 (almost 13%) come grantees that received a recent donation. It is interesting to compare the total number of annual comments by firms (1,457.8) to the number of comments by recent grantees (645.6) which, as we will see, submit comments with similar content.

Finally, Table 3 presents annual donations, which average $9 million per firm, and the donations associated with grantees that comment on the same rules as the firm, which average $700,000. The average firm contributes 8% of its funds to grantees who comment on the same rules (16% to grantees commenting to the same agency). We can conclude that co-commenting represents a meaningful share of both firms' and grantees' activity. Appendix Tables C.1 and C.2 report the same firm commenting and co-commenting quantities for rules that have been classified as "significant" under Executive Order 12866, because of the scale of their impacts.[27] Significant rules make up approximately 10% of all rules that receive at least one organization comment, but they receive almost half of all firm comments. Within significant rules, for every five firm comments received by a regulator, the regulator also receives three comments from non-profits with a financial tie to the firms they are co-commenting with, roughly half of these involving a donation in the concurrent or previous year (i.e., a recent donation).

It is useful to compare the dollar amounts of these donations with federal lobbying expenditures, using a dataset maintained by the Center for Responsive Politics.[28] The amount that firms in our sample spent lobbying all federal institutions during our reference period (2003-2014) was $772 million per year. Assuming that those funds were split evenly among all of the institutions listed in each lobbying report filing, we obtain a rough estimate of $538 million per year spent by our sample firms lobbying our sample agencies. The equivalent estimate for the total amount of money donated to non-profits that co-comment with their donor firms is $251 million, or about

[26] A recent donation, as we discuss later, refers to a grant received in the year of the comment or the year before.

[27] One common reason for being classified as significant is that the rule has "an annual effect on the economy of $100 million or more."

[28] See https://www.opensecrets.org/federal-lobbying. Last accessed 5/4/2021.

47% of total federal lobbying expenditures. For an additional comparison, firm political action committee (PAC) campaign contributions in a typical congressional cycle average 10% of total lobbying expenditures, or about a fifth of the donations that we consider in this article.

3 Evidence based on charitable giving and non-profit commenting on regulations

This section focuses on the link between firms and non-profits through charitable grants, and establishes a relationship between firm-grantee financial ties and their tendency to comment on the same regulations. We denote firms/foundations by $f \in F$ and grant-receiving non-profits (grantees) by $g \in G$. The following analysis employs all firms and non-profits available in our datasets, which includes the 11,531 non-profits that receive at least one grant from any charitable foundation in our sample and that comment at least on one rule since 2003.

Let D_{fgt} be an indicator function that takes a value of 1 if we observe a donation from firm f to grantee g in year t, and 0 otherwise. The indicator function C_{frt} is equal to 1 if firm f comments on rule r in year t, and 0 otherwise. The indicator function C_{grt} is defined similarly and is equal to 1 if grantee g comments on rule r in year t, and 0 otherwise. We define $CC_{fgrt} = C_{frt} \times C_{grt}$ as an indicator equal to 1 when donor f and grantee g comment on the same rule r at time t. We adopt two types of specifications: a "co-commenting" specification and a "rule" specification.

3.1 Co-commenting specification

We begin by relating the event of a firm and a grantee commenting on the same rule to a recent financial tie between the two in the form of a charitable donation. In particular, we examine whether co-commenting is more likely in the year of, or year immediately following, a donation.

Let $CC_{fgt} = I\left(\sum_r CC_{fgrt} > 0\right)$ indicate whether firm f and grantee g comment on the same rule at time t. Our benchmark specification is:

$$CC_{fgt} = \beta_0 + \beta_1 D_{fgt-1} + \delta_{fg} + \delta_t + \varepsilon_{fgt} \tag{1}$$

where δ_{fg} indicates firm-grantee pair fixed effects, δ_t time fixed effects, and D_{fgt-1} is equal to 1 if we observe a donation from f to g in the year that is concurrent with (t) or preceding $(t-1)$ the comments, and 0 otherwise. We group together years t and $t-1$ donations due to the coarseness of the data along the time dimension. We only observe the year of a comment, so it is possible for a comment to be made in, say, January of 2006 and a donation in June 2006; hence we can only be certain that the lagged-year donation took place prior to co-commenting.[29]

The four columns in Table 4 report different sets of fixed effects in order of increasing stringency. In column (1) we only include time fixed effects δ_t, while in column (2) we include separate grantee, firm, and time fixed effects, which account for the average tendency of certain firms and grantees to be more active in grant-making and receiving, and also in commenting on rules.

One may still be concerned that the pattern of co-commenting may result from firms contributing to non-profits that share similar objectives and views, or non-profits that operate in similar sectors. For instance, the Bayer Science & Education Foundation associated with Bayer US, a pharmaceutical company, may be more likely to donate to healthcare-related research non-profits, and both Bayer and healthcare-related non-profits may be more likely to comment on healthcare-related regulations than an average organization. For this reason, our preferred specification in column (3) of Table 4 includes firm-grantee fixed effects and time fixed effects. In this specification, β_1 is estimated employing only within-pair variation over time in donations and co-commenting. In particular, β_1 will detect whether, controlling for the average tendency of a certain firm f to co-comment with and donate to a specific non-profit g, we observe co-comments occurring immediately after a donation from f to g. Column (4) is an even more demanding specification, as we introduce grantee-year and firm-year fixed effects, which control for firm- and grantee-specific changes in commenting and giving/receiving over time. Standard errors are clustered at the grantee-firm pair level for all columns.

We find a robust and economically significant association between recent donations and the likelihood of co-commenting. Co-commenting is sparse when considering all possible firm-grantee-year triples: 0.175% feature co-commenting. In column (3) a recent donation is associated with a 76% increase in the likelihood of co-commenting, even after controlling for the general propensity of a specific firm to give to as well as co-comment with a specific grantee. Even in the saturated specification of column (4), a recent donation increases the probability of co-commenting by 46%.

As a further robustness exercise, Appendix Table C.4 includes, along with dummies for donations at time t and $t-1$, a dummy for whether firm f donated to g in year $t+1$. The set of fixed effects in this table is analogous to Table 4. In column (4) of that table, with the most restrictive set of fixed effects (i.e. pair, grantee-year and firm-year fixed effects), we find that donations made immediately after the commenting period are not associated with co-commenting, whereas only immediately preceding donations are. This pattern further confirms the particular timing we emphasize here, with co-commenting more prevalent only after we observe a recent donation from firm to grantee (though it is theoretically possible that firms might reward non-profits only after comments are made, in which case we would observe a positive coefficient). Figure 1 illustrates

[29] In Appendix Table C.4 we separate contemporaneous and lagged donations and find that lagged donations strongly predict co-commenting, while contemporaneous donations are a weak predictor of co-commenting.

this intuition graphically, by applying an event study approach to the data. The figure displays the sharp increase in likelihood of co-commenting relative to the period before the donation.

3.2 Rule specification

In the specifications we have considered thus far, we have aggregated co-commenting across different rules at the firm-grantee-year (fgt) level. For robustness, we now present an alternative approach that allows us to control for the average level of commenting on a given rule r. This "rule" specification relates the probability of commenting by a grantee on r to donations received:

$$C_{gr} = \beta_0 + \beta_1 \underbrace{I\left(\sum_f D_{fg} \times C_{fr} > 0\right)}_{DonorComment_{gr}} + \delta_g + \delta_r + \eta_{gr}, \tag{2}$$

where C_{gr} is equal to 1 if g comments on rule r (0 otherwise) and $DonorComment_{gr} = I\left(\sum_f D_{fg} \times C_{fr} > 0\right)$ is equal to 1 if g receives a donation from any firm that comments on r, and 0 otherwise. In its most saturated version, this specification includes rule fixed effects δ_r, which capture the extent to which certain rules are subject to more intense commenting, and grantee fixed effects δ_g, to account for factors like resources and size of the non-profit, which may make g both more visible (to corporate donors) and more likely to comment on any rule.

Table 5 reports estimates of β_1 under different fixed effects and with two-way clustered standard errors at the grantee and rule level. Our preferred specification in column (4) has rule and grantee fixed effects. When considering all the possible pairs of grantees and rules, we find a comment in 0.043% of cases. It is not surprising that this number is small, since the universe of all possible grantee-rule pairings involve non-profits like, say, the Red Cross, that we would not expect to comment on, say, financial regulation. Starting from this baseline probability of commenting on a specific rule, we find that the probability that a non-profit comments on a particular rule is 3 to 5.5 times higher when a donor firm commented on the same rule, a quantitatively sizable result that accords with our previous results under specification (1).

4 Quantifying the similarity in content across regulatory comments

Thus far our analysis has demonstrated that financial connections between firms and non-profits are associated with an increase in the propensity to co-comment on the same rules. We now show that the content of non-profits' messages to regulators are also related to these non-profits' financial connections to firms.

To build intuition (and without intent to claim any deliberate deception by the parties involved in this particular instance), consider the example of Bank of America's $150,000 donation to the Greenlining Institute in 2010. Bank of America is the second largest bank in the United States by total assets and is a central player in housing finance; the Greenlining Institute is a non-profit focused on improving access to affordable housing and credit for low-income families and minorities. In 2011 both organizations commented on the Office of the Comptroller of the Currency's Credit Risk Retention (CCR) rule, Docket ID OCC-2011-0002 initiated under the Dodd-Frank Act of 2010 (Title IX, Subtitle D, Section 941). CCR, also known as the "skin in the game" rule, imposed a 5% retention requirement on all mortgage loans originated by lenders in the United States to moderate "originate-to-distribute" moral hazard problems pervasive in the build-up to the 2008 financial crisis. The main comment submitted by Bank of America[30] observed that, in relation to relaxing the definition of qualified mortgages exempted from retention requirements on the issuing bank's balance sheet (i.e., mortgages deemed safe enough to warrant exemption from the restriction): "*...the PCCRA provision will cause some borrowers to be unable to obtain a loan at all. In the currently tight private residential mortgage market, borrowers already must provide significant down payments.*" The Greenlining Institute provided a similar assessment in its comment,[31] expressing the opinion that "*by raising the barrier to affordable home ownership with an unreasonable 20% down payment requirement, we will not only keep families from rebuilding after foreclosure, but we will prohibit an entire generation of first time borrowers from owning a home, despite lower home prices across the country.*" In sum, both organizations appeared to advocate openly for laxer definitions of the CCR exemptions, limiting the rule's bite, and allowing assets with substantially lower quality and higher risk to be exempt.[32]

In this section, we provide a framework for examining the content and textual similarity of comments filed by non-profits and firms, and show that, upon receipt of a donation from a firm's foundation, comments by a non-profit are more similar to those of its donor, suggesting that the Bank of America-Greenlining example may hold more broadly in the data.

We compute approximate measures of semantic similarity of pairs of public comments using Latent Semantic Analysis (LSA) with bag-of-words features. LSA is an established technique in the natural language processing (NLP) literature and it has been shown to perform well on a variety of document classification and retrieval tasks.[33] In our own tests, we found LSA worked significantly

[30]Document ID OCC-2011-0002-0141

[31]Document ID OCC-2011-0002-0353

[32]These efforts ultimately succeeded in entirely defanging the rule. For a discussion, see Floyd Norris for the *New York Times*, October 23, 2014, Page B1 "Banks Again Avoid Having Any 'Skin in the Game'," available at https://www.nytimes.com/2014/10/24/business/banks-again-avoid-having-any-skin-in-the-game.html. Last accessed 5/4/2021.

[33]See Dumais et al. (1988) and Deerwester et al. (1990). For a discussion of latent semantic analysis, see Dumais

better than some alternatives on a benchmark classification task we developed with our data (see Appendix B for details). We also verified that we obtain very similar results when using Latent Dirichlet Allocation (Blei et al., 2003), another popular approach to modeling document similarity (see Appendix D). We proceed in three steps in the construction of our measures. First, we collect all comments from all organizations with at least two comments in all rules, and collapse the documents to organization-rule-year level observations by concatenating the text from all attachments and submissions from a single organization on a given rule in a particular calendar year. Next, we apply LSA to construct a document vector for each rule-year comment which summarizes the distribution of words in each comment. As is common in LSA, we use term frequency inverse document frequency (TF-IDF) weighting, to emphasize the importance of words which appear in a small number of documents. Finally, we construct a scalar similarity measure from the cosine angle between the document vectors corresponding to firm and grantee comments, and scale this measure to have a standard deviation of one across all firm-grantee co-comment pairs.

Our benchmark comment similarity specification is:

$$S_{fgr} = \beta_0 + \beta_1 D_{fgt-1} + \delta_{fg} + \delta_r + \varepsilon_{fgr} \tag{3}$$

where S_{fgr} is the similarity of comments of grantee g and firm f commenting on the same rule r finalized in year t, D_{fgt-1} is indicator variable that equals 1 if firm f donated to grantee g in either year t or year $t-1$ and 0 otherwise, and the coefficient of interest is β_1. As each rule r is finalized in a specific year t, year fixed effects are spanned by rule fixed effects and are therefore omitted. The dataset we employ for this analysis includes all possible firm-grantee pairs of comments conditional on commenting on the same rule r (note that this is a small subset of the firm-grantee-year data employed in the Table 4 analyses, since co-commenting is a relatively rare occurrence).[34]

The results for equation (3) with separate firm, grantee and rule fixed effects are presented in column (1) of Table 6. We find that firm and grantee comments are 4.7% of a standard deviation more similar after a recent donation.

One potential concern is that the results in column (1) are driven by firms preferentially donating to grantees that have more similar comments on average. We thus include a firm-grantee pair fixed effect in column (2). This specification, with more restrictive fixed effects, exploits only variation within a firm-grantee pair over time and thus measures whether the similarity of comments is higher than average *for a specific pair* when there is a recent donation linking the

(2004). All details for our analysis are in Appendix B.

[34] As a complement to the approach in equation (3), Appendix E reports results from a matching estimator that only uses comments from the most similar regulations to estimate the paired untreated counterfactual. The results are consistent with the evidence reported in this section.

two. A recent donation in this specification is associated with an increase in the similarity of comments by 6.1% of a standard deviation, a significant effect.

Even though we find similarity increasing after a recent donation in the fixed effect specification, it is conceivable that donations may happen only at the exact time when the firm and the grantee serendipitously agree on a specific topic of regulation. A more stringent bar to clear would be to hold the topic constant and test whether a non-profit's comments become more similar to those of the firm after receiving a donation, relative to their standard level of similarity when commenting on that specific topic. To put it differently, we would ideally assess whether a grantee changes its position on the identical topic on which it typically comments just after receiving a donation, along the lines of the Coca-Cola and AAPD example discussed in the introduction.

By construction, we do not have multiple comments on the same rule by the same entities. However, the specification in column (3) aims to approximate this thought experiment, by adding fixed effects for agency (a proxy for the topic) times sector (NAICS 6 digit code) of the firm times IRS's National Taxonomy of Exempt Entities Classification (NTEEC) code of the non-profit. This specification therefore exploits only variation in similarity and donations within a set of firms, grantees and issues that are homogeneous. We find that even in this specification, recent donations are associated with an increase in similarity.[35]

In columns (4)-(6), we maintain the specifications in columns (1)-(3) with an additional modification to the document vectors that is intended to correct for potential bias introduced by similarities in the firm's and grantee's commenting style. Here, we use the term "style" broadly to mean any aspect of the comment text that tends to be repeated across comments by the same organization. For example, there can be large differences in the amount of technical language and jargon employed by different commenters. Our solution is to control for each organization's style by subtracting their mean comment document vector from all of their comments before computing cosine similarities between document vectors (see Appendix B for details). The resulting similarity measure then focuses on the parts of comments that vary over time rather than fixed aspects of commenting style. We find that controlling for style in this way only increases the implied association between a recent donation and co-comment similarity.

In Appendix C we also present analyses that underscore the very specific timing of the link from donation to comment similarity. In particular, we modify our definition of donations to focus on the period immediately *after* the regulatory commenting phase. Appendix Table C.5 reports these results, using specifications that parallel those presented for the co-commenting results in Section 4. The estimated coefficient on future donations is much smaller in magnitude than

[35] Although not shown for the sake of brevity, most variation in results with different fixed effects is due to the regression specification rather than changes in the sample. The difference in results in columns (4) and (5) are one exception: the estimated change in similarity associated with a recent donation is 7.1% when using the specification from column (4) and sample from column (5).

that of recent donations, though for this set of results neither coefficient is generally statistically significant. If we run the same comment similarity regressions on future donations alone, the estimated coefficients are small and never statistically significant (in contrast to recent donations). This placebo exercise is informative along several dimensions. As future donations are close in time to the commentary activity, but statistically and economically insignificant, these findings further assuage the concern that our results may be driven by some underlying shared tendencies of firms and grantees operating in related areas. The systematic timing of excess similarity between comments' texts just following the disbursement of a charitable grant offers more support to the view that donations provide firms with some influence over grantees' expressed viewpoints.

It is natural to ask whether an increased similarity of the text of comments necessarily implies more similar positions on an issue. We construct a test to assess the possibility that firms and grantees may employ a similar terminology, while nonetheless delivering opposing messages to regulators. Our test is based on an analysis of comment sentiment, which relies on established NLP scholarship. Semantic orientation exercises are common in the NLP literature (e.g., the unsupervised classification of book reviews as positive or negative), including applications to economics and finance, for example in the classification of monetary policy announcements as hawkish or dovish, in the study of the tone of financial news, or in partisan speech (Lucca and Trebbi, 2009; Gentzkow et al., 2019).[36] Using these tools, our goal is to rule out the possibility that the comments of non-profits receiving grants may use similar words, but express views that are in opposition to their corporate donors.

Table 7 maintains the same design and structure of fixed effects as Table 6, but replaces the similarity score S_{fgr} with a semantic orientation concurrence score W_{fgr} as our dependent variable. W_{fgr} is defined as the negative absolute difference between the individual sentiment scores computed for the comments submitted by firm f and grantee g on rule r. To construct the sentiment scores of each comment, we follow Loughran and McDonald (2011) and employ their recommended Fin-Neg word list and TF-IDF weighting scheme: first we compute a weight for each word in the comment that reflects its frequency in the document, and relative rarity in other documents. Then the comment sentiment is computed as the sum of weights for words in the Fin-Neg dictionary, divided by the sum of weights for all words in the comment. The Fin-Neg dictionary is based on a negative sentiment word list from the Harvard Psychosociological Dictionary, but corrects the scoring of words that often occur in business settings (for example, the Harvard list codes "*foreign*" as negative) and Loughran and McDonald (2011) demonstrate that the resulting sentiment scores predict firm financial outcomes when applied to the text of

[36] In general, by semantic orientation we refer to the direction (polarity) of words, phrases or longer pieces of text in a semantic space or context (e.g., friendly/adversarial, dovish/hawkish, positive/negative) calculated based on a reference lexicon of words or n-grams over which directionality is carefully labeled by a pool of researchers.

SEC filings. The comments in our data discuss a wider range of topics than corporate finance, but we believe the Fin-Neg word list is more suitable than comparable sentiment dictionaries which do not correct for common business language. We interpret each comment sentiment score as a measure of how negative the comment is towards the rule. The interpretation of the coefficient of interest β_1 is therefore the effect of a charitable donation on the alignment of sentiment across firm and non-profit (i.e., the excess co-movement of sentiment in the two comments relative to any randomly generated pair of firm and grantee comments on that rule).

The data do not support the view that donations systematically reach grantees expressing opposing views to the firm providing the grant relative to a random grantee. The sign of β_1 is inconsistent across specifications and never statistically or economically significant. In Appendix F we show that the results in Table 7 also hold if we use different dictionaries and approaches for measuring sentiment, including measuring partisan alignment following Gentzkow et al. (2016). Overall, we conclude that there is no systematic relationship between comment sentiment and donations, and that our findings are unlikely to be explained by firm and grantee comments carrying similarly worded, but antagonistic messages.

5 Comments and final rules

The evidence provided thus far points to firms and their recent grantees commenting more often on the same rules and with more similar language. Circling back to our initial motivation, these patterns may be of concern only if they have an impact on final regulations.

At this point it is important to distinguish between two very different pieces of text that appear in the Federal Register when the final rule is published: i) the *final regulatory text* is designed to formulate, amend, or repeal sections of the Code of Federal Regulations (5 U.S.C. § 551(5)) and is written with a terminology and structure, at times dictating a change in a single word, that makes it very different from comments submitted and hence unsuitable to our analysis; ii) the *discussion of the rule* tends to be longer and presents arguments in favor of, or against, specific choices that may have been brought forward by firms, non-profits, and other entities in their attempts to persuade the regulator. We therefore focus on this latter part of the final rule.[37]

Typically, it is extremely hard to assess the effects of lobbying on policy outcomes (Kang, 2016). Much lobbying activity is designed to block change (so no policy differences are observed in equilibrium) and information flows are immaterial and undisclosed (e.g., meetings and phone calls). In our context, though, it is possible to measure the weight placed on each firm's comments

[37]The discussion of the rule is found in the Supplementary Information section, which is part of the preamble to the final rule and typically constitutes its most important component. See https://www.federalregister.gov/uploads/2011/01/the_rulemaking_process.pdf. Last accessed 5/4/2021.

by employing two proxies: the similarity between the final rule discussion by the regulatory agency and the firm's own comments, and the frequency with which a firm is cited by name in the agency's discussion of the final rule. We aim to assess whether, when a firm's grantee comments on the same rule as the firm, the regulator's published discussion of the final rule appears more similar to the firm's comments, and whether the regulator cites that firm more frequently in its discussion.

As an example, consider the concern expressed by Wells Fargo, one the largest depository institutions in the U.S., on a specific regulatory burden that the bank believed was implied by the proposed version of the so-called Volcker Rule of the Dodd-Frank Act of 2010. The Volcker Rule aimed to prohibit depository institutions from engaging in the use of part of their depository funding for speculative trading (proprietary trading).[38] Wells Fargo expressed the concern that the proposal required transaction-by-transaction oversight: "*We also do not believe that the Proposed Rule's transaction-by-transaction approach, which would require analyzing permitted customer trading, market making, underwriting and hedging activities on a transaction-by-transaction basis, is the best way for the Agencies to implement the Proposed Rule...*"[39] The OCC addressed this concern directly and conceded some changes to the rule: "*A number of commenters expressed general concern that the proposed underwriting exemption's references to a 'purchase or sale of a covered financial position' could be interpreted to require compliance with the proposed rule on a transaction-by-transaction basis. These commenters indicated that such an approach would be overly burdensome. . . . [T]o address commenters' confusion about whether the underwriting exemption applies on a transaction-by-transaction basis, the phrase 'purchase or sale' has been modified to instead refer to the trading desk's 'underwriting position.'*" The two texts appear related.[40]

We begin by constructing S_{fr}, the similarity score between the discussion of rule r and firm f's comment, using the same LSA-based approach as for our co-comment similarity analysis.[41] In contrast to the similarity score constructed in section 4, S_{fr} measures the similarity between a comment and the discussion of comments in the final rule, rather than the similarity between the texts of two comments on a proposed rule. We interpret S_{fr} as a proxy for the salience and effectiveness of the firm's comment in shaping the regulator's decisions.

Let us posit that S_{fr} is a function of the commenting efforts of the firm and of grantees connected to the firm by donations:

$$S_{fr} = \beta_1 GranteeCocomment_{fr} + \delta_f + \delta_r + \varepsilon_{fr} \tag{4}$$

[38]Rule 79 FR 5535

[39]Document ID OCC-2011-0014-0285

[40]Interestingly, the Black Economic Council, a recent Wells Fargo grantee, also expressed concerns on the same rule on grounds of excessive complexity. See Document ID OCC-2011-0014-0024.

[41]Because of the specific focus on the exact wording of the discussion of rule r, in this section we take r to refer to each separate final rule discussion, including the minority of cases where there are multiple final rules in a docket. Appendix A provides more details on the correspondence between rules and dockets.

The variable of interest is the dummy $GranteeCocomment_{fr} = I\left(\sum_g \sum_t C_{grt} \times D_{fg,t-1} > 0\right)$, which is equal to 1 if we observe that a grantee, commenting on the same rule as the firm, also received a donation from the firm in the same or previous year as the grantee submitted their comment, and 0 otherwise. If there is excess similarity between rule discussion and a firm's comment when grantees connected to the firm by donation also comment on that rule, β_1 should be positive. We interpret an increase in S_{fr} as a proxy that, at a minimum, captures the firm having the attention of the regulator. We note, however, that S_{fr} could conceivably correlate with influence in shaping the content of the final rule or in keeping out certain provisions. Importantly, given that we control for rule fixed effect in equation (4), our empirical test asks whether the comment-rule similarity is larger for firms that have a recent grantee commenting on the same rule *relative* to the comment-rule similarity for firms that also commented on that rule but did not have a recent grantee commenting as well.

We also examine whether firms are cited more often in final rule discussions in which we observe a comment by one of their grantees, employing $log(1 + citations)$.[42] Firm fixed effects in this specification capture the extent to which certain firms are systematically more likely to be cited by regulators across all rules. Similarly, rule fixed effects control for the fact that some rule discussions may include on average more numerous references to firms' comments. Note that we limit the citation analysis to the subset of agencies where there is a norm of citing specific commenters – in many agencies such citation behavior is very rare. We focus on agencies that cite at least one commenter per rule.[43]

Table 8 presents our regression results. We find that the similarity between firm comments and the rule discussion is 16% of a standard deviation higher when at least one grantee commenting on the same rule has received a recent donation from the firm. Similarly, firms are cited more frequently (33% more often) within each rule, and are more than twice as likely to be cited at all.

One of the main difficulties with interpreting these results as causal is that we do not observe all channels of communication from the firm to the regulator (a form of omitted variable bias). However, we do have information about lobbying contacts between the firm and regulator from lobbying disclosure reports filed with the Senate's Office of Public Records.[44] For columns (2), (4), (6), and (8), we control for the estimated expenditure on lobbyists hired to communicate with

[42] To the extent that the comments by grantees could be cited in place of a firm, we will underestimate the true extent to which a firm's view is cited in the final rule.

[43] One reason for this behavior is that, generically discussing comments instead of naming specific commenters may limit ex post legal action against the regulator. An instance is action brought for arbitrary and capricious behavior arising for agency's failure to address dissenting comments to a proposed rule. Note that if we do *not* limit the sample to agencies with a citation norm, the point estimates on these results are much smaller and some are not statistically significant. See Bertrand et al. (2018) for these results.

[44] We use bulk lobbying data that has been cleaned and organized by the Center For Responsive Politics, available through www.opensecrets.org.

the agency that published the rule in question.[45] Our results are robust to controlling for lobbying expenditures over the same time period as donations. In fact, conditional on our fixed effects specification, the inclusion of lobbying expenditures as a control does not change the estimated effect of grantee co-commenting at all. This adds weight to the interpretation that the channel of influence we capture in our analysis is through the submitted comments.[46]

As with our co-commenting and comment similarity results, these rule outcomes do not appear to be driven by future donations. In Appendix Table C.6 we add an indicator for future donations to grantee co-commenters. When both variables are included, it is the variable based on recent donations that predicts final rule similarity.

6 Heterogeneity analysis

In a final set of empirical analyses, we examine whether there exists heterogeneity in the relationship between donations and co-commenting behavior. Details on these analyses may be found in Appendix G; we summarize here our main findings for brevity. As our data span different dimensions, we explored heterogeneity by regulatory agency, by importance of rules, by grantee characteristics, and by industry/firm characteristics. Overall, Appendix G shows that our results are not driven by selected subsamples, but also that estimated coefficients respond in intuitive directions in terms of magnitudes.

In part G.1 of the Appendix, we show that for high-stake rules (i.e. rules that attract attention, with higher than median number of grantee comments), the extent of co-commenting we describe above is much stronger, both statistically and quantitatively. This is intuitive, as the use of charitable grants as influence is inherently costly, and firms will be more motivated to deploy these grants in situations in which the outcome is particularly contentious or important.

Part G.2 shows that grantees with agency-specific expertise are more frequent targets of donations at times when the firm comments on regulation. We also discuss how certain dimensions of heterogeneity, for example based on the interaction of charitable donations with the degree of expertise of a grantee and its engagement with specific regulatory agencies, may help rule out alternative mechanisms, including a potential for "hush money" to silence experts.[47]

One potentially important firm characteristic that may affect the extent and efficacy of the

[45] Lobbying disclosure reports do not contain per-agency expenditures, but each filing lists the branches of government contacted, and the total amount spent. We divide total expenditures for each filing evenly between all branches listed. In practice, our results are not sensitive to how this lobbying amount is constructed.

[46] We note, however, that our measure of influence via lobbying is only at the agency-level, so that our test of whether there exist correlated margins of influence-seeking is an imperfect one.

[47] In part G.5 of the Appendix we look at heterogeneity based on two other grantee attributes: research- and policy-orientation. Both types of non-profits are more likely than average to comment on regulation, but also less "persuadable" to comment via donation. These results are marginally significant at best, however.

behavior we document is concentration of the commenting firm's industry. As shown in part G.3, our estimated coefficients tend to be quantitatively and statistically stronger in more concentrated industries, based on top 4 and top 8 revenue concentration ratios. As concentrated industries offer the most natural environment for a collective action solution and for lobbying according to the standard logic of Olson (1965), this result appears to align with the intuition that there is a strategic element to the co-commenting phenomenon we document.

Regarding regulatory agencies, in part G.4 we focus on whether one can detect any asymmetry across party lines in the behavior of agencies under different administrations. As our sample covers both Republican and Democratic Presidents, we focus on the partisan affiliation of the President who appoints executive branch and independent agency commissioners during each electoral cycle. We show that regulatory agencies with commissioners appointed under a Republican administration appear less sensitive to the co-commenting behavior of grantees and that firms make less use of co-commenting under Republican administrations. One explanation for this result may be that Republican appointees may be less sensitive to special interests beyond the business sector relative to Democrats, so there is lesser value to co-opting non-profits.[48]

7 Case studies

In this section we provide case study evidence to complement our econometric analysis, and to inform the discussion about the welfare implications of our findings. The case studies that we discuss entered in the public domain either through court filings or based on documents uncovered by public interest organizations and journalists. These examples allow us to observe directly the types of activities that one is otherwise required to infer based on statistical analysis. The cases are also sufficiently widespread across industries and over time to underscore how the behavior we document may be more diffuse than previously considered, and they are sufficiently compelling in nature, that one may not wish to dismiss a priori the "comments-for-sale" view.

7.1 Soft drink companies and public health policy

The Coca-Cola Foundation/AAPD example presented in the introduction illustrates a case of a sizable donation followed by a shift in recommendations by a nominally arms-length grantee. It is important to underscore that such events are not necessarily anomalies.

Aaron and Siegel (2017), in their analysis of sponsorships by the two major soda companies between 2011 and 2015, report how "*Save the Children, a group that promoted soda taxes, suddenly dropped this effort in 2010 after receiving more than \$5 million from the Coca-Cola Company and*

[48]See Yackee and Yackee (2006) for a discussion.

PepsiCo in 2009." Save the Children had previously campaigned for soda taxes in the District of Columbia, Mississippi, New Mexico, Philadelphia, and Washington State, and in 2010 abruptly changed course after receiving the grant.

Emails recovered by The Associated Press show more direct evidence of Coca-Cola donations helping to influence policy positions at another non-profit, the Global Energy Balance Network (GEBN), an anti-obesity group run by a professor at the University of Colorado. The emails reveal that, concurrent with a $1.5 million gift from the company to GEBN, Coca-Cola's chief health and science officer suggested content for the non-profit's website, provided input into the selection of GEBN's senior leadership, and edited GEBN's mission statement, which was primarily focused on shifting the blame for obesity towards lack of physical exercise.[49]

7.2 Non-profit support for power utilities' regulatory agenda

In 2019, the Energy and Policy Institute (EPI) released a report titled, "*How Utilities use Charitable Giving to Influence Politics and Increase Investor Profits.*" EPI surveyed the philanthropic activities of 10 utilities – whose total giving between 2013-2017 exceeded $1 billion – using their IRS Form 990s and FERC Form 1 and Form 60 (EPI, 2019). As participants in heavily regulated industries, utility companies are prime candidates for the types of regulatory influence-seeking behaviors that are our focus.

Electric utilities routinely buttress their requests for rate increases or public subsidies with letters of support from local non-profits, often representing minorities or disadvantaged groups. The EPI report revealed grants by Ameren in Illinois to the NAACP, The Black Chamber of Commerce, and the Springfield Urban League, all given around deliberations for weakening energy efficiency rules in the state. Similarly, the Arizona Public Service (APS) Company, an electricity utility, enlisted Chicanos Por La Causa and the Phoenix Indian Center in its letter supporting rate increases – both APS grantees. More starkly, in 2016, the leader of the Greater Abyssinia Baptist Church in Cleveland, Ohio, was the lead signatory of a letter sent to the state's governor from the Cleveland Clergy Council in support of an Electric Security Plan proposed by FirstEnergy of Akron, Ohio. In 2016 the Church had received a $100,000 donation from FirstEnergy's foundation, and another in 2017. However, just before these donations, the church leader had expressed concerns and members of his congregation had marched in protest against the plan.

There is also distinct evidence that some of the messaging from grantees may be manipulated by firms. For instance, in May 2019, EPI analyzed several public written testimonies by grantees speaking favorably about the bailout of FirstEnergy Solutions, a bankrupt utility in Ohio. The

[49]Nancy Fink Huehnergarth, "Emails Reveal How Coca-Cola Shaped The Anti-Obesity Global Energy Balance Network", *Forbes*, November 24, 2015. After news of Coke's involvement became public, GEBN was shut down.

examination of the pdf files' metadata revealed that the documents were all created by a lobbyist hired by FirstEnergy Solutions.

The case of the NAACP in particular warrants further elaboration as an unlikely ally in companies' pushback against unfavorable regulation or legislation. A 2020 *New York Times* column focused on corporate donations to the NAACP describes, for example, something approaching an explicit *quid pro quo* involving the NAACP's Florida conference, which had received $225,000 from Florida Power and Light.[50] As the *Times* reports, "*donations doubled in 2014 just as the utility was pressing state regulators to restrict rooftop solar power and weaken the state's energy efficiency goals*," while in the same year, according to the *Times* report, NAACP Florida filed comments in support of the company's position with the state Public Service Commission, taken verbatim from Florida Power and Light lobbying materials. The NAACP's comments were later cited by the commission in the ruling in favor of utilities' demands (the commission cut its energy-efficiency goals by 90%). The organization's director later observed that it was clear that, *"if we wanted the money, we had to [support the utilities' position]*."

The NAACP's national office saw these types of concerns as sufficiently pervasive and problematic that in 2019 it published a white paper for their local chapters warning of the various ways that energy companies would try to co-opt non-profits in pursuing fossil-fuel-friendly policies.[51] Funding is given as a key mechanism, with the document providing the example of St. Louis Missouri branch, which was cut off by Peabody Coal, a frequent donor, after voicing opposition to fossil fuel interests in comments to the EPA.[52]

7.3 Non-profit support for telecommunications mergers

Peng (2016) describes the efforts of telecommunications firms to win merger approvals from the Federal Communication Commission (FCC), in part by assembling diverse and vocal coalitions of supporters. Peng quotes Crawford (2013) on the Comcast-NBCU merger, in which *"[t]he company encouraged letters to the FCC from more than one thousand non-profits...including community centers, rehabilitation centers, civil rights groups, community colleges, sports programs, [and] senior citizen groups*." For the AT&T/T-Mobile merger, Peng similarly documents letters of support addressed to the FCC from non-profits that, at first glance, would appear to have little interest or expertise in telecommunications policy, including a homeless shelter in Louisiana, a special needs

[50] Ivan Penn, "N.A.A.C.P. tells local chapters: Don't let energy industry manipulate you," *New York Times*, January 5, 2020.

[51] Jacqueline Patterson. 4/1/2019. "Fossil Fuel Foolery: An Illustrated Primer on the Top 10 Manipulation Tactics of the Fossil Fuel Industry". NAACP.

[52] The report states (p.11) that Peabody Coal's reply to the NAACP St. Louis Branch president upon inquiry about a missing grant was: "*We only give money to our friends and you folks went down and talked bad about coal to the EPA*."

employment agency in Michigan, and the Gay & Lesbian Alliance Against Defamation (GLAAD). The non-profits were all AT&T Foundation grantees (in the case of the homeless shelter, the donation had come in just five months before the merger was announced). In no case did the non-profits disclose their AT&T funding in their comments to the FCC, and in at least one instance, the comments did not appear to represent the views of the non-profit membership. According to Peng, "*GLAAD's president and six board members resigned when its merger endorsement made headlines and revealed that the organization had received AT&T funds.*"

7.4 Tobacco Industry

The tobacco industry was a pioneer in the sort of indirect influence we document in this section. Via previously confidential British American Tobacco's (BAT) documents, released publicly during the tobacco health damages litigation of the early 2000s, Fooks and Gilmore (2013) find evidence that, "*[d]onations [are] used to facilitate closer relationships with recipient organisations by generating trust and support and shape their organisational priorities. Organisations are encouraged to lobby and advocate on behalf of the industry, thereby expanding political conflicts around tobacco control.*" They document[53] that BAT's donations were "*allocated to some groups on the basis of their potential to shape policy agenda though their influence on government thinking and news reporting*" and that they were "*...made to shift thinking on the importance of tobacco control regulation by influencing perceptions of the relative risks of tobacco to population level health.*"[54]

Similar conclusions are reached in Tesler and Malone (2008), McDaniel and Malone (2009), and McDaniel and Malone (2012), using documents from other tobacco corporations. For instance, McDaniel and Malone (2009) report how Philip Morris's funding to the Young Women's Christian Association national organization disappeared after the organization signed a public letter that was critical of tobacco marketing practices.

The strategic use by tobacco companies of charitable giving as an influence tool over third party grantees is now so heavily documented (and deemed ultimately detrimental to public welfare) that the World Health Organization's (WHO) Articles 5.3 and 13 of the Framework Convention on Tobacco Control (FCTC)[55] specifically aim to limit the political effects of tobacco industry philanthropy.[56]

[53]See Legacy Tobacco Documents Library at http://legacy.library.ucsf.edu.

[54]For example the authors report that in China "*BAT supported the Beijing Liver Foundation...to lobby the Ministry of Public Health to "maintain a perspective on health issues," recognising that the company could not "credibly, directly communicate with the Ministry""* with the goal of shifting public health concerns from smoking to other non-tobacco related issues, such as hepatitis.

[55]World Health Organization (2013). "Guidelines for Implementation of Article 5.3 of the WHO Framework Convention on Tobacco Control." Retrieved from https://escholarship.org/uc/item/1vw586jd. p.7.

[56]In its guidelines to the implementation of Article 5.3 of the FCTC, the WHO provides the recommendation to, "*[d]enormalize and, to the extent possible, regulate activities described as socially responsible by*

7.5 Mobil Foundation

A document leaked from the Mobil Foundation provides detailed written justification for each of the grants that it made in 1994.[57] For the vast majority of these grants, the document includes a paragraph with the heading "*Benefits to Mobil*" that delineates the reasons that supporting a given charity may be advantageous to the Mobil Corporation. These reasons often go beyond the oft-cited rationales for corporate philanthropy of brand recognition and goodwill.

Of particular interest to our setting are instances in the document in which attempts at indirect influence over regulation appear as an explicit rationale. Excerpts from entries in the 1994 Budget Recommendations of Mobil Foundation, Inc. most pertinent to our discussion on regulation are reported in Table 9.

Some of these read as rather anodyne explanations for donations to promote the use of science in environmental risk assessment. For example, a donation to the Academy of Natural Sciences (unaffiliated with the National Academy of Sciences nor with the American Academy of Arts and Sciences) is justified based on the organization's ability "*to challenge the EPA behind-the-scenes on the effectiveness of a regulation for the environment and whether sound science supports the proposed law.*" A similar rationale is provided for a grant to the Harvard Center for Risk Analysis among various others, to support the promotion of "scientific risk assessment" which, "*will benefit Mobil through the adoption of more cost-effective laws and regulations.*" In other instances, the potential to influence a grantee appears more directly, as in the case of a grant to the National Research Council in support of a study on groundwater treatment, where the benefits to Mobil include the possibility that, "*[b]y helping to fund the study, Mobil may be offered the opportunity to participate or to receive early access to the findings*," or in the case of the National Safety Council (which has a Mobil employee on its board) where Mobil was "*successful in 1989 in having the National Safety Council Board of Directors pass a resolution opposing the mandating of any alternative fuel, such as methanol, until studies demonstrated a reduced risk of death, illness or injury.*"

The entries in Table 9 alone account for about 10% of total charitable activity of the Mobil Foundation that year (about $1.2 million in 1993 dollars). The document also provides the names of other significant corporate donors to each organization (in addition to a time series of donations by Mobil to that specific grantee). In most cases, these other donors are other oil, chemical, or industrial firms, indicating that Mobil is unlikely to be the only business aiming to forge ties with potentially useful non-profits.[58]

---

the tobacco industry, including but not limited to activities described as corporate social responsibility." See https://www.who.int/fctc/guidelines/article_5_3.pdf. Last accessed May 4, 2021.

[57] Sharon Kelly. 6/12/2019. The Guardian. "How Mobil pushed its oil agenda through 'charitable giving'".

[58] The type of hidden influence-seeking we describe – in addition to being widespread across firms – may also

8 Conclusions

Politicians and voters are frequent targets of messages aimed at persuading them of the merits of specific policy positions. While in most cases the identity of the sender is disclosed, allowing an assessment of the bias and interests of a message's originator, in other instances the identity may be unavailable or even deliberately obscured. These situations range from the use of dark money in U.S. electoral politics in the aftermath of the Supreme Court's *Citizens United v. Federal Election Commissio*n and *McCutcheon v. Federal Election Commission* cases, to the circulation of white papers by think-tanks and other non-profits.

Independent arms-length organizations may extend the credibility of the positions held by special interests. Our paper argues that one has to be careful in assessing the information provided by these apparently independent organizations when this information comes in close proximity to monetary transfers from firms. Such transfers, often in the form of charitable grants, are virtually undetectable by private citizens and civil servants without access to detailed tax returns information.

In order to provide a quantitative and systematic perspective to this issue, this paper studies the interaction of non-profit organizations and large corporations within the United States federal regulatory environment. The paper presents evidence that corporate foundations' charitable grants reach targeted non-profits just before those same non-profits engage in public commentary. The availability of a large set of public comments by non-profits and by corporations on a diverse set of rules and regulations, ranging from banking to environmental regulation, makes for a rich and virtually untapped empirical environment.

The content of the comments simultaneously communicated by non-profits and by corporations appears systematically closer (in terms of textual similarity) in the presence of a charitable contribution provided immediately before those comments are filed. While circumstantial, the evidence points to potential concerns in the assessment of this *prima facie* independent information by targeted regulators, who may be unaware of the philanthropic grants that take place out of direct view. The regulator may thus interpret similar comments from diverse sources as independent, when in fact they are linked via financial ties.

The paper also tries to evaluate whether there exist benefits to large business interests from enlisting allied advocates who may be perceived as more balanced and less biased. We focus

not be limited to the U.S. context. For example, a Greenpeace Canada report released in 2020 provides details of a a confidential consultant's presentation which lays out a strategy for influencing Canada's clean fuel standards. The presentation emphasizes the need for the appearance of "*diverse voices...including credible experts and third parties*," while industry's role seem to remain as "secondary." One key prong of this approach is facilitating relations with think tanks and NGOs, and the report lists a number of organizations that would likely be supportive. See "Leaked document details industry's secret plan to defeat Clean Fuel Standard: 'Fighting climate change is a losing battle'" Jesse Firempong, October 7, 2020.

on textual similarity between the commenting firm and final rule discussion to gauge influence of comments over regulation. We find evidence consistent with co-comments from non-profits providing additional visibility to the messages sent by the firms themselves, measured in terms of comment similarity to the final rule or likelihood of citation of a donor firm. The ultimate economic returns to regulatory influence remains complex to assess, and we see this as an area of empirical investigation in need of further research.

References

Aaron, Daniel G and Michael B Siegel, "Sponsorship of national health organizations by two major soda companies," *American journal of preventive medicine*, 2017, *52* (1), 20–30.

Austen-Smith, David, "Information and influence: Lobbying for agendas and votes," *American Journal of Political Science*, 1993, pp. 799–833.

_ , "Campaign contributions and access," *American Political Science Review*, 1995, *89* (3), 566–581.

_ **and John R Wright**, "Counteractive lobbying," *American Journal of Political Science*, 1994, pp. 25–44.

Baron, David P, "Private politics, corporate social responsibility, and integrated strategy," *Journal of Economics & Management Strategy*, 2001, *10* (1), 7–45.

Baumgartner, Frank R, Jeffrey M Berry, Marie Hojnacki, Beth L Leech, and David C Kimball, *Lobbying and policy change: Who wins, who loses, and why*, University of Chicago Press, 2009.

Bebchuk, Lucian and Robert J. Jackson, "Shining Light on Corporate Political Spending," *Georgetown Law Journal*, 04 2013, *101*, 923–967.

Becker, Gary S, "A theory of competition among pressure groups for political influence," *The quarterly journal of economics*, 1983, *98* (3), 371–400.

Bertrand, Marianne, Matilde Bombardini, and Francesco Trebbi, "Is it whom you know or what you know? An empirical assessment of the lobbying process," *The American Economic Review*, 2014, *104* (12), 3885–3920.

_ , _ , Raymond Fisman, and Francesco Trebbi, "Tax-exempt lobbying: Corporate philanthropy as a tool for political influence," *American Economic Review*, 2020, *110* (7), 2065–2102.

_ , _ , _ , Bradley Hackinen, and Francesco Trebbi, "Hall of mirrors: Corporate philanthropy and strategic advocacy," Technical Report, National Bureau of Economic Research 2018.

Besley, Timothy and Stephen Coate, "Elected versus appointed regulators: Theory and evidence," *Journal of the European Economic Association*, 2003, *1* (5), 1176–1206.

Blei, David M, Andrew Y Ng, and Michael I Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, 2003, *3*, 993–1022.

Bombardini, Matilde and Francesco Trebbi, "Votes or money? Theory and evidence from the US Congress," *Journal of Public Economics*, 2011, *95* (7), 587–611.

_ and _ , "Competition and political organization: Together or alone in lobbying for trade policy?," *Journal of International Economics*, 2012, *87* (1), 18–26.

Brown, William O, Eric Helland, and Janet Kiholm Smith, "Corporate philanthropic practices," *Journal of corporate finance*, 2006, *12* (5), 855–877.

Crawford, Susan P, *Captive audience: The telecom industry and monopoly power in the new gilded age*, Yale University Press, 2013.

Deerwester, S, S Dumais, G Furnas, T Landauer, and R Harshman, "Indexing by latent semantic analysis," *Journal of the American Society for Information Science*, 1990, *41* (6), 391–407.

DeGregorio, Christine A, *Networks of champions: leadership, access, and advocacy in the US House of Representatives*, University of Michigan Press, 2010.

Dewatripont, Mathias and Jean Tirole, "Advocates," *Journal of political economy*, 1999, *107* (1), 1–39.

Drutman, Lee, *The business of America is lobbying: How corporations became politicized and politics became more corporate*, Oxford University Press, 2015.

Dumais, S, G Furnas, T Landauer, and S Deerwester, "Using latent semantic analysis to improve information retrieval," *Proceedings of CHI 88 Conference on Human Factors in Computing Systems*, 1988, pp. 281–285.

Dumais, Susan T, "Latent semantic analysis," *Annual review of information science and technology*, 2004, *38* (1), 188–230.

Easley, David and Maureen O'Hara, "The economic role of the nonprofit firm," *The Bell Journal of Economics*, 1983, pp. 531–538.

EPI, "Strings Attached. How utilities use charitable giving to influence politics and increase investor profits," 2019.

Fooks, Gary J and Anna B Gilmore, "Corporate philanthropy, political influence, and health policy," *PloS one*, 2013, *8* (11), e80864.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, "Text as data," *Journal of Economic Literature*, 2019, *57* (3), 535–74.

_ , Jesse M Shapiro, and Matt Taddy, "Measuring polarization in high-dimensional data: Method and application to congressional speech," Technical Report, National Bureau of Economic Research 2016.

Gillam, Carey, *Whitewash: The story of a weed killer, cancer, and the corruption of science*, Island Press, 2017.

Glaeser, Edward L, "The governance of not-for-profit firms," Technical Report, National Bureau of Economic Research 2002.

_ and Andrei Shleifer, "Not-for-profit entrepreneurs," *Journal of public economics*, 2001, *81* (1), 99–115.

_ and _ , "The rise of the regulatory state," *Journal of economic literature*, 2003, *41* (2), 401–425.

Gordon, Sanford C and Howard Rosenthal, "Cross-ideological coordination by private interests: Evidence from mortgage market regulation under Dodd-Frank," *Business and Politics*, 2018, pp. 1–29.

Grossman, Gene M and Elhanan Helpman, "Protection for sale," *The American Economic Review*, 1994, *84* (4), 833–850.

_ and _ , *Special interest politics*, MIT press, 2001.

Hanson, Samuel G and Adi Sunderam, "The variance of non-parametric treatment effect estimators in the presence of clustering," *Review of Economics and Statistics*, 2012, *94* (4), 1197–1201.

Jacobson, Michael F, "Lifting the veil of secrecy from industry funding of nonprofit health organizations," *International journal of occupational and environmental health*, 2005, *11* (4), 349–355.

Johnson, Simon and James Kwak, *13 bankers: The Wall Street takeover and the next financial meltdown*, Vintage, 2010.

Kang, Karam, "Policy influence and private returns from lobbying in the energy sector," *Review of Economic Studies*, 2016, *83* (1), 269–305.

_ and Hye Young You, "The Value of Connections in Lobbying," *working paper*, 2016.

Kendall, Chad, Tommaso Nannicini, and Francesco Trebbi, "How do voters respond to information? Evidence from a randomized campaign," *American Economic Review*, 2015, *105* (1), 322–53.

Kitzmueller, Markus and Jay Shimshack, "Economic perspectives on corporate social responsibility," *Journal of Economic Literature*, 2012, *50* (1), 51–84.

Krishna, Vijay and John Morgan, "A model of expertise," *The Quarterly Journal of Economics*, 2001, *116* (2), 747–775.

Kwak, James, "Cultural capture and the financial crisis," *Preventing regulatory capture: Special interest influence and how to limit it*, 2014, *71*, 79–81.

Laffont, Jean-Jacques and Jean Tirole, *A theory of incentives in procurement and regulation*, MIT press, 1993.

Lando, Harry A, "Toward a comprehensive strategy for reducing the health burden of tobacco," *British journal of addiction*, 1991, *86* (5), 649–652.

Le, Quoc and Tomas Mikolov, "Distributed Representations of Sentences and Documents," in "Proceedings of the 31st International Conference on International Conference on Machine Learning Volume 32" ICML'14 JMLR.org 2014, pp. II–1188–II–1196.

Lohmann, Susanne, "Information, access, and contributions: A signaling model of lobbying," *Public Choice*, 1995, *85* (3-4), 267–284.

Lorenz, Geoffrey Miles, "Prioritized interests: Diverse lobbying coalitions and congressional committee agenda setting," *The Journal of Politics*, 2020, *82* (1), 225–240.

Loughran, Tim and Bill McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *The Journal of finance*, 2011, *66* (1), 35–65.

Lucca, David O and Francesco Trebbi, "Measuring central bank communication: an automated approach with application to FOMC statements," Technical Report, National Bureau of Economic Research 2009.

Mahoney, Christine and Frank R Baumgartner, "Partners in advocacy: Lobbyists and government officials in Washington," *The Journal of Politics*, 2015, *77* (1), 202–215.

McDaniel, Patricia A and Ruth E Malone, "Creating the ï¿œdesired mindsetï¿œ: Philip Morris's efforts to improve its corporate image among women," *Women & health*, 2009, *49* (5), 441–474.

_ **and** _ , "British American Tobacco's partnership with Earthwatch Europe and its implications for public health," *Global Public Health*, 2012, *7* (1), 14–28.

Nelson, David and Susan Webb Yackee, "Lobbying coalitions and government policy change: An analysis of federal agency rulemaking," *The Journal of Politics*, 2012, *74* (2), 339–353.

Olson, Mancur, *The Logic of Collective Action*, Harvard University Press, 1965.

Peltzman, Sam, "Toward a more general theory of regulation," *The Journal of Law and Economics*, 1976, *19* (2), 211–240.

Peng, Victoria, "Astroturf Campaigns: Transparency in Telecom Merger Review," *University of Michigan Journal of Law Reform*, 2016, *49* (2).

Phinney, Robin, *Strange bedfellows: Interest group coalitions, diverse partners, and influence in American social policy*, Cambridge University Press, 2017.

Potters, Jan and Frans Van Winden, "Lobbying and asymmetric information," *Public choice*, 1992, *74* (3), 269–292.

Řehůřek, Radim and Petr Sojka, "Software Framework for Topic Modelling with Large Corpora," in "Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks" ELRA Valletta, Malta May 2010, pp. 45–50. http://is.muni.cz/publication/884893/en.

Rose-Ackerman, Susan, "Altruism, nonprofits, and economic theory," *Journal of economic literature*, 1996, *34* (2), 701–728.

Ross, Lainie Friedman, JW Norton, SA Young et al., "Is academic medicine for sale," *N Engl J Med*, 2000, *343* (7), 508–10.

Stigler, George J, "The theory of economic regulation," *The Bell journal of economics and management science*, 1971, pp. 3–21.

Strauss, Peter L, "From expertise to politics: the transformation of American rulemaking," *Wake Forest L. Rev.*, 1996, *31*, 745.

Sunstein, Cass R, "The Office of Information and Regulatory Affairs: myths and realities," *Harv. L. Rev.*, 2012, *126*, 1838.

Tesler, Laura E and Ruth E Malone, "Corporate philanthropy, lobbying, and public health policy," *American journal of public health*, 2008, *98* (12), 2123–2133.

Vidal, Jordi Blanes I, Mirko Draca, and Christian Fons-Rosen, "Revolving door lobbyists," *The American Economic Review*, 2012, *102* (7), 3731–3748.

West, William F and Connor Raso, "Who shapes the rulemaking agenda? Implications for bureaucratic responsiveness and bureaucratic control," *Journal of Public Administration Research and Theory*, 2013, *23* (3), 495–519.

Yackee, Jason Webb and Susan Webb Yackee, "A bias towards business? Assessing interest group influence on the US bureaucracy," *The Journal of Politics*, 2006, *68* (1), 128–139.

Yackee, Susan Webb, "The politics of rulemaking in the United States," *Annual Review of Political Science*, 2019, *22*, 37–55.

Yurochkin, Mikhail, Sebastian Claici, Edward Chien, Farzaneh Mirzazadeh, and Justin Solomon, "Hierarchical Optimal Transport for Document Representation," 2019.

Table 1: Annual firm comment count distribution by commenting relationship

	Annual firm comment counts (rules per firm/year)[1]							
	Mean	Std. Dev.	Min	Max	P50	P90	P99	Total[2]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Annual comments from each firm on:								
Any rule	1.9	4.9	0.1	108.4	0.6	4.4	20.1	1457.8
Rules where at least one grantee also comments	1.3	2.2	0	18.9	0.6	3.4	12.3	1051.0
Rules where at least one grantee who receives a donation from the firm at any time also comments	0.3	1.0	0	12.3	0	0.7	4.9	229.9
Rules where at least one grantee who has received a recent[3] donation from the firm also comments	0.2	0.7	0	10.9	0	0.3	3.3	136.3

Notes: This table summarizes the number of comments submitted by each firm in a representative year (computed as the average across years 2008-2014, the period during which our data are most complete).
[1] Each firm-rule-year observation is counted as one comment. Firms that submit multiple documents (or multiple form letters as part of a coordinated campaign) on the same rule in the same calendar year are counted as submitting one comment on that rule.
[2] Total comment count for all firms in our sample.
[3] We use the term "recent" to refer to any donation which occurs in the same or previous calendar year relative to the comment year.

Table 2: Annual grantee comment count distribution by commenting relationship

	Annual grantee comment counts (rules per grantee/year)[1]							
	Mean	Std. Dev.	Min	Max	P50	P90	P99	Total[2]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Annual comments from each grantee on:								
Any rule	0.6	1.9	0.1	71.6	0.1	1.0	6.1	5073.0
Rules where at least one firm also comments	0.3	1.1	0	32.6	0.1	0.6	4.0	3040.0
Rules where at least one firm who donates to the grantee at any time also comments	0.1	0.8	0	33.1	0	0.3	2.9	1255.6
Rules where at least one firm who has recently[3] donated to the grantee also comments	0.1	0.5	0	31.4	0	.1	1.4	645.6

Notes: This table summarizes the number of comments submitted by each grantee in a representative year (computed as the average across years 2008-2014, the period during which our data are most complete). The set of grantees include those that comment on at least one rule during 2003-2016.
[1] Each grantee-rule-year observation is counted as one comment. Grantees that submit multiple documents (or multiple form letters as part of a coordinated campaign) on the same rule in the same calendar year are counted as submitting one comment on that rule.
[2] Total comment count for all grantees in our sample.
[3] We use the term "recent" to refer to any donation which occurs in the same or previous calendar year relative to the comment year.

Table 3: Annual firm donation distribution by commenting relationship

	Annual donations (millions $/year)							
	Mean	Std. Dev.	Min	Max	P50	P90	P99	Total[1]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Annual donations from each firm to:								
All grantees	9.0	29.1	0	407	2.3	18.7	124.9	3430.0
Grantees that comment at least once	2.5	7.5	0	78.4	0.5	5.2	39.5	936.1
Grantees that ever submit a comment to the same agency as the firm	1.4	5.9	0	77.4	0.1	2.5	30.3	544.3
Grantees that ever comment on the same rule as the firm	0.7	4.3	0	75.4	0	.9	12.8	247.4

Notes: This table summarizes the distribution of annual firm donations for a representative year for our sample of firms that comment at least once (computed by averaging across years 2008-2014, the period during which our data are most complete).
[1] Total donations for all firms in our sample.

Table 4: Co-commenting - Recent donation

Dependent variable	Firm f and grantee g commented on the same rule in year $t(\times 100)$			
Mean			0.175	
	(1)	(2)	(3)	(4)
Firm f contributed to grantee g in year t or $t-1$	1.167***	0.727***	0.133***	0.080**
	(0.038)	(0.035)	(0.038)	(0.036)
Fixed effects				
Year	Y	Y	Y	
Grantee		Y		
Firm		Y		
Grantee-Firm Pair			Y	Y
Grantee-Year				Y
Firm-Year				Y
Observations	122,287,230	122,287,230	122,232,220	122,232,220
R^2	0.003	0.019	0.133	0.201

Notes: The dependent variable is equal to 100 if grantee g and firm f comment on the same rule r in year t. The independent variable is equal to one if grantee g received a donation from firm f at year t or $t-1$. The sample includes the set of firm-grantee pairs constructed as follows: foundations whose firms comment on at least one rule during 2003-2016, and these foundations' grantees who commented on at least one rule during the same period. Standard errors are clustered at the grantee-firm pair level. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Commenting on rules

Dependent variable	Grantee g commented on rule $r \times 100$			
Mean	0.043			
	(1)	(2)	(3)	(4)
Grantee g received donation	0.237***	0.177***	0.209***	0.142***
from any firm commenting on r	(0.022)	(0.018)	(0.022)	(0.016)
Fixed effects				
Grantee		Y		Y
Regulation			Y	Y
Observations	117,545,368	117,545,368	117,545,368	117,545,368

Notes: The dependent variable is equal to 100 if grantee g comments on rule r. The independent variable is equal to one if grantee g received in any year 2003-2016 a donation from a firm that commented on r. The sample includes the set of firm-grantee pairs constructed as follows: foundations whose firms comment on at least one rule during 2003-2016, and these foundations' grantees who commented on at least one rule during the same period. Standard errors are two-way clustered at the rule and grantee level. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Similarity of comments - Recent donation

Dependent variable	Similarity of comments by grantee g and firm f on same rule					
	(1)	(2)	(3)	(4)	(5)	(6)
Grantee g received donation from firm f at t or $t-1$	0.047*** (0.016)	0.061* (0.035)	0.032* (0.020)	0.057*** (0.017)	0.065* (0.039)	0.040* (0.022)
Fixed Effects						
Rule	Y	Y	Y	Y	Y	Y
Firm	Y			Y		
Grantee	Y			Y		
Firm-Grantee Pair		Y			Y	
Agency×NAICS×NTEEC			Y			Y
Comment style control				Y	Y	Y
Observations	168,347	71,195	81,851	168,347	71,195	81,851

Notes: The dependent variable is a similarity index between the comment of firm f and the comment of grantee g in the same rule r, scaled to have a standard deviation of one. The independent variable is equal to one if grantee g received a donation from firm f in the year when the comment appears or the year before. The sample includes the subset of firm-grantee observations in which firm and grantee comment on the same regulation. Standard errors use two-way clustering by rule and firm-grantee pair. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Comment sentiment alignment - Recent donation

Dependent variable	Sentiment alignment of comments by grantee g and firm f on same rule-year					
	(1)	(2)	(3)	(4)	(5)	(6)
Grantee g received donation	-0.009	-0.040	-0.014	0.006	-0.028	-0.028
from firm f at t or $t-1$	(0.020)	(0.039)	(0.020)	(0.031)	(0.039)	(0.023)
Fixed Effects						
Rule	Y	Y	Y	Y	Y	Y
Firm	Y			Y		
Grantee	Y			Y		
Firm-Grantee Pair		Y			Y	
Agency×NAICS×NTEEC			Y			Y
Commenter style control				Y	Y	Y
Observations	168,341	71,189	81,851	168,341	71,189	81,851

Notes: The dependent variable is the negative absolute difference between the sentiment score assigned to the comment of firm f and the comment of grantee g in the same rule-year rt, using TF-IDF weighted Fin-Neg wordlist created by Loughran & McDonald (2011), with the dependent variable re-scaled to have a standard deviation of one. The independent variable is equal to one if grantee g received a donation from firm f in the year when the comment appears or the year before. The sample includes the subset of firm-grantee observations in which firm and grantee comment on the same regulation. Standard errors use two-way clustering by rule and firm-grantee pair. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Rule outcomes - Recent donation

Dependent variable	Similarity between comment submitted by firm f and discussion text in rule r				Log citation count		Any citation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
At least one grantee g co-commenting and receiving donation from firm f in year t or $t-1$	0.156***	0.155***	0.110**	0.107**	0.288**	0.287**	0.126**	0.129**
	(0.051)	(0.051)	(0.049)	(0.049)	(0.134)	(0.134)	(0.061)	(0.060)
Log expenditure lobbying agency in t and $t-1$		0.002		0.007**		0.002		-0.005
		(0.004)		(0.004)		(0.007)		(0.003)
Fixed Effects								
Rule	Y	Y	Y	Y	Y	Y	Y	Y
Firm	Y	Y	Y	Y	Y	Y	Y	Y
Commenter Style Control			Y	Y	*	*	*	*
Observations	4,675	4,675	4,665	4,665	1,212	1,212	1,212	1,212

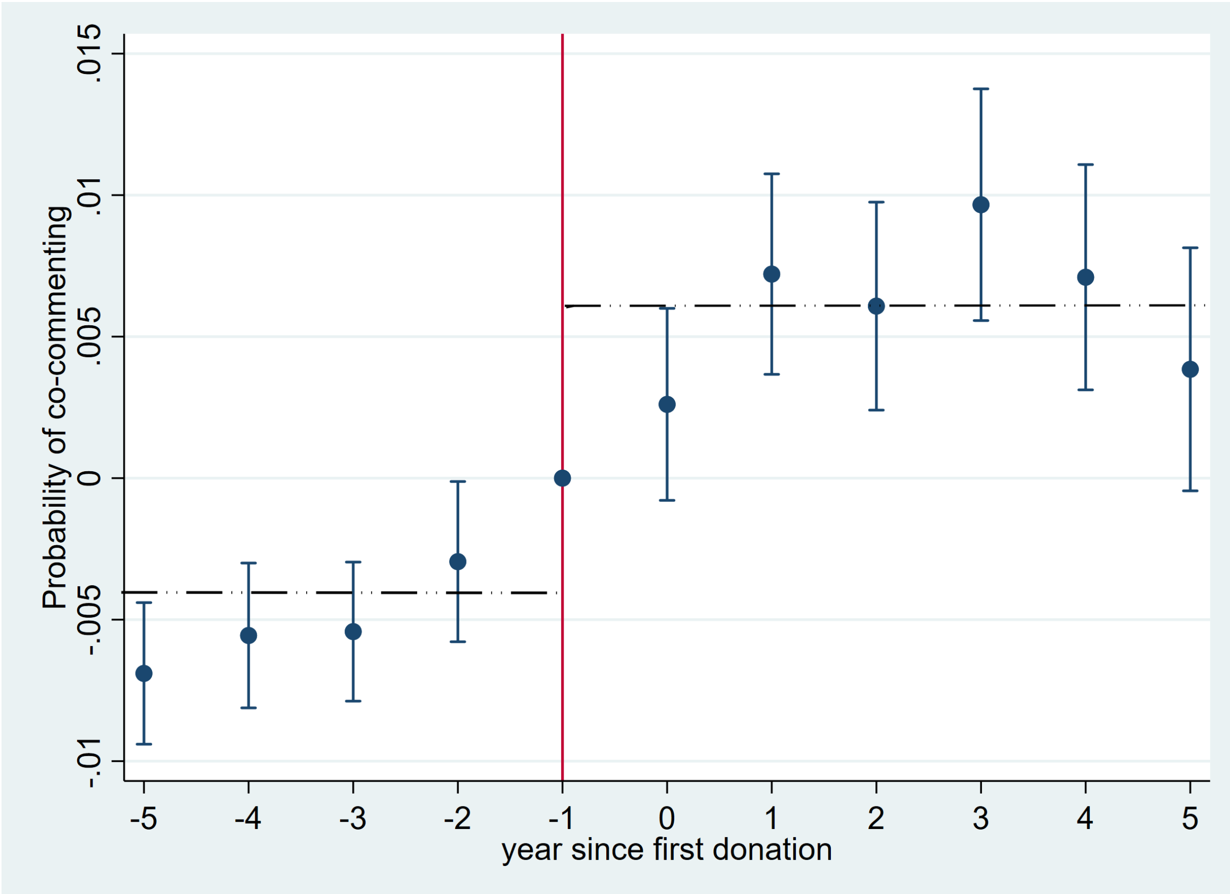
Notes: The dependent variables are several measures of the relationship between firm comments and the discussion of comments in subsequent rules. For columns 1-4 the outcome is the overall similarity of the text, for columns 5 and 6, the outcome is log of the number of detected occurrences of the firm's name in the discussion text, and for columns 7 and 8 the outcome is an indicator for the presence of at least one occurrence of the firm's name in the discussion text. Some agencies rarely cite any commenters by name, so in columns 5-8 we restrict the sample to agencies whose rules contain an average of at least one citation of a firm per rule. The independent variable is equal to one if there is at least one grantee g co-commenting on regulation r and receiving a grant from firm f in year t or $t-1$. The asterisks (*) for Commenter Style Control in the citation columns indicates that the outcome measure is not adjusted, but the comment similarity with style control is used to find best matched rule for each comment. The sample includes all rule-firm pairs, conditional on the firm commenting on the rule. Standard errors use two-way clustering by rule and firm. *** p<0.01, ** p<0.05, * p<0.1

Table 9: 1994 Budget Recommendations Mobil Foundation, Inc.

Grantee Organization	**Listed Benefits to Mobil**	**Amount in Budget Year 1994**	**Grants by Other Corporations in Budget Year 1993**
American Council on Science and Health	*One of the major benefits is due to the Executive Director, Dr. Elizabeth Whelan, an articulate spoke person who often appears as a counterpoint to so-called "public interest" groups.* [...]	$15,000	Stauffer $10,000; Chevron $10,000; Ashland $25,000; Union Carbide $10,000; ARCO $10,000; Shell $15,000; Exxon $10,000
National Safety Council	*Mobil is currently represented on the Board of Directors by David E. Miller and has representatives from the operating divisions on various working committees. By supporting this unique life-saving organization, Mobil is identified as a co-leader in the Council's life-saving mission. Additionally, we were successful in 1989 in having the National Safety Council Board of Directors pass a resolution opposing the mandating of any alternative fuel, such as methanol, until studies demonstrated a reduced risk of death, illness, or injury.* [...]	$35,000	DuPont $30,000; Chevron $17,000; Shell $10,000; Amoco $10,000; Texaco $7,500; Ford $10,000; IBM $20,000; AT&T $12,500
The Academy of Natural Sciences	*Based on the contacts of the Academy, the Environmental Associates Program has the potential to challenge the EPA behind-the-scenes on the effectiveness of a regulation for the environment and whether sound science supports the proposed law.*	$15,000	Air Products and Chemicals $15,000; ARCO $15,000; DuPont $15,000; FMC Corp $15,000; [...]
Bermuda Biological Station for Research	[...] *Dr. Knap's findings are generally quite supportive of oil industry activities and could also influence legislation and regulations favorable to Mobil's U.S. offshore operations. [...] This type of expertise would be most helpful to Mobil not only during the cleanup portion of a spill, but to gather data and provide testimony during litigation concerning environmental damage. Awards from this type of litigation can exceed the total cost of clean-up and mitigation activities.*	$15,000	General Atlantic Grp $100,000; Exxon $50,000; Texaco $40,000; X. L. Foundation $40,000; Bacardi int. $50,000; Transworld Oil Ltd. $25,000
Greater Caribbean Energy and Environment Foundation	[...] *Dr. Thorhaug represents a valuable Mobil interface with UNEP and UNDP oil spill related activities particularly in third world tropical countries. Her publications and input can have a positive influence in evolving regulations in lesser developed nations.*	$10,000	Exxon $6,000; World Bank $25,000; UN Unesco $15,000; Kenya Govt. $30,000; Philippines Govt. $6,000; UNEP $8,000
Harvard School of Public Health Center for Risk Analysis (HCRA)	*HCRA (and Director John Graham in particular) is recognized as a leading authority in the application of risk analysis to public policy. It has been effective in influencing Clean Air Act legislation on toxic emission standards and pointing out the safety risks associated with excessively stringent fuel economy standards. We expect the Center to play an influential role in consideration of future environmental legislative and regulatory actions.*	$12,500	N/A

Source: Research, Engineering & Environmental Affairs, July 1993, "1994 Budget Recommendations Mobil Foundation, Inc.". Note: Total Budget = $1,217,200

Figure 1: Event Study for Co-Comment Activity After a Donation



Notes: The unit of observation for this analysis is the firm-grantee-year. The dependent variable is a dummy variable that equals 1 if the grantee and the firm co-comment at least once in that year, 0 otherwise. For firm-grantee pairs where there is at least one donation over the sample period, we define the event date as the time of the first donation. To focus on a subset of 'clean' events, we exclude from the event study firm-grantee pairs where a second donation occurs within 5 years of the first donation. We further restrict the event study to firm-grantee pairs for which we have at least 5 years of data prior and post the first donation. Finally, for the subsample of firm-grantee pairs that meet the above criteria for inclusion in the event study, we only include 5 years of data prior and post event. Firm-grantee pairs for which we observe no donations over the sample period are used as controls. We then regress the co-comment dummy on a vector of dummies for 5 leads and 5 lags indicators, the event dummy, calendar year fixed effects and firm-grantee pair fixed effects, clustering standard errors at the firm-grantee level. The event study graph reports the estimated coefficients on the lead, event and lag dummies, all relative to one year before the donation, as well as a 95% confidence interval.

ONLINE APPENDIX - NOT FOR PUBLICATION

Hall of Mirrors: Corporate Philanthropy and Strategic Advocacy

Marianne Bertrand, Matilde Bombardini,
Raymond Fisman, Brad Hackinen, Francesco Trebbi

A Appendix: Regulation comments

Our data on regulatory comments come from regulations.gov. Under the Administrative Procedures Act (APA), federal agencies must provide a means for the public to submit comments on proposed rules and other regulatory changes. Regulations.gov is a shared platform that is now used by most federal agencies to facilitate submission and public review of comments. Information about submitted comments, including the original text and attachments, can be viewed through a web browser. The site also provides an API that allows for more efficient data access, particularly for collecting simple comment metadata such as the title of the comment and posted date.

A.1 Overview

Our sample starts with the complete collection of metadata for all comments posted to regulations.gov in the years 2003-2017 (inclusive), yielding a total of 6,871,697 unique documents. From these, we identify 981,232 comments that appear to be authored by organizations rather than private individuals (org comments). We download the complete text for all org comments using common file formats, giving us about 90% of comment text for the org comment sample.

Before moving to a more detailed description of the comment and rule text collection it is worth describing the time dimension of the data. In the early period we are limited by the availability of comment data. Regulations.gov went online in 2003, but it was initially used by only a handful of agencies. Figure C.1 shows the number of proposals published in the Federal Register that direct commenters to regulations.gov. Proposals without a regulations.gov link would have provided alternate contact information such as an agency email address or internal comment management system, and comments submitted on these proposals are not available in our data. The plot shows that the fraction of proposals with a regulations.gov link increased gradually over time, reaching about 80% in 2008. The fraction rose to nearly 90% by 2018, but we have only limited comment data for the 2003-2008 period. In more recent years we are limited by the fact that FoundationSearch may take several years to post data on each firm. Overall, these constraints mean that we have only partial data for 2003-2007 and 2014-2015, and our best coverage is in the 2008-2014 period. This pattern is presented graphically in Figure C.2 which plots the number of co-comments with financial ties by year. The clear hump shape is driven by data availability. In our regressions we generally include the whole 2003-2016 sample, but drop firm-year observations with missing donation data and use year or other year-interacted fixed effects to control for time-varying comment coverage (and other time trends). Finally, when linking comments to rules, we use all rules published in the Federal Register in any year up to 2017. We include this extra year of data because it often takes a long time for agencies to develop the final rule after receiving comments, and some comments from 2016 could be linked to a rule published in 2017.

A.2 Collecting metadata

The regulations.gov API provides a search function for document metadata. We retrieved the metadata for all public submission documents posted since the site came online in 2003, and include all years up to and including 2017. Some agencies have begun digitizing older comments and posting them to regulations.gov retroactively. However, an EPA spokesperson stated (in personal email correspondence) that this work is currently incomplete, and that the text of some older comments will never be released digitally since the submitters were not aware of this possibility at the time. Thus we consider data on pre-2003 comments on regulations.gov unreliable and do not include them.

A.3 Identifying org comments

Authorship information can appear in three different metadata fields: "title", "organization", or "submitterName." Comments appear to fall into two main types: those that contain "organization" and/or "submitterName" information, and those that only contain authorship information in the title. First, we drop all comments that have "submitterName" information, but no organization. These appear to be written by private individuals. For the remaining comments, we look for an organization name in either the organization field or the title (if the organization field is blank). We use a custom neural-network-based classifier to extract organization names from the selected field (classification is necessary for the organization field because it contains many false positives such as "self" or "none"). The classifier converts each title string to ASCII characters and predicts whether each character is part of an organization string. Contiguous chunks of characters with predicted probability greater than 0.5 are counted as organization names. The classifier is multi-layer bi-directional Gated Recurrent Unit (GRU), implemented in *PyTorch*[59]. Code is available on Brad Hackinen's github page[60]. The classifier is trained on almost 9000 manually constructed training examples. This training set was constructed iteratively by starting with easily parsed titles, fitting the neural network, estimating the classifier's uncertainty from the total entropy of the character-level predicted probabilities, reviewing a sample of high-entropy titles, adding them to the training set, and repeating until the error rate was acceptably low. We also manually classified an additional set of 1000 random titles as a test set. The results of the test are shown below. 93% of titles are classified without error. 83% of titles with an organization are extracted exactly correctly, while 98.5% of titles with no org are extracted correctly (in other

[59]https://pytorch.org/
[60]https://github.com/bradhackinen/subex

words, the classifier avoids 98.5% of false positives).

A.4 Collecting comment text

Comments on regulations.gov can have comment text in two locations: a "text" field in the comment metadata, or in one or more attachments. The "text" field contains text that submitters have entered on a web form. It is often as brief as "see attached." Most substantial text is contained in the comment attachments where submitters can upload PDFs, Word documents, other other file formats. We download all attachments of the following formats: PDF, MS Word 8, MS Word 12, and simple .txt files. The majority of attachments are in PDF format.

We use the XpdfReader *pdftotext* [61] command-line utility to extract text from most PDFs. Some PDFs contain only images of each page. In this case we must fall back on Optical Character Recognition (OCR), which we implement with a combination of *GhostScript*[62] (to render page images) and *Tesseract-OCR*[63]. We use *Apache Tika*[64] to extract text from MS Word formats, and the *chardet*[65] Python package to detect formatting of simple text files. All of these tools are open source.

A.5 Linking comments to rules

This section discusses the link between comments and final rule discussion, which forms the basis of the analysis in Section 5. A practical challenge for this analysis is that regulators do not provide clear direction regarding which comments are addressed in which rule. We use a variety of document identifiers to link comments to rules, in most cases narrowing the set of possibilities to a one or two rules for each comment. We describe the procedure in three steps:

i. Comments have two pieces of information to facilitate the link to a rule: a docket identifier and a submission date (the date that the comment is posted by the agency to regulations.gov). Unfortunately, many rules have a different docket identifier than the preceding document which called for comments. As a result we first need to link Federal Register documents together to determine which comments could potentially be cited. This is a surprisingly difficult task, as agencies are quite inconsistent in how they use dockets and other identifiers. The federalregister.gov API provides a variety of useful information about Federal Register documents including publication date, associated dockets identified by the agency, affected sections of the Code of Federal Regulations (CFR), regulation identifier numbers, title,

[61] https://www.xpdfreader.com/pdftotext-man.html
[62] https://www.ghostscript.com/
[63] https://github.com/tesseract-ocr
[64] http://tika.apache.org/
[65] https://pypi.org/project/chardet/

action description, and topic keywords. Our script uses fuzzy matching techniques to find documents that are linked by keywords and other identifiers which are associated with a small number of Federal Register documents (ideally, only two). For example, one proposal document might be linked to a rule document because they share the same title. Another pair of documents might be linked because they share a docket number and affect the same CFR sections. Some documents share unusual identifiers with a relatively large number of other documents, and the script attempts to reduce the number of large linked clusters by down-weighting documents that have many potential matches.

ii. Once we have linked Federal Register documents to each other, we link comments to Federal Register documents by docket, and assume that the comment could be discussed by any rule that is a) connected to the original call for comments by a chain of linked documents, and b) published after the comment was submitted.

iii. Given this imperfect matching, we take an additional step before running each regression: when comments are potentially linked to multiple rules, we match the comment to the rule discussion with the highest similarity to the comment content (according to whichever version of the similarity measure is in use). Thus, S_{fr} can also be interpreted as the maximum similarity with any subsequent rule linked to the comment. In this context, we define a grantee as co-commenting with a firm if they commented on the same docket and in the same year as the firm comment which was linked to the rule.

Table A.1: Organization name extraction accuracy

Sample	Count	Character Accuracy	String Accuracy
All test titles	1000	0.970	0.928
Test titles containing org	371	0.935	0.830
Test titles with no org	629	0.991	0.986

Notes: Character accuracy is the average fraction of characters classifier correctly in each title. *String accuracy* is the fraction of titles with every character correctly classified.

Table A.2: List of Agencies on regulations.gov (A-F)

ACF	Children and Families Administration	DOI	Interior Department
AHRQ	Agency for Healthcare Research and Quality	DOJ	Justice Department
AID	Agency for International Development	DOL	Employment Standards Administration
AMS	Agricultural Marketing Service	DOS	State Department
AOA	Aging Administration	DOT	Transportation Department
APHIS	Animal and Plant Health Inspection Service	EAB	Economic Analysis Bureau
ARS	Agricultural Research Service	EAC	Election Assistance Commission
ASC	Appraisal Subcommittee	EBSA	Employee Benefits Security Administration
ATBCB	Archit. and Transportation Barriers Compliance Board	ED	Education Department
ATF	Alcohol, Tobacco, Firearms, and Explosives Bureau	EDA	Economic Development Administration
ATSDR	Agency for Toxic Substances and Disease Registry	EEOC	Equal Employment Opportunity Commission
BIA	Indian Affairs Bureau	EERE	Off. Energy Efficiency and Renewable Energy
BIS	Industry and Security Bureau	EIB	Import Export Bank of the United States
BLM	Land Management Bureau	EOIR	Executive Office for Immigration Review
BOEM	Ocean Energy Management Bureau	EPA	Environmental Protection Agency
BOP	Prisons Bureau	ESA	Employment Standards Administration
BOR	Reclamation Bureau	ETA	Employment and Training Administration
BPD	Public Debt Bureau	FAA	Federal Aviation Administration
BSEE	Safety and Environmental Enforcement Bureau	FAR	Federal Acquisition Regulation System
CCC	Commodity Credit Corporation	FBI	Federal Bureau of Investigation
CDC	Centers for Disease Control and Prevention	FCIC	Federal Crop Insurance Corporation
CDFI	Community Development Financial Institutions Fund	FDA	Food and Drug Administration
CFPB	Consumer Financial Protection Bureau	FEMA	Federal Emergency Management Agency
CMS	Centers for Medicare Medicaid Services	FFIEC	Federal Financial Institutions Exam. Council
CNCS	Corporation for National and Security Service	FHWA	Federal Highway Administration
COE	Engineers Corps	FINCEN	Financial Crimes Enforcement Network
COLC	U.S. Copyright Office, Library of Congress	FISCAL	Bureau of the Fiscal Service
CPSC	Consumer Product Safety Commission	FMCSA	Federal Motor Carrier Safety Administration
CSREES	Coop. State Research, Education, and Extension Service	FNS	Food and Nutrition Service
DARS	Defense Acquisition Regulations System	FRA	Federal Railroad Administration
DEA	Drug Enforcement Administration	FS	Fiscal Service
DHS	Homeland Security Department	FSA	Farm Service Agency
DOC	Commerce Department	FSIS	Food Safety and Inspection Service
DOD	Defense Department	FSOC	Financial Stability Oversight Council
DOE	Energy Department	FTA	Federal Transit Administration

Table A.3: List of Agencies on regulations.gov (F-Z)

FTC	Federal Trade Commission	OJP	Justice Programs Office
FWS	Fish and Wildlife Service	OMB	Management and Budget Office
GIPSA	Grain Inspection, Packers and Stockyards Adm.	ONRR	Natural Resources Revenue Office
GSA	General Services Administration	OPM	Personnel Management Office
HHS	Health and Human Services Department	OPPM	Procurement and Property Management, Office of
HHSIG	Inspector General, Health and Human Serv Dept	OSHA	Occupational Safety and Health Administration
HRSA	Health Resources and Services Administration	OSM	Surface Mining Reclamation and Enforcement Office
HUD	Housing and Urban Development Department	OTS	Thrift Supervision Office
ICEB	Immigration and Customs Enforcement Bureau	PBGC	Pension Benefit Guaranty Corporation
IHS	Indian Health Service	PCLOB	Privacy and Civil Liberties Oversight Board
IRS	Internal Revenue Service	PHMSA	Pipeline and Hazardous Materials Safety Adm.
ITA	International Trade Administration	PTO	Patent and Trademark Office
LMSO	Labor-Management Standards Office	RBS	Rural Business-Cooperative Service
MARAD	Maritime Administration	RHS	Rural Housing Service
MMS	Minerals Management Service	RITA	Research and Innovative Technology Administration
MSHA	Mine Safety and Health Administration	RUS	Rural Utilities Service
NHTSA	National Highway Traffic Safety Administration	SAMHSA	Substance Abuse and Mental Health Services Adm.
NIFA	National Institute of Food and Agriculture	SBA	Small Business Administration
NIGC	National Indian Gaming Commission	SLSDC	Saint Lawrence Seaway Development Corporation
NIH	National Institutes of Health	SSA	Social Security Administration
NIST	National Institute of Standards and Technology	TREAS	Treasury Department
NLRB	National Labor Relations Board	TSA	Transportation Security Administration
NOAA	National Oceanic and Atmospheric Administration	TTB	Alcohol and Tobacco Tax and Trade Bureau
NPS	National Park Service	USC	United States Courts
NRC	Nuclear Regulatory Commission	USCBP	U.S. Customs and Border Protection
NRCS	Natural Resources Conservation Service	USCG	Coast Guard
NSF	National Science Foundation	USCIS	U.S. Citizenship and Immigration Services
NTIA	National Telecommunications and Information Adm.	USDA	Agriculture Department
NTSB	National Transportation Safety Board	USPC	Parole Commission
OCC	Comptroller of the Currency	USTR	Trade Representative, Office of United States
OFAC	Foreign Assets Control Office	VA	Veterans Affairs Department
OFCCP	Federal Contract Compliance Programs Office	VETS	Veterans Employment and Training Service
OFPP	Federal Procurement Policy Office	WCPO	Workers Compensation Programs Office
OJJDP	Juvenile Justice and Delinquency Prevention Office	WHD	Wage and Hour Division

B Appendix: Construction of comment similarity measures

In sections 4 and 5 of the paper we compare the content of firm comments with the content of grantee comments and regulator discussion text. In the first case, our goal is to capture similarities between in the policies advocated for (or against) by different commenters. In the second, it is to measure how much attention the regulator has paid to different comments. Complete solutions to these problems (in the sense of replicating what a literate and informed human could deduce from reading the text) are currently beyond the frontier of natural language processing (NLP) technology. Instead, we approximate these notions with a simple and robust method of text analysis called Latent Semantic Analysis (LSA, or sometimes called Latent Semantic Indexing) with bag-of-words features. We also introduce a small but novel adjustment to the LSA algorithm which controls for each author's average commenting style, to reduce the possibility that our results are driven by spurious correlations between fixed aspects of the text like writing style or document formatting.

The basic recipe is as follows: After collecting and cleaning the comment text (to remove headers, page numbers, and so forth), we convert each comment into a vector of word counts. We drop very rare and very common words and weight the remaining counts using a standard term frequency-inverse document frequency (TF-IDF) function to emphasize the words that are most useful in distinguishing between documents. These weighted word counts are combined into a large, sparse term-document matrix, which is then factored using singular value decomposition to generate vectors representing each document. Finally, the pairwise document similarity is computed as the cosine similarity between the document vectors. The rest of this section explains these steps in greater detail, and describes a docket classification test we conducted to verify the effectiveness of the approach and choose the dimensionality of the document vectors.

B.1 Sample construction

Both comments and rules contain text that is not relevant for our desired similarity measure. Comments are usually formatted as letters with addresses at the top, page headers and footers, and sometimes additional contact information at the end. Optical character recognition also sometimes generates "junk" text when it encounters images with text, or poor quality scans. We use regular expressions to detect common opening and closing phrases such as "To Whom It May Concern," and "Sincerely," that occur near the beginning and end of the document, and trim away text that comes before or after these phrases. We drop any line that has less than 50% alphanumeric characters (after removing white-space), and also search for lines that occur multiple times (allowing for changes to numbers and punctuation characters) at the beginning and end of each page to filter out headers and footers. Altogether, some amount of irrelevant text remains in the sample, but it is significantly reduced relative to the raw extracted text.

Rule documents published in the Federal Register are much cleaner than comments. We use bulk XML files provided by the Government Print Office which identify individual paragraphs and headers. We start by dropping certain sections that appear in many rules but do not include discussion of comments (Agency, Action, Dates, Summary, Addresses, Contact sections, as well as all appendices and tables of contents). We then search for the keywords "comment" and "letter" (also allowing matches to any words such as "commenters" or "commented" that contain those words) to identify paragraphs, footnotes, and headers that are likely to contain discussion text. For headers containing these terms, we add all paragraphs under that header to the discussion text for that rule. For paragraphs containing these terms, we select all adjacent paragraphs that fall under the same header and add them to the discussion text. Finally, we check that the agency uses at least one of the words "commenter," "commented," "response," or "received" somewhere in the selected text. This step is useful for dropping the rules that mention "comment" or "letter" but do not actually discuss comments that have been received (for example, this sometimes occurs when the document includes a call for new comments to be submitted).

Once we have selected the text for each comment and rule, we compile all of the text files into a single corpus. Some comments have multiple attachments, and commenters occasionally submit multiple times to a single docket (though this is quite rare). We concatenate all text submitted by each organization to a single docket within a calendar year and treat each of these concatenated texts as a single document. We drop any comments that are highly repetitive (in which the set of unique lines that are more than 25 characters long is less than a third of the total number of lines that are more than 25 characters long). This step drops a small number of comments in which the agency combined many form-letter submissions into one very long file. Then we clean the text by removing all punctuation except that which occurs inside words as a part of acronyms like "U.S.", or hyphenated terms. Finally, we convert all mixed-case words to lower-case, and keep all-uppercase words as is (so that "US" is not converted to "us" for example).

B.2 Generating document vectors

Given the size of our dataset, both in terms of the number and the length of documents, it was important for us to identify an algorithm that is computationally very efficient. Some algorithms require independent comparisons of each document pair, thus making them very costly for our problem (for example, recent methods involving optimal transport distance measures, or older set-based measures like the Jaccard Index). We focused instead on algorithms that generate dense vector representations of each document. These document vectors can then be used to quickly compute cosine similarity measures between many pairs of documents in parallel. We initially considered three candidate algorithms: Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), and doc2vec. LDA is analyzed in Appendix D. LSA and doc2vec were both able to efficiently generate large document vectors in a reasonable amount of time, so we ran a systematic test to examine the performance of both algorithms for our data.

B.2.1 LSA implementation

Our LSA implementation is standard. We load the corpus, split the text on whitespace to break it into discrete tokens, count the number of times each token occurs in each document, and the number of documents in which each token occurs. We drop all tokens that occur in only one document (they cannot provide any information about similarity), and all tokens that appear in more than 80% of documents (these are also not very informative). Then we convert each count c_{ij} of token i in document j into a feature weight w_{ij} using a common form of TF-IDF weighting:

$$w_{ij} = c_{ij} ln(\frac{N}{n_i})$$

where n_i is the number of documents containing at least one occurrence of token i, and N is the total number of documents in the corpus. We then stack these weights into a large, sparse, feature-document matrix M and apply a truncated singular value decomposition (SVD) to compute a rank D approximation of M:

$$M \approx A_D \Sigma_D B_D^T$$

where Σ_D is a diagonal matrix containing the D largest singular values of M. We discard A_D and take the singular value-scaled matrix $V := B_D \Sigma_D$ as our set of LSA document vectors. The word “latent” in “Latent Semantic Analysis” refers to the idea that compressing the full feature-document matrix to a lower-dimensional approximation often squeezes synonyms and other co-occurring words into the same singular vectors, improving the quality of the document model. The amount of compression is determined by the parameter D, which we choose using an empirical test described below.

B.2.2 Doc2Vec implementation

Doc2vec is an algorithm for constructing vector representations of documents by learning to predict word occurrences in the text (Le and Mikolov, 2014). It is attractive because it is computationally efficient and scales well for large corpus sizes. We rely on the gensim implementation (Řehůřek and Sojka, 2010). We train the model for 10 epochs, using the negative sampling version of the algorithm with 10 negative samples, a window size of 10, and minimum word count of 5. As with LSA, we experiment with different values for the vector size D.

B.3 Similarity measures

B.3.1 Cosine similarity

For any given document vectors v_i and v_j, our standard measure of document similarity is the cosine of the two vectors:

$$\theta_{ij} = \frac{v_i \cdot v_j}{\|v_i\| \, \|v_j\|}$$

B.3.2 Controlling for commenting style

One of the major challenges in working with the comment data is that the free-form nature of the comment documents makes it difficult to distinguish between substantive content and superfluous text. In our sample construction step, we remove as much extraneous material as possible. But some superfluous text is harder to detect. For example, many organizations spend the first paragraph or two describing themselves – how large they are, where they operate, what products they provide, how many workers they employ. Superficially, these paragraphs do not look any different from later paragraphs which describe the organization's positions on the regulation under discussion, so it is hard to remove them without a deep understanding of the text. But similarities between these paragraphs and other text are not what we wish to capture in our similarity measure. For example, we would not want our co-commenting similarity results to be driven by firms donating to grantees with similar self-description paragraphs. Another concern relates to the very diverse set of organizations that submit comments. When reading through comments, it quickly becomes apparent that some organizations use complex scientific and legal jargon, while others write in plain, even casual, language. We do not want our comment similarity measure to be biased by firms preferentially donating to grantees with a similar level of linguistic complexity.

One improvement we can make is to ensure that our similarity measure focuses on content and linguistic patterns that are not part of a recurring pattern for a particular organization. The solution is analogous to fixed effects in panel data. We often believe that individuals have a specific average outcome that is separate from the variation we aim to measure. Including individual fixed effects in the regression controls for this average outcome, and the resulting estimates depend only on within-individual variation. In the case of comments, we can think of each commenter as having an average commenting style that incorporates the boilerplate text, self-description content, and tendency to use more or less sophisticated language. If we "subtract" each organization's average comment, we control for these stylistic dimensions and ensure that our measure depends only on within-commenter variation. Depending on how text is represented, it might not be clear how to subtract one comment from another, or take an average across documents. Fortunately, one advantage of our document vector-based approach is that linear operations on these vectors are simple and conceptually clear. Suppose that v_{it} is the document vector corresponding to organization i's comment in docket-year t. Then we control for commenting style of organization i by constructing the de-meaned vector

$$\tilde{v}_{it} = v_{it} - \frac{1}{|T_i|} \sum_{t \in T_i} v_{it}$$

where T_i is the set of periods when i submitted a comment. For both LSA and doc2vec document vectors, this operation is roughly equivalent to subtracting the average number of occurrences of each token across documents by the same organization before computing the vectors (but much more computationally efficient). We then compute our new similarity measure that controls for comment style as the cosine similarity between demeaned vectors:

$$\tilde{\theta}_{ijt} = \frac{\tilde{v}_{it} \cdot \tilde{v}_{jt}}{\|\tilde{v}_{it}\| \, \|\tilde{v}_{jt}\|}$$

At this point, the analogy with fixed effects breaks down somewhat since cosine similarity is a non-linear operation. However, we believe the intuition holds: de-meaning the comments within each organization prior to estimating the relationship between comment similarity and donations prevents many spurious correlations that could be driven by similarities between the average comment style of firms and grantees or between firms and regulators, and instead focuses the similarity measure on aspects of the text that change from comment to comment.

It is worth noting that using this procedure to control for comment style is different from including separate firm and grantee fixed effects in the similarity regressions. Separate firm and grantee fixed effects can only control for the average similarity of a particular firm or grantee to all organizations with which it co-comments. However, because co-commenting is not random, this average similarity could equally be driven by variation in similarity over time within a firm-grantee pair (what we aim to measure), or cross-sectional correlations between the commenting style of firms and grantees who co-comment (what we want to avoid). On the other hand, firm-grantee pair fixed effects do eliminate the same cross-sectional variation in comment similarity (as well as cross-sectional variation in donations). However, these pair fixed effects are only identified for a small portion of our firm-grantee pairs, and so are necessarily limited in precision. Controlling for comment style by demeaning vectors within organization offers an intermediate level of control between separate firm and grantee fixed effects and pair fixed effects specifications, and can be estimated even when organizations co-comment only once.

B.4 Docket-match prediction test

There are many ways to construct a similarity measure between documents and this flexibility introduces extra degrees of freedom into our analysis of comments and rules. At the same time, it seems reasonable to expect that some approaches to measuring similarity will work better than others by some objective metric. The trouble is that selecting the similarity measure based on our regression results would surely introduce bias. To solve this problem, we decided to select our similarity measure according to performance on a completely separate benchmark task.

We evaluate similarity measures according to how useful they are in predicting whether two randomly chosen comments come from the same docket. We feel this serves as an appropriate benchmark for two reasons. First, it can be computed using our actual data. The performance of different similarity measures can vary considerably across datasets (for example, see the performance comparisons in Yurochkin et al., 2019), so it is important that we do not need to extrapolate from some other data that may have different properties. Second, all comments have docket information, and these labels are among the most reliable pieces of information about the comment. We may thus run the test at large scale, without worrying about additional noise introduced by imperfect linking or missing data. A similarity measure that provides good predictive information about whether comments come from the same docket must necessarily be capturing whether comments discuss the same narrow topics. This is not exactly the same as our goal of detecting parallel arguments between comments, but we believe it is sufficiently comparable to be a useful benchmark for evaluating similarity measures.

To construct the sample for the test, we select one pair of comments from every unique docket-year (after dropping both docket-years and commenters with only one comment). These become our "matched" observations. We then sample an equal number of random comment pairs in which the two comments come from different dockets. These become our "unmatched" observations. We run two versions of the test: "random pairs" and "same-agency pairs." In one version of the test, the unmatched comments can come from any other docket in our data. In a harder version of the test, the unmatched comment pairs are restricted so that both comments were submitted to the same agency. For example, one comment might have been submitted to the EPA regarding an air quality regulation, and the other submitted to the EPA regarding a water quality regulation. These comments are likely to be more similar to each other than to comments submitted to the FDA regarding medical device testing, or the FAA regarding the maintenance of a specific airplane part. Thus, the "same-agency" version of the test emphasizes distinctions between comments that are already relatively similar, potentially making it a better match for our co-comment analysis.

To measure the accuracy of a given similarity measure, we first construct document vectors using the entire corpus as our training data, then compute the cosine similarity between the document vectors for the comment pairs in the test sample. We score each similarity measure based on how well a fitted logistic regression can predict out-of-sample pairs, using the cosine similarity measure as the only feature. We use a 5-fold hold-out strategy, fitting the model on 80% of the data and predicting the remaining 20% to generate one prediction for each observation. Observations are predicted to be a "match" if the predicted probability given the similarity is greater than 0.5, and our reported accuracy score is the fraction of pairs for which the predicted match value is equal to the true match value. Given our balanced sample, a completely uninformed guess would obtain 50% accuracy. This measure essentially asks how well the comment pairs can be sorted into matched and unmatched pairs by choosing a single threshold similarity and classifying all pairs with similarity higher than the threshold as matched, and lower than the threshold as unmatched. It would be very surprising if comments in a given docket are so similar to each other and so different from comments in other dockets that the classifier could achieve 100% prediction accuracy using only a one-dimensional similarity measure.

We test two algorithms, LSA and doc2vec, with 5 logarithmically spaced values for D ranging from 64 to 1024. This parameter effectively controls the amount of information that can be contained in the document vectors, and setting D appropriately has a large effect on the accuracy. Intuitively, there is potentially a trade-off between the benefits of compressing the data to reduce noise (low D) and allowing the vectors to capture enough detail to discern between similar documents (high D). For each algorithm and D, we compare the performance of both the random pairs and same-agency pairs, with and without organization demeaning.

Figure C.3 shows the results of the test. We observe several interesting patterns. First, LSA always performs better than doc2vec, unless D is very small. Second, the performance of LSA is highest when D is large. Given the slope of the curve, it seems possible that even larger vectors would further improve performance. However, $D = 1024$ was the largest LSA vector size we were able to compute on a fairly capable computer with 128 GB RAM. LSA with $D = 1024$ achieves an 93% accuracy on the basic docket-match prediction task with random unmatched comments. We find this quite reassuring, as it suggests that LSA is very good at detecting systematic similarities and differences in the content of comments. As expected, the task is harder when unmatched comments come from the same agency. Here LSA achieves 78% accuracy. Demeaning the document vectors by organization also consistently makes the task harder. Fortunately, LSA with $D = 1024$ achieves the best accuracy on every version of the test, making it a clear choice to use for our analysis.

B.5 Constructing co-comment and rule-comment similarities

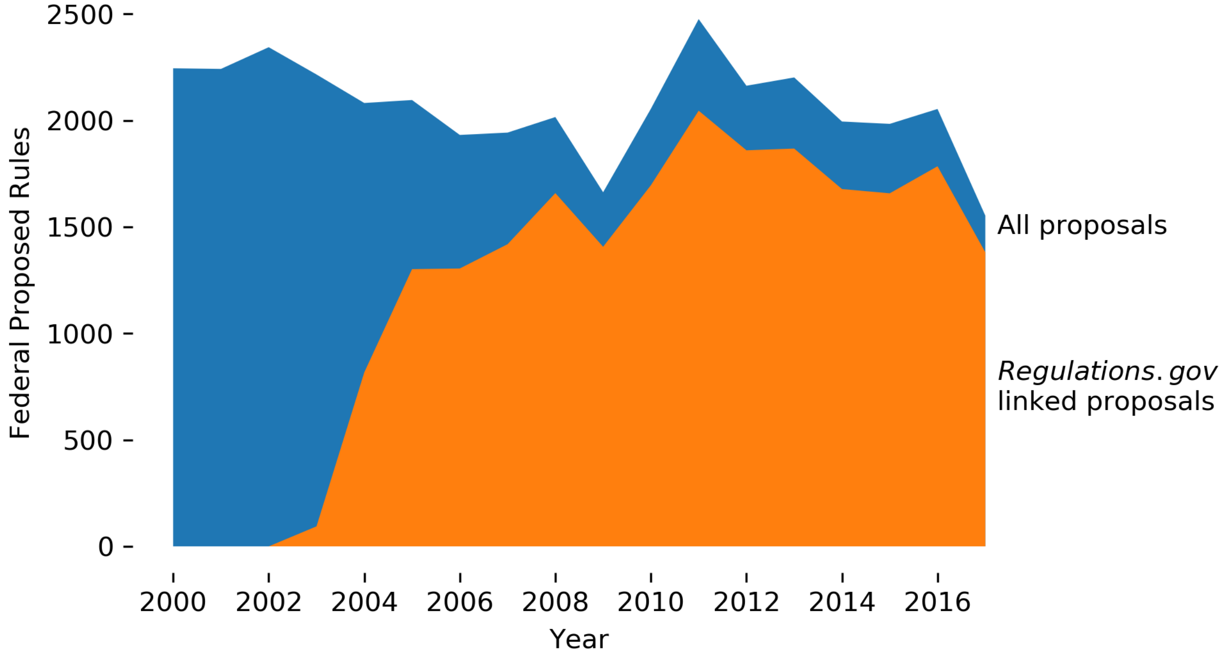
Based on our results from the docket-match prediction test, we use LSA document vectors with $D = 1024$ to construct all of our similarity measures. Constructing the co-comment similarity measures is straightforward. We build a corpus as described above using all organization comments for the largest possible sample such that all commenters and dockets have at least two comments (where a single "comment" is actually all the text submitted by a particular organization to a specific docket in one calendar year), and construct document vectors for each comment. From these comment vectors we produce an additional set that are de-meaned by organization. We then compute the cosine similarity for every co-comment pair.

Estimating the similarity between the rule discussion and an organization's comment(s) is a little more complicated. We start by compiling a slightly larger corpus that contains all comments and all rule discussions. We construct a new set of LSA vectors with $D = 1024$, and then compute the cosine similarity between every linked firm-rule pair. In the case that there are multiple rules linked to a comment, we compute all possible similarities, and then select the observation with the highest similarity to include in the regression. This step is a solution – albeit an imperfect one – to dealing with cases in which the correct match is unclear. There are several reasons why comments might be linked to multiple rules. First, it is possible that the comment-rule linking algorithm generated one or more false positive matches. However, even when the matching is perfect, it is possible to have multiple rules linked to a comment. For example, agencies occasionally publish a short rule that delays implementation of the new regulation without a full discussion of the comments. It is also possible for agencies to publish corrections after the main rule is published. We omit minor corrections from our data, but larger corrections, adjustments, or interpretative guidance may motivate the agency to publish another version of the rule without discussing the prior comments. In each of these examples, only one of the rules actually discusses the linked comment leading to a meaningful similarity measure, while the other only adds noise. Selecting the rule with the highest similarity for each comment should (on average) identify the rule where that comment is actually being discussed. Even when this criterion fails to identify the correct match, there is no obvious reason that it would generate a spurious relationship between document similarity and donations.

C Appendix: Additional tables and figures

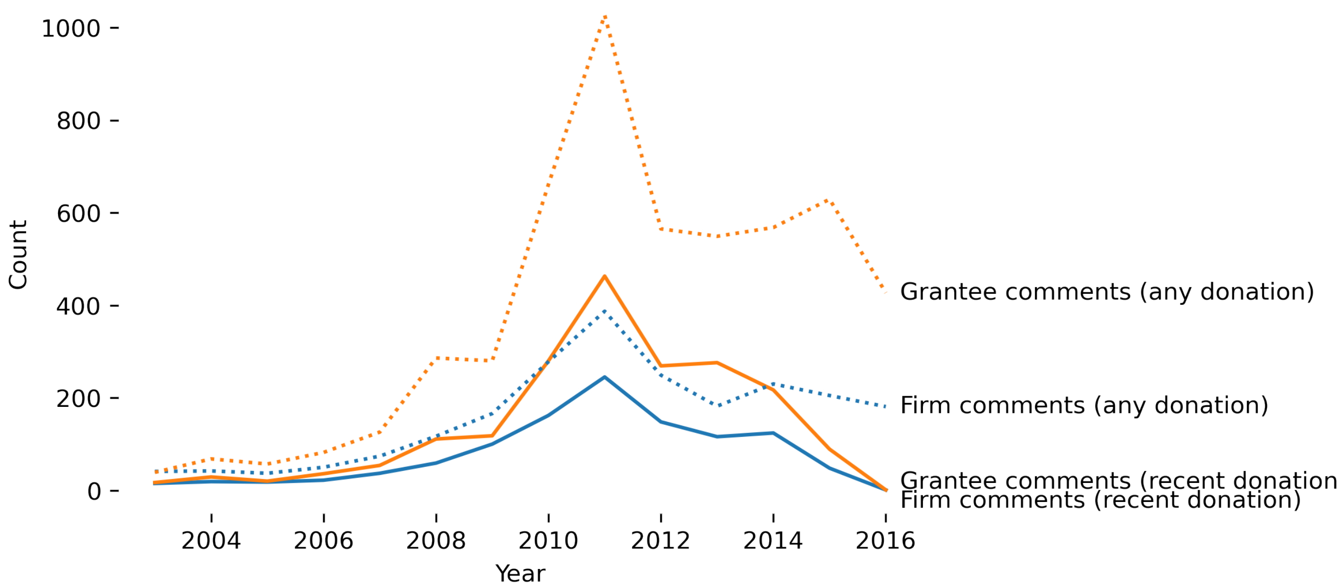
We report here various additional figures and tables mentioned in the text.

Figure C.1: Regulations.gov comment coverage



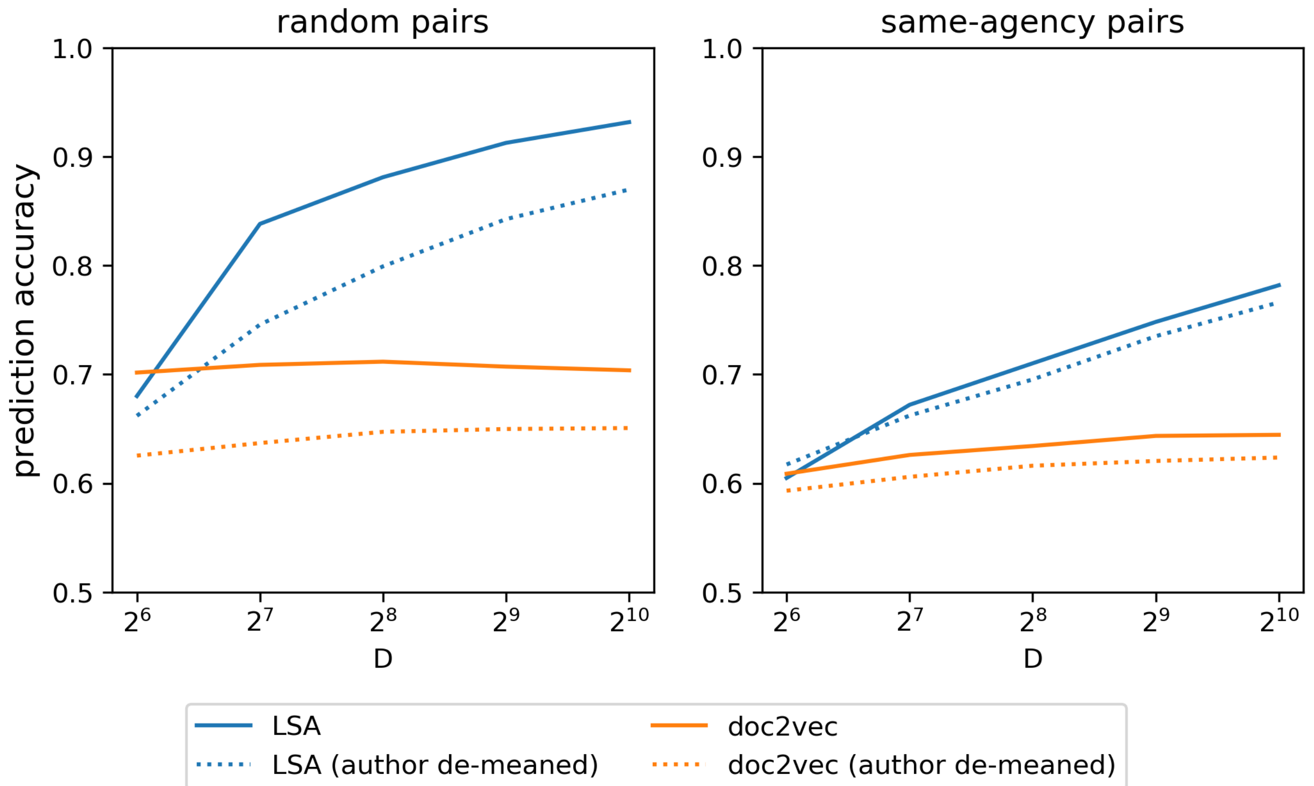
Notes: This figure shows the number of proposed regulations published on regulations.gov each year in blue. The portion that have a regulations.gov link are in orange. Those proposals that do not a have a regulations.gov link represent rule-making activity that is omitted from our data.

Figure C.2: Annual Donation Co-Comment Counts



Notes: This figure shows the number of donation co-comment events (when a firm donates to a grantee and then both comment on the same rule) by comment year. Dotted lines indicate co-comments that are associated with a donation at any point in the sample, while solid lines indicate co-comments that occur in the same year or the year following a donation. The hump shape is driven by data availability: early in the sample we are missing comment data, and late in the sample we are missing donation data.

Figure C.3: Docket-match Detection Test Results



Notes: This figure shows the results of the docket-match prediction test, in which the goal is to predict whether a given pair of comments come from the same docket using a logistic classifier with cosine similarity between the two document vectors as the only feature. The accuracy of each algorithm is plotted as a function of D, the number of dimensions in the document vectors. Accuracy is defined as the fraction of correct predictions made when fitting on 80% of the sample and making predictions on the remaining 20%. Results for LSA are in blue, doc2vec in orange. Solid lines indicate the results for unmodified document vectors, while the results using organization-demeaned vectors are plotted as dotted lines. The left panel shows results for the test where unmatched pairs are completely random, and the right panel shows results for the harder task where unmatched pairs were selected from comments submitted to the same agency.

Table C.1: Annual firm comment count distribution by commenting relationship (Significant rules only)

	Annual firm comment counts (rules per firm/year)[1]							
	Mean	Std. Dev.	Min	Max	P50	P90	P99	Total[2]
Annual comments from each firm on:								
Any rule	0.8	1.3	0	11.6	0.3	2.0	6.6	596.9
Rules where at least one grantee also comments	0.7	1.1	0	10.0	0.3	1.7	5.7	549.3
Rules where at least one grantee who receives a donation from the firm at any time also comments	0.2	0.6	0	6.6	0	0.4	2.6	134.3
Rules where at least one grantee who has received a recent[3] donation from the firm also comments	0.1	0.4	0	5.7	0	0.3	2.0	84.7

Notes: This table summarizes the number of comments submitted by each firm in a representative year on rules that are deemed "significant" under EO 12866 (computed as the average across years 2008-2014 where our data is most complete).
[1] Each firm-rule-year observation is counted as one comment. Firms that submit multiple documents (or multiple form letters as part of a coordinated campaign) on the same rule in the same calendar year are counted as submitting one comment on that rule.
[2] Total comment count for all firms in our sample.
[3] We use the term "recent" to refer to any donation which occurs in the same or previous calendar year relative to the comment year.

Table C.2: Annual grantee comment count distribution by commenting relationship (Significant rules only)

	Annual grantee comment counts (rules per grantee/year)[1]							
	Mean	Std. Dev.	Min	Max	P50	P90	P99	Total[2]
Annual comments from each grantee on:								
Any rule	0.3	0.7	0	22.7	0.1	0.6	2.7	2401.7
Rules where at least one firm also comments	0.2	0.5	0	17.9	0.1	0.4	2.0	1670.3
Rules where at least one firm who donates to the grantee at any time also comments	0.1	0.3	0	8.1	0	0.1	1.1	553.4
Rules where at least one firm who has recently[3] donated to the grantee also comments	0	0.2	0	7.4	0	0	0.7	265.3

Notes: This table summarizes the number of comments submitted by each grantee in a representative year on rules that are deemed "significant" under EO 12866 (computed as the average across years 2008-2014 where our data is most complete).

[1] Each grantee-rule-year observation is counted as one comment. Grantees that submit multiple documents (or multiple form letters as part of a coordinated campaign) on the same rule in the same calendar year are counted as submitting one comment on that rule.

[2] Total comment count for all grantees in our sample.

[3] We use the term "recent" to refer to any donation which occurs in the same or previous calendar year relative to the comment year.

Table C.3: Top Agencies by Number of Comments

Top 30 agencies in firms' comments	Number of comments	Top 30 agencies in grantees' comments	Number of comments
EPA	5242	EPA	8941
FAA	2446	CMS	5213
FDA	1600	FDA	3725
CMS	556	ED	3060
EERE	498	HHS	2827
OCC	382	FWS	2159
PHMSA	341	HUD	1638
OSHA	314	NOAA	1496
HHS	258	FNS	1414
CFPB	248	APHIS	1295
NHTSA	240	IRS	1200
NLRB	237	FAA	918
USTR	228	CFPB	788
DOT	213	DOJ	604
EBSA	207	EERE	589
IRS	184	SSA	555
FWS	170	BOEM	528
FMCSA	148	USCIS	527
USCG	130	NLRB	521
HUD	125	OSHA	512
APHIS	94	AMS	511
BIS	88	OPM	489
EIB	87	OCC	478
TSA	85	DOT	477
LMSO	84	HRSA	470
FRA	72	USCG	457
TREAS	66	OMB	456
ESA	66	ETA	441
USCBP	64	FHWA	421
AMS	58	USTR	413

Notes: This table reports the 30 top agencies as ranked by the number of comments they receive by firms (first two columns) or by grantees (last two columns). Note that we count multiple documents submitted by the same organization regarding the same rule in the same calendar year as a single comment.

Table C.4: Co-commenting and donations - Future, contemporaneous and lagged donations

Dependent variable	Firm f and grantee g commented on the same rule in year $t \times 100$			
Mean	0.175			
	(1)	(2)	(3)	(4)
Firm f contributed to grantee g in year $t+1$	0.572***	0.341***	-0.018	-0.028
	(0.042)	(0.041)	(0.046)	(0.045)
Firm f contributed to grantee g in year t	0.488***	0.276***	-0.029	-0.048
	(0.042)	(0.042)	(0.045)	(0.043)
Firm f contributed to grantee g in year $t-1$	0.801***	0.534***	0.180***	0.142***
	(0.047)	(0.046)	(0.049)	(0.047)
Fixed effects				
Year	Y	Y	Y	
Grantee		Y		
Firm		Y		
Grantee-Firm Pair			Y	Y
Grantee-Year				Y
Firm-Year				Y
Observations	102,714,672	102,714,672	102,637,658	102,637,658

Notes: Standard errors are clustered at the grantee×firm level. *** p<0.01, ** p<0.05, * p<0.1

Table C.5: Comment similarity - Recent and future Donations

Dependent variable	Similarity of comments by grantee g and firm f on same rule					
	(1)	(2)	(3)	(4)	(5)	(6)
Grantee g received donation	0.029	0.037	0.022	0.036*	0.032	0.027
from firm f at t or $t-1$	(0.018)	(0.040)	(0.022)	(0.019)	(0.044)	(0.025)
Grantee g received donation	0.004	0.002	-0.010	0.007	-0.011	0.003
from firm f at $t+1$	(0.021)	(0.043)	(0.022)	(0.022)	(0.046)	(0.024)
Fixed Effects						
Rule	Y	Y	Y	Y	Y	Y
Firm	Y			Y		
Grantee	Y			Y		
Firm-Grantee Pair		Y			Y	
Agency×NAICS×NTEEC			Y			Y
Commenter style control				Y	Y	Y
Observations	156,263	63,496	75,205	156,263	63,496	75,205

Notes: The dependent variable is a similarity index between the comment of firm f and the comment of grantee g on in the same rule r, scaled to have a standard deviation of one. The independent variables are equal to one if grantee g received a donation from firm f in the year when the comment appears or the year before (a recent donation), or the year after the comment appears (a future donation). Standard errors use two-way clustering by rule and firm-grantee pair. *** p<0.01, ** p<0.05, * p<0.1.
*** p<0.01, ** p<0.05, * p<0.1

Table C.6: Rule outcomes - Recent and future donations

Dependent variable	Similarity between comment submitted by firm f and discussion text in rule r				Log citation count		Any citation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
At least one grantee g co-commenting and receiving donation from firm f in year t or $t-1$	0.148**	0.148**	0.129**	0.129**	0.251	0.261*	0.083	0.094
	(0.063)	(0.063)	(0.057)	(0.058)	(0.158)	(0.157)	(0.069)	(0.067)
At least one grantee g co-commenting and receiving donation from firm f in year $t+1$	0.018	0.017	-0.033	-0.037	0.063	0.057	0.044	0.038
	(0.054)	(0.054)	(0.055)	(0.055)	(0.126)	(0.125)	(0.053)	(0.051)
Log expenditure lobbying agency in t and $t-1$		0.005		0.009***		-0.006		-0.007*
		(0.004)		(0.003)		(0.007)		(0.004)
Fixed Effects								
Rule	Y	Y	Y	Y	Y	Y	Y	Y
Firm	Y	Y	Y	Y	Y	Y	Y	Y
Commenter Style Control			Y	Y	*	*	*	*
Observations	4,310	4,310	4,300	4,300	1,121	1,121	1,121	1,121

Notes: The dependent variables are several measures of the relationship between firm comments and the discussion of comments in subsequent rules. For columns 1-4 the outcome is the overall similarity of the text, for columns 5 and 6, the outcome is log of the number of detected occurrences of the firm's name in the discussion text, for columns 7 and 8 the outcome is an indicator for the presence of at least one occurrence of the firm's name in the discussion text. Some agencies rarely cite any commenters by name, so in columns 5-8 we restrict the sample to agencies whose rules contain an average of at least one citation of a firm per rule. The independent variables are equal to one if grantee g received a donation from firm f in the year when the comment appears or the year before (a recent donation), or the year after the comment appears (a future donation). The asterisks (*) for Commenter Style Control in the citation columns indicates that the outcome measure is not adjusted, but the comment similarity with style control is used to find best matched rule for each comment. Standard errors use two-way clustering by rule and firm. *** p<0.01, ** p<0.05, * p<0.1

D Appendix: Applying Latent Dirichlet Allocation

Latent Dirichlet Allocation, originally developed by Blei et al. (2003) is a popular approach to topic modeling that uses Bayesian methods to group words into topics and assign topics to documents. At the most basic level, there are more similarities than differences between LDA and LSA. Both essentially perform dimensionality reduction on a large, sparse, term-document matrix to find vectors that summarize the distribution of words in each topic and document. But where LSA uses singular value decomposition to find the vectors, LDA fits a high-dimensional Bayesian model with a Dirichlet prior on the vector parameters. As such, LDA models have a clearer probabilistic interpretation and can avoid overfitting in small samples, but are also much more computationally demanding to fit. In our case, we estimate that fitting an LDA model on the full sample of comments used in our LSA analysis would take months of CPU time. To make fitting the LDA model more manageable, we restricted the sample to only comments submitted by firms and grantees. This allows us to construct the same set of pairwise firm-grantee similarity observations, only giving up the additional information about term frequencies that are contained in the non-firm/non-grantee organization comments. We chose to fit an LDA model with 1024 topics.[66] This is a relatively large number of topics, but we believe it is a good choice based on size and complexity of our corpus, while also maintaining consistency with the LSA vectors. Table D.1 presents the results of replicating our co-comment similarity regressions with the LDA vectors. The two sets of results are remarkably similar, suggesting that our co-comment similarity results are not sensitive to the choice of topic modeling approach.

[66] We used the python "lda" module (available at https://pypi.org/project/lda/) with default settings. We also experimented with Gensim and Mallet, but found them to be unstable and had trouble fitting a model on our corpus.

Table D.1: Similarity of comments - Recent donation (LDA replication)

Dependent variable	Similarity of comments by grantee g and firm f on same rule					
	(1)	(2)	(3)	(4)	(5)	(6)
Grantee g received donation	0.052***	0.062*	0.042**	0.054***	0.070*	0.049**
from firm f at t or $t-1$	(0.019)	(0.033)	(0.020)	(0.020)	(0.038)	(0.021)
Fixed Effects						
Rule	Y	Y	Y	Y	Y	Y
Firm	Y			Y		
Grantee	Y			Y		
Firm-Grantee Pair		Y			Y	
Agency×NAICS×NTEEC			Y			Y
Comment style control				Y	Y	Y
Observations	168,347	71,195	81,851	168,347	71,195	81,851

Notes: The dependent variable is a similarity index between the comment of firm f and the comment of grantee g in the same rule-year rt, scaled to have a standard deviation of one. The independent variable is equal to one if grantee g received a donation from firm f in the year when the comment appears or the year before. Standard errors use two-way clustering by rule and firm-grantee pair. *** p<0.01, ** p<0.05, * p<0.1

E Estimating comment similarity by matching on similar dockets

In this section we present the results of estimating the difference in textual similarity of comments between a firm and grantee with and without a recent donation using a matching estimator. We attempt to estimate δ, the average treatment effect of recent donations on the similarity of comments between firms and grantees, conditional on co-commenting. The advantage of using a matching estimator is that we can be careful about how we construct a counterfactual similarity for each treated observation. In our fixed effects regressions it is difficult to distinguish between two competing explanations for the apparent changes in similarity: one explanation is that donations cause grantees to change the content of their comments. An alternative explanation is that there is a great deal of diversity in regulations and firms choose to donate to grantees that will be aligned with them on upcoming regulatory topics. In this second model, the apparent increase in comment similarity is actually due to a selection effect that would not be present if we could also condition on regulatory topics. A matching estimator allows us to address this concern by only using comments from the most similar regulations to estimate the untreated counterfactual. While we cannot be certain that these similar regulations are sufficiently comparable so as to eliminate the possibility of selection. the matched estimate may nonetheless be an improvement over the fixed effects regressions on this dimension.

We construct our matching estimator as follows. Each observation is a similarity measurement s_{fgkt} between a firm f and grantee g comments in docket k and year t. The goal is to estimate δ, which is defined as the expected difference in similarity for an observation that has been "treated" by a donation in either year t or $t-1$ ($D_{fgt} = 1$) and the similarity that would be observed in a counterfactual untreated state ($D_{fgt} = 0$). Given that we can only observe comment similarities between organizations that commented on the same docket, we also condition on the fact that both co-commented on docket k in year t ($CC_{fgkt} = 1$). Thus, our estimand can be written as:

$$\delta = E[s^1_{fgkt}|D_{fgt} = 1, CC_{fgkt} = 1] - E[s^0_{fgkt}|D_{fgt} = 1, CC_{fgkt} = 1]$$

where s^1_{fgkt} and s^0_{fgkt} are the treated and untreated potential outcomes respectively. We substitute each counterfactual untreated observation with one or more untreated "control" observations involving the same firm-grantee pair commenting in a different year when there is no donation. When identifying untreated control observations, we also exclude all observations involving the same grantee and docket as a treated observation (if grantees change their commenting behavior in response to recent donations, comparisons with other firm comments in the same docket will also be distorted). Out of an abundance of caution, we also exclude cases in which a firm's donation data is missing, but they give to the grantee in some other year. There are 2,647 treated comment similarity observations in our data, and 183,070 potential control observations. However, it is only possible to find valid controls within the same firm-grantee pair for 648 treated observations.

We construct a set of docket features from data about Federal Register documents linked to that docket. The Federal Register website provides several standardized fields for each published document, including the title, a short abstract, topic keywords, the sections of the Code of Federal Regulations that are affected by the rule-making, and a Regulation Identification Number (RIN) when available. We extract keywords from the titles and abstracts, and pool these together with the other identifiers for all documents linked to a docket. To identify similar dockets, we take the weighted Jaccard similarity index of their features, where the weight for feature f is given as

$$w_f = \frac{1}{log(1 + n_f)}$$

and n_f is the number of dockets with feature f.

The docket similarity measure allows us to identify the control observations for each treated observation that involve the most closely related topics and domain of rule-making. In our primary specification we keep only control observations that have the maximum docket similarity for each treated observation (ties are included, but given fractional weight). For the sake of comparison, we also compute estimates using the *worst* docket matches, as well as estimates using all available controls. We estimate standard errors using the non-parametric approach developed in Hanson and Sunderam (2012) to cluster at the firm-grantee pair level.

Table E.1 presents these results for both our basic LSA cosine similarity measure and our similarity measure that is adjusted to remove the effects of author's average commenting style. We find that the estimated δ is positive, and of a similar order of magnitude to our fixed effects regressions in the main text of the article. Again, it appears that firm-grantee co-comments are more similar after a recent donation. However, the matching estimator results reveal that this difference is primarily driven by comparisons within the most similar dockets. Specifically, $\hat{\delta}$ is relatively large and statistically significant when matching on the most similar dockets, but close to zero and not statistically significant when matching on the worst dockets (with the estimate for all controls falling in between). This evidence strengthens the case for a causal interpretation of the increase in similarity. It is still possible that firms selectively donate to grantees based on upcoming regulatory topics, but these results would only be consistent with that selection story if it operates at a very fine level of regulatory topics that are too subtle to detect with our sample.

Table E.1: Matching Estimator Co-Comment Similarity Estimates

	Control observations used for each treated observation (within firm-grantee pairs):					
	Most Similar Docket Only		Least Similar Docket Only		All Valid Controls	
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\delta}$	0.115*	0.170**	0.025	0.019	0.057	0.087
	(0.061)	(0.077)	(0.099)	(0.115)	(0.105)	(0.136)
Commenter style control		Y		Y		Y
$N_{Treated}$	648	648	648	648	648	648
$N_{Control}$	342	342	321	321	573	573

Notes: This table presents the results of estimating the difference in textual similarity of comments between a firm and grantee with and without a recent donation. $\hat{\delta}$ is the mean difference between the treated outcomes and the mean outcome for each treated outcome's matched control observations. Each observation is a measure of comment similarity between comments submitted by firm f and grantee g on docket k in year t. Observations are considered treated if the firm gave a donation to the grantee in year t or $t-1$ and untreated otherwise. However, we exclude all control units that i) include the same grantee and regulation as a treated observation, ii) include a firm-grantee pair for which the firm is missing donation data but has gives to the grantee at least once in some other year. Control observations are always matched exactly by firm and grantee. In columns 1 and 2, only the control observations with the highest docket similarity are included. In columns 3 and 4 only the control observations with the lowest docket similarity are included. In columns 5 and 6 all valid control observations are included, and given equal weight. $N_{Treated}$ and $N_{Control}$ report the number of unique treated and control observations that are successfully matched. Odd numbered columns show the estimates for simple cosine distances between comment document vectors, while even numbered columns show the estimates for comment vectors that have been adjusted to subtract each organizations average comment. Confidence intervals and statistical significance are computed using the non-parametric variance estimation approach for matching estimators developed in Hanson & Sunderam (2011), clustering by firm-grantee pair. * = p<0.1, **=p<0.05, ***=p<0.01.

F Appendix: Additional sentiment analysis

This appendix presents the results of three additional co-comment sentiment regressions complementing the analysis in Table 7 by using different dictionaries and approaches, including a measure of partisanship based on Gentzkow et al. (2016).

The first two tables replicate our main sentiment analysis results using alternative sentiment measures. Table F.1 is a simple robustness check using the original Harvard Psychosociological Dictionary word list that served as a starting point in Loughran and McDonald (2011). Here we compute the sentiment scores in the same way as our main analysis, using the Harvard-IV-4 TagNeg word list instead of the Fin-Neg word list. Table F.2 shows regression results computed using the popular AFINN sentiment lexicon, an alternative sentiment scoring dictionary that provides valence scores ranging between -5 (negative) and 5 (positive) for each labeled word. We construct each comment sentiment value as the mean sentiment score for words in the comment text. Though not shown here, we also experimented with the simple proportional scoring method used in Loughran and McDonald (2011) and TF-IDF weighted versions of the AFINN sentiment measure. Overall, it appears our original result stands using a variety of sentiment measures: There are no substantial changes in the relative sentiment of comments submitted by firms and grantees when there has been a recent donation.

We also experimented with measuring commenter partisanship using wordlists provided by Gentzkow et al. (2016) based on the 1000 most partisan phrases spoken by members of congress. We pre-process our comment text in the same way as they did for congressional speech, removing stop-words, stemming words, and constructing bi-grams. Then we match the comment bi-grams to the GST phrase lists and compute the mean partisanship score for bigrams in the comment. Not many phrases in the GST list can be found in our comments, so this measure is very noisy. Finally, we compute the differences between partisanship scores as our measure of co-comment similarity and normalize this measure by standard deviation in the sample, as in our other regressions.

Table F.3 presents these results on partisanship and suggest that changes in partisan alignment relating to a recent donation are not readily detectable - at least not as measured by the strong linguistic patterns that occur in congressional speech.

Table F.1: Comment sentiment alignment - Recent donation, Harvard-IV-4 TagNeg wordlist

Dependent variable	Sentiment alignment of comments by grantee g and firm f on same rule-year					
	(1)	(2)	(3)	(4)	(5)	(6)
Grantee g received donation	0.012	0.006	-0.003	0.027	0.019	0.010
from firm f at t or $t-1$	(0.022)	(0.033)	(0.015)	(0.025)	(0.033)	(0.018)
Fixed Effects						
Rule	Y	Y	Y	Y	Y	Y
Firm	Y			Y		
Grantee	Y			Y		
Firm-Grantee Pair		Y			Y	
Agency×NAICS×NTEEC			Y			Y
Commenter style control				Y	Y	Y
Observations	168,341	71,189	81,851	168,341	71,189	81,851

Notes: The dependent variable is the negative absolute difference between the sentiment score assigned to the comment of firm f and the comment of grantee g in the same rule-year rt, using TF-IDF weighted Harvard Psychosociological Dictionary Harvard-IV-4 TagNeg word list (as provided by Loughran & McDonald, 2011), with the dependent variable re-scaled to have a standard deviation of one. The independent variable is equal to one if grantee g received a donation from firm f in the year when the comment appears or the year before. Standard errors use two-way clustering by rule and firm-grantee pair. *** p<0.01, ** p<0.05, * p<0.1

Table F.2: Comment sentiment alignment - Recent donation, AFINN sentiment lexicon

Dependent variable	Sentiment alignment of comments by grantee g and firm f on same rule-year					
	(1)	(2)	(3)	(4)	(5)	(6)
Grantee g received donation	0.052*	0.016	0.009	0.053	0.000	0.008
from firm f at t or $t-1$	(0.029)	(0.040)	(0.023)	(0.033)	(0.043)	(0.024)
Fixed Effects						
Rule	Y	Y	Y	Y	Y	Y
Firm	Y			Y		
Grantee	Y			Y		
Firm-Grantee Pair		Y			Y	
Agency×NAICS×NTEEC			Y			Y
Commenter style control				Y	Y	Y
Observations	168,341	71,189	81,851	168,341	71,189	81,851

Notes: The dependent variable is the negative absolute difference between the sentiment score assigned to the comment of firm f and the comment of grantee g in the same rule-year rt, AFINN sentiment lexicon, with the dependent variable re-scaled to have a standard deviation of one. The independent variable is equal to one if grantee g received a donation from firm f in the year when the comment appears or the year before. Standard errors use two-way clustering by rule and firm-grantee pair. *** p<0.01, ** p<0.05, * p<0.1

Table F.3: Comment partisan alignment - Recent donation

Dependent variable	Partisan alignment of comments by grantee g and firm f on same rule-year					
	(1)	(2)	(3)	(4)	(5)	(6)
Grantee g received donation from firm f at t or $t-1$	-0.014 (0.018)	-0.071 (0.058)	-0.001 (0.009)	-0.019 (0.021)	-0.090 (0.066)	-0.009 (0.010)
Fixed Effects						
Rule	Y	Y	Y	Y	Y	Y
Firm	Y			Y		
Grantee	Y			Y		
Firm-Grantee Pair		Y			Y	
Agency×NAICS×NTEEC			Y			Y
Commenter style control				Y	Y	Y
Observations	168,341	71,189	81,851	168,341	71,189	81,851

Notes: The dependent variable is the negative absolute difference between the party score assigned to the comment of firm f and the comment of grantee g in the same rule-year rt. Each comment party score is created by processing the comment text to extract bigram phrases as in Gentzkow, Shapiro & Taddy (2019), and them matching these phrases to the list of top 1000 most partisan phrases that they provide for each session of congress. Alignment is calculated as the negative absolute difference between comment scores, re-scaled to have a standard deviation of one. The independent variable is equal to one if grantee g received a donation from firm f in the year when the comment appears or the year before. Standard errors use two-way clustering by rule and firm-grantee pair. *** p<0.01, ** p<0.05, * p<0.1

G Appendix: Heterogeneity Analysis

G.1 Analysis of “high stakes” rules

If the regularity that we have uncovered in the relationship between donations and grantees’ comments is a strategic use of charitable giving by corporations, we conjecture that may be particularly prominent in the comments of highly debated rules. More specifically, we suggest that there are rules that attract more attention from various groups in society, with potentially opposing views. If firms make strategic use of their donations, then the mechanism we suggest should be particularly pronounced in the commenting patterns for those rules. We conduct a simple heterogeneity analysis by comparing the magnitude of the effects in our rule specification in Table 5 in these “high stakes” rules. We attempt to capture the importance and potential heterogeneity of views on a given rule by counting the total number of comments submitted by firms and grantees,[67] although we probe that the results are robust to using other criteria.[68] In table G.1, columns (1)-(4) report coefficients for specification (2) for the set of rules that received more than the median total number of comments. Columns (5)-(8) report results for the remaining rules. Across all specifications, the coefficient of interest, which captures the association between the probability of commenting for a grantee and the presence of a donation from a firm that also commented on the same rule, is substantially higher for “high stakes” rules. Focusing on the specifications in columns (4) and (8) the coefficient in the former is 0.176 compared with 0.004 in the latter (44 times larger). Even adjusting for the lower baseline probability of commenting (0.096 vs 0.007) the coefficient in (4) is approximately three times larger than the corresponding value in column (8).[69]

Finally, in table G.4 we show that the probability that grantee g receives a grant is higher when rules are “high stakes.” In all columns the dependent variable is equal to one if grantee g received a donation from a firm commenting on rule r. The independent variables in columns (1)-(3) are the same dummy variables that are used in the heterogeneity analysis in Tables G.1,

[67]We exclude from the count of grantee comments those submitted by grantees that received a donation from a firm that commented on the same rule.

[68]In Table G.2 we define as “high stakes” rules those with more than the median number of grantee comments. In Table G.3 we define a measure of the balanced presence of both grantee and firm comments by calculating the absolute value of the difference between the grantee share of comments on a rule and the mean grantee share of comments across all rules. Then we consider as “high stakes” rules that have a below-median value of this measure (a proxy for a relatively balanced number of firms and grantee comments) and an above-median total number of comments. It is natural also to think about a measure of *political* contentiousness based on whether a rule faced a close congressional vote. However, only about a third of rules are initiated directly by Congress (West and Raso, 2013).

[69]Normalizing by the baseline commenting probability, the ratio of the two coefficients in (4) and (8) is $\frac{0.176/0.096}{0.004/0.007} = 3.2025$.

G.2 and G.3, respectively. All specifications include grantee fixed effects. Column (4) employs three continuous variables that unpack the definition of "high stakes": the number of comments, the share of grantee comments and its square. We find a a positive effect of the total number of comments and a hump-shaped relationship with the share of grantee comments. The probability of a grantee receiving a grant is highest at a share of grantee comments of 38%.[70] This result shows that grantees that comment on rules in which there are more split comments are more likely to receive donations compared to rules in which either only grantees or only firms comment. We conclude that our paper's main findings are primarily driven by relatively more controversial rules that draw more attention and those for which firms can plausibly benefit more from amplifying their messages.

[70]This is calculated as $0.098/(0.129 \times 2) \approx 0.38$

Table G.1: Commenting on high-comment rules

Dependent variable	Grantee g commented on rule $r \times 100$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Rules w/ *above* median # of comments				Rules w/ *below* median # of comments			
Mean	0.096				0.007			
Grantee g received donation	0.306***	0.258***	0.237***	0.176***	0.003	0.008***	-0.002	0.004*
from any firm commenting on r	(0.018)	(0.019)	(0.014)	(0.022)	(0.002)	(0.001)	(0.002)	(0.002)
Fixed effects								
Grantee		Y		Y		Y		Y
Regulation			Y	Y			Y	Y
Observations	47,814,692	47,814,692	47,814,692	47,814,692	69,730,676	69,730,676	69,730,676	69,730,676

Notes: The dependent variable is equal to 100 if grantee g comments on rule r. The independent variable is equal to one if grantee g received in any year 2003-2016 a donation from a firm that commented on r. Standard errors are two-way clustered at the rule and grantee level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table G.2: Commenting on high-grantee-comment rules

Dependent variable	Grantee g commented on rule $r \times 100$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Rules w/ *above* median # of grantee comments				Rules w/ *below* median # of grantee comments			
Mean	0.094				0.005			
Grantee g received donation from any firm commenting on r	0.433***	0.342***	0.370***	0.264***	0.015***	0.017***	0.013***	0.015***
	(0.024)	(0.026)	(0.020)	(0.030)	(0.002)	(0.002)	(0.002)	(0.003)
Fixed effects								
Grantee		Y		Y		Y		Y
Regulation			Y	Y			Y	Y
Observations	50,169,120	50,169,120	50,169,120	50,169,120	67,376,248	67,376,248	67,376,248	67,376,248

Notes: The dependent variable is equal to 100 if grantee g comments on rule r. The independent variable is equal to one if grantee g received in any year 2003-2016 a donation from a firm that commented on r. Standard errors are two-way clustered at the rule and grantee level. *** p<0.01, ** p<0.05, * p<0.1

Table G.3: Commenting on “high stakes” rules

Dependent variable	Grantee g commented on rule $r \times 100$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Rules w/ *above* median # of comments and *balanced*				Other rules			
Mean	0.118				0.007			
Grantee g received donation from any firm commenting on r	0.346***	0.295***	0.273***	0.205***	-0.009***	0.012***	-0.034***	-0.013***
	(0.020)	(0.022)	(0.016)	(0.024)	(0.002)	(0.001)	(0.004)	(0.004)
Fixed effects								
Grantee		Y		Y		Y		Y
Regulation			Y	Y			Y	Y
Observations	25,370,612	25,370,612	25,370,612	25,370,612	92,174,756	92,174,756	92,174,756	92,174,756

Notes: The dependent variable is equal to 100 if grantee g comments on rule r. The independent variable is equal to one if grantee g received in any year 2003-2016 a donation from a firm that commented on r. A rule is defined as *balanced* if it has a value of the variable |%grantee comments−average%grantee comments| below its median. Standard errors are two-way clustered at the rule and grantee level. *** p<0.01, ** p<0.05, * p<0.1

Table G.4: Grant probability for "high stakes" rules

Dependent variable	Grantee g received donation from any firm commenting on r			
	(1)	(2)	(3)	(4)
Rule r has *above* median # of comments	0.021***			
	(0.001)			
Rule r has *above* median # of grantee comments		0.006***		
		(0.001)		
Rule r has *above* median # of comments and is *balanced*			0.040***	
			(0.002)	
Total number of comments on rule r				0.001***
				(0.000)
Share of grantee comments on rule r				0.098***
				(0.007)
Square of share of grantee comments on rule r				-0.129***
				(0.006)
Observations	117,545,368	117,545,368	117,545,368	117,545,368

Notes: The dependent variable is equal to one if grantee g received in any year 2003-2016 a donation from a firm that commented on r. A rule is defined as *balanced* if it has a value of the variable $|$%grantee comments$-$average%grantee comments$|$ below its median. All columns include grantee fixed effects. Standard errors are two-way clustered at the rule and grantee level. *** p<0.01, ** p<0.05, * p<0.1

G.2 Hush money

Sections 3-5 focus on the role of donations from corporations to non-profits in generating additional messages that are more similar to the donor's position. This section examines whether corporations also use donations for a distinct strategic purpose: to silence opposing opinions.

It is plausible to envision an informational lobbying environment in which agents supporting a specific action opposed by a counterpart may be motivated to suppress the opposing viewpoint (and compensate the counterpart for its silence). For example, in a discussion of the strategies employed in the multi-year campaign of the tobacco industry against greater regulation, Lando (1991) writes: "*The tobacco industry has been effective in purchasing what has been described as 'innocence by association'. Tobacco industry sponsorship of sports events is notorious. The industry has also contributed substantially to the arts, to women's groups, and to organizations representing minorities. These types of pernicious industry activities have been successful in buying the silence or the tacit support of some groups that have suffered a disproportionate share of the tobacco burden.*" Payment in exchange for inaction and silence is commonplace in the market (e.g., noncompete and nondisclosure agreements, non-disparagement clauses, etc.) and such private agreements or clauses do not represent *per se* invalid contracts or violations of free speech. They may be, however, private agreements that are undisclosed to regulators, who may interpret the silence of some parties to the regulatory process as informative.[71]

The role of such "negative" strategies is thought to be crucial to the success of special interest groups in politics. Blocking unfavorable bills from ever seeing the light of day (committee discharge) in the U.S. Congress is as much a part of lobbying as facilitating the passage of bills favorable to an industry. Similarly, interest group comments in rule-making often aim to kill unfavorable provisions or stall the implementation of rules. ("Nothing happening" is almost always the desirable policy outcome for incumbent firms; see Baumgartner et al., 2009.)

To test for the presence of "hush money" in rule-making, we propose an extension of our empirical framework in section 3. In particular, we modify the rule specification in section 3.2 as follows:

$$C_{gr} = \beta_0 + \beta_1 DonorComment_{gr} + \beta_2 DonorComment_{gr} \times Comments_{ga} + \delta_g + \delta_r + \eta_{gr} \qquad (5)$$

where $DonorComment_{gr}$ is equal to 1 if grantee g received a donation from a firm that also commented on the same regulation, and 0 otherwise, and $Comments_{ga}$ is a measure of how frequently g comments to regulatory agency a. We consider three different measures for $Comments_{ga}$: the total number of comments submitted by g to a, the share of g's comments that are submitted to

[71] Absence of a signal is in fact informative in games of incomplete information in which Bayesian agents are assumed. For an application to political campaigns, see Kendall et al. (2015).

a, and the share of all comments submitted to a that come from g.

To understand the intuition behind this test, observe that certain non-profits may have particular expertise or focus in a specific area of regulation, which we approximate by the identity of the agency overseeing the rule (e.g., the Sierra Club commenting on rules proposed by the EPA).[72] Interacting $Comments_{ga}$ with the donation from a commenting firm, $DonorComment_{gr}$, aims to establish whether such donations have a differential effect on the likelihood of commenting for grantees that typically comment on rules considered by agency a, versus grantees that normally do not comment on rules by a. We argue that this interaction is useful for assessing the potential role of hush money, since within the set of issue experts (high $ShareComments_{ga}$) it is more likely that donations are made with the aim of inducing silence and muting expert commentary. A plausible null hypothesis supporting the presence of hush money is therefore $\beta_2 < 0$, as charitable donations may be more likely to be hush money for grantees that routinely comment on rules from a.

Our results based on this specification suggest that hush money is not a common strategy in our setting. In Table G.5 we present results using all three measures of $Comments_{ga}$ with and without rule fixed effects. The evidence points clearly in the direction of donations increasing co-commenting from grantees that routinely comment on rules from the regulator proposing r. The coefficient $\beta_2 > 0$ is systematically positive and highly statistically significant, indicating that firms are more likely to induce – rather than stifle – comments from such grantees. While this does not rule out the existence of hush money, it nevertheless suggests that this behavior might be less prevalent than the co-commenting behavior documented in sections 3-5.

[72]Bertrand et al. (2014) follow a similar approach to define issue expertise of individual lobbyists from federal lobbying reports.

Table G.5: Hush money

Dependent variable	Grantee g commented on rule $r \times 100$					
Mean	0.043					
	(1)	(2)	(3)	(4)	(5)	(6)
$DonorComment_{gr}$	0.101***	0.072***	0.100***	0.030	0.006	0.031*
	(0.017)	(0.007)	(0.016)	(0.019)	(0.011)	(0.018)
$DonorComment_{gr}$	0.183***			0.183***		
$\times NumberComments_{ga}$	(0.033)			(0.035)		
$DonorComment_{gr}$		2.789***			2.747***	
$\times$Share g comments to a		(0.167)			(0.268)	
$100 * DonorComment_{gr}$			5.617***			5.618***
$\times$Share a comments from g			(0.891)			(0.924)
Fixed effects						
Grantee	Y	Y	Y	Y	Y	Y
Rule				Y	Y	Y
Observations	117,545,368	117,545,368	117,545,368	117,545,368	117,545,368	117,545,368

Notes: The dependent variable is equal to 100 if grantee r comments on rule r. The $DonorComment_{gr}$ is equal to one if grantee g received in any year 2003-2016 a donation from firm f that also commented on rule r. a indicates the agency receiving the comments. Standard errors use two-way clustering by grantee and rule. *** p<0.01, ** p<0.05, * p<0.1

G.3 Industry concentration

This subsection illustrates heterogeneity of our results along industry concentration measures. The main finding in this analysis is that the main co-commenting regularity uncovered in the paper is evident in highly concentrated sectors and less so in sectors at low levels of revenue concentration.

Tables G.6 and G.7 report our main findings splitting by either above/below median CR4 NAICS 3 industry or above/below median CR8 NAICS 3 industry, both standard measures of industry-level revenue share concentration in the literature. Using either definition, we observe that our paper's main results are driven by the subsample where collective action and lobbying efforts are intuitively more effectively organized per standard Olson (1965) rationale. More specifically, only high concentration industries present a clear pattern linking co-comments to recent charitable grants when using our more exhaustive sets of fixed effects (columns (3) and (4) in each table versus columns (7) and (8)).

This heterogeneity result is useful in highlighting a connection between the phenomenon uncovered in the paper and more traditional forces at play in standard special interest politics analysis (Grossman and Helpman, 2001) and in this sense it produces additional evidence in corroboration of the discussion presented in the Introduction.

Table G.6: Heterogeneity in the grant-co-comment relationship by industry concentration level

Dependent variable	Firm f and grantee g commented on the same rule in year $t \times 100$							
	Above median CR4 NAICS 3 industry				Below median CR4 NAICS 3 industry			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grantee g received donation from firm f at t or $t-1$	0.995***	0.679***	0.299***	0.206***	1.3477***	0.8484***	0.0256	-0.0258
	(0.052)	(0.049)	(0.051)	(0.048)	(0.0786)	(0.0723)	(0.0831)	(0.0779)
Fixed effects								
Year	Y	Y	Y		Y	Y	Y	
Grantee		Y				Y		
Firm		Y				Y		
Grantee-Firm Pair			Y	Y			Y	Y
Grantee-Year				Y				Y
Firm-Year				Y				Y
Observations	26,412,156	26,412,156	26,383,968	26,383,968	33,726,942	33,726,942	33,719,895	33,719,895
R-Squared	0.004	0.018	0.116	0.195	0.0045	0.0198	0.1258	0.2110

Notes: The dependent variable is equal to 100 if grantee g and firm f comment on the same rule in year t. The independent variable is equal to one if grantee g received a donation from firm f at year t or $t-1$. CR4 is the share of the largest four firms' revenues in a NAICS3 industry. Standard errors are clustered at the firm-grantee pair level in all columns. *** p<0.01, ** p<0.05, * p<0.1

Table G.7: Heterogeneity in the grant-co-comment relationship by industry concentration level

Dependent variable	Firm f and grantee g commented on the same rule in year t×100							
	Above median CR8 NAICS 3 industry				Below median CR8 NAICS 3 industry			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grantee g received donation from firm f at t or $t-1$	0.965***	0.658***	0.285***	0.192***	1.3781***	0.8740***	0.0580	0.0011
	(0.052)	(0.049)	(0.051)	(0.048)	(0.0770)	(0.0708)	(0.0816)	(0.0767)
Fixed effects								
Year	Y	Y	Y		Y	Y	Y	
Grantee		Y				Y		
Firm		Y				Y		
Grantee-Firm Pair			Y	Y			Y	Y
Grantee-Year				Y				Y
Firm-Year				Y				Y
Observations	25,045,038	25,045,038	25,016,850	25,016,850	35,094,060	35,094,060	35,087,013	35,087,013
R-Squared	0.005	0.020	0.118	0.216	0.0040	0.0181	0.1236	0.2000

Notes: The dependent variable is equal to 100 if grantee g and firm f comment on the same rule in year t. The independent variable is equal to one if grantee g received a donation from firm f at year t or $t-1$. CR8 is the share of the largest eight firms' revenues in a NAICS3 industry. Standard errors are clustered at the firm-grantee pair level in all columns. *** p<0.01, ** p<0.05, * p<0.1

G.4 Republican-Democratic agency split

Given the importance of partisanship and polarization in current U.S. politics, we consider whether the relationships between commenting and donations we document are different under Republican versus Democratic regulators.

In general, agencies are led by political appointees selected by the President of the United States and confirmed by the Senate. Institutional details vary by agency: leaders of cabinet departments serve at the pleasure of the President, while the leaders of independent agencies often serve fixed terms. However, even in independent agencies such as the FCC, which are nominally led by a bipartisan commission, the President selects the chair of the commission and appoints the most recent member of the commission, giving majority control to the President's party. Altogether, the structure of the executive branch is such that the party affiliation of agency leaders aligns with that of the current president.

While the majority of the comments in our data were submitted in 2008 or later, we have some comment data as early as 2003. This means that we can compare our results before and after President Obama took office in 2009 to explore how comments relate to donations under regulators of different parties. Of course, this simple analysis must be interpreted carefully: we cannot distinguish between the effects of party-based differences and other simultaneous time trends, and our data is much more sparse in the 2003-2008 period, when comment data is unavailable on regulations.gov for some agencies.

With these caveats in mind, tables G.8, G.9, and G.10 present replications of our main results on the relationship between recent donations and co-commenting, co-comment similarity, and rule outcomes when we include an additional interaction term for recent donations when the president was Republican (an indicator variable). We find that the interaction term is consistently negative, suggesting that the relationship between recent donations and each of our outcomes is weaker under Republican regulators. The interaction term is significant at standard statistical levels in the co-comment regressions, where we have the largest sample size. Reassuringly, in each regression, the magnitude of the estimated interaction term increases as we add sets of fixed effects.

We hypothesize that these results may reflect an intuitive pattern: Republican regulators could be more "pro-business" and more politically aligned with firms, while Democratic regulators could be less aligned with, and less trustful of, communications that come directly from business special interests. This would mean that firms have less incentive to coordinate with third parties to obfuscate the source of their advocacy when dealing with Republican regulators, while these strategies could be more beneficial when dealing with Democratic regulators. This asymmetry begins to answer a question posed by (among others) Yackee and Yackee (2006), on what advocacy strategies by firms may be more appropriate under regulators with heterogeneous beliefs. As more information on the comments filed after 2017 under President Trump becomes available, this asymmetry may be further confirmed.

Table G.8: Co-commenting - Recent donation, President Party interactions

Dependent variable	Firm f and grantee g commented on the same rule in year $t(\times 100)$			
Mean	0.175			
	(1)	(2)	(3)	(4)
Firm f contributed to grantee g in year t or $t-1$	1.833***	1.413***	0.933***	0.579***
	(0.063)	(0.060)	(0.055)	(0.053)
Firm f contributed to grantee g in year t or $t-1$ and President is Republican	-1.451***	-1.494***	-1.758***	-1.092***
	(0.064)	(0.064)	(0.071)	(0.066)
Fixed effects				
Year	Y	Y	Y	
Grantee		Y		
Firm		Y		
Grantee-Firm Pair			Y	Y
Grantee-Year				Y
Firm-Year				Y
Observations	122,287,230	122,287,230	122,232,220	122,232,220

Notes: The dependent variable is equal to 100 if grantee g and firm f comment on the same rule r in year t. The independent variable is equal to one if grantee g received a donation from firm f at year t or $t-1$. President party is determined by the year comments were posted (i.e., for years 2001-2008 the president was Republican, and for 2009-2016 the president was a Democrat). Standard errors are clustered at the grantee-firm pair level. *** p<0.01, ** p<0.05, * p<0.1

Table G.9: Similarity of comments - Recent donation, President Party Interactions

Dependent variable	Similarity of comments by grantee g and firm f on same rule					
	(1)	(2)	(3)	(4)	(5)	(6)
Grantee g received donation from firm f at t or $t-1$	0.049***	0.066*	0.038*	0.061***	0.072*	0.048**
	(0.017)	(0.038)	(0.021)	(0.017)	(0.041)	(0.024)
Grantee g received donation from firm f at t or $t-1$ and President is Republican	-0.014	-0.030	-0.048	-0.032	-0.048	-0.077
	(0.077)	(0.078)	(0.073)	(0.075)	(0.081)	(0.075)
Fixed Effects						
Rule	Y	Y	Y	Y	Y	Y
Firm	Y			Y		
Grantee	Y			Y		
Firm-Grantee Pair		Y			Y	
Agency×NAICS×NTEEC			Y			Y
Comment style control				Y	Y	Y
Observations	168,347	71,195	81,851	168,347	71,195	81,851

Notes: The dependent variable is a similarity index between the comment of firm f and the comment of grantee g in the same rule r, scaled to have a standard deviation of one. The independent variable is equal to one if grantee g received a donation from firm f in the year when the comment appears or the year before. Standard errors use two-way clustering by rule and firm-grantee pair. President party is determined by the year comments were posted (i.e., for years 2001-2008 the president was Republican, and for 2009-2016 the president was a Democrat). *** p<0.01, ** p<0.05, * p<0.1

Table G.10: Rule outcomes - Recent donation, Presidental Party Interaction

Dependent variable	Similarity between comment submitted by firm f and discussion text in rule r				Log citation count		Any citation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
At least one grantee g co-commenting and receiving donation from firm f in year t or $t-1$	0.173***	0.173***	0.133**	0.131**	0.344**	0.343**	0.141**	0.145**
	(0.055)	(0.055)	(0.055)	(0.055)	(0.159)	(0.159)	(0.069)	(0.068)
At least one grantee g co-commenting and receiving donation from firm f in year t or $t-1$ *and president is Republican*	-0.115	-0.116	-0.147	-0.150	-0.187	-0.187	-0.052	-0.052
	(0.105)	(0.105)	(0.093)	(0.093)	(0.215)	(0.214)	(0.096)	(0.093)
Log expenditure lobbying agency in t and $t-1$		0.002		0.007**		0.002		-0.005
		(0.004)		(0.004)		(0.007)		(0.003)
Fixed Effects								
Rule	Y	Y	Y	Y	Y	Y	Y	Y
Firm	Y	Y	Y	Y	Y	Y	Y	Y
Commenter Style Control			Y	Y	*	*	*	*
Observations	4,675	4,675	4,665	4,665	1,212	1,212	1,212	1,212

Notes: The dependent variables are several measures of the relationship between firm comments and the discussion of comments in subsequent rules. For columns 1-4 the outcome is the overall similarity of the text, for columns 5 and 6, the outcome is log of the number of detected occurrences of the firm's name in the discussion text, for columns 7 and 8 the outcome is an indicator for the presence of at least one occurrence of the firm's name in the discussion text. Some agencies rarely cite any commenters by name, so in columns 5-8 we restrict the sample to agencies whose rules contain an average of at least one citation of a firm per rule. The independent variable is equal to one if there is at least one grantee g co-commenting on regulation r and receiving a grant from firm f in year t or $t-1$. President party is determined by the year comments were posted (i.e., for years 2001-2008 the president was Republican, and for 2009-2016 the president was a Democrat). Standard errors use two-way clustering by rule and firm. *** p<0.01, ** p<0.05, * p<0.1

G.5 Other grantee attributes

We conclude by considering two other main dimensions of heterogeneity that may be informative as to the types of non-profits that may be most susceptible to co-opting by corporate funders. Specifically, we consider research-focused organizations, and organizations focused on shaping policy, both by influencing public opinion and directly lobbying governments on legislation. In both cases, differences in the effect of money on co-commenting behavior is ambiguous. Consider first research-focused organizations. On the one hand, such entities ostensibly provide neutral expert input on regulations that lie within their purview; on the other hand, research organizations such as think tanks may be targeted with donations from firms which exploit preexisting sympathies to nudge them toward providing supportive comments.[73] Advocacy organizations share a similar ambiguity, as a result of forceful prior policy positions which, on the one hand, should make them less persuadable, but may also lead to donations that aim to nudge them toward supportive commentary.

We define research- and advocacy-focus based on the IRS's National Taxonomy of Exempt Entities (NTEE) code, a three-digit activity classification system for non-profits. The first digit, a letter, denotes the organization's main area (e.g., arts, medical, environment, etc), whereas the second two are numerical digits which capture the type of activity. For example, A denotes all arts organizations, while A50 is the category for museums. We define *Research* as an indicator variable that takes on a value of 1 for each of the following groups: all non-profits in the main areas of medical (H), science (U), and social science (V); non-profits across all main sectors with the activity code for Research Institutes & Public Policy Analysis (05); and institutions of higher education with a research focus (B43 and B50, universities offering graduate programs, and graduate/professional schools respectively). We will further distinguish between comments from higher education organizations (B43 and B50, in which case the commenter is usually a faculty member) and all other research-focused entities. The indicator variable *Advocacy* captures all non-profits with the activity code for Alliances and Advocacy (01)[74] as well as all non-profits in the main area of Civil Rights, Social Action & Advocacy (R).

In Table G.11, we present results that parallel those of Table 4. We add to this specification a series of interaction terms to explore heterogeneous effects of grants on co-commenting. Standard errors are clustered at the grantee-firm pair level for all columns. First, in column (1) we present results with only year fixed effects, to examine differences in the average level of co-commenting

[73] See, for example, Eric Lipton and Brooke Williams, "How Think Tanks Amplify Corporate America's Influence," *The New York Times*, August 7, 2016.

[74] The definition of this category is as follows – for the education category it reads, "Organizations whose activities focus on influencing public policy within the Education major group area. Includes a variety of activities from public education and influencing public opinion to lobbying national and state legislatures." The definition is similar for other major area (first-digit) groups.

for research and advocacy organizations. We also include as a control $log(Income)$ of the grantee, to control for size. As expected, advocacy organizations – which, recall, are defined by a mission of affecting policy – are far more likely to co-comment on regulations, a direct result of their frequent commenting more generally. Similarly (though of a much smaller magnitude) research-focused organizations are more likely to co-comment. Co-commenting is also correlated with size, as expected.

Column (2) examines whether there is differential co-commenting behavior for *Research* organizations; this specification includes firm-grantee fixed effects.The interaction term is negative, marginally significant ($p < 0.10$), and large in magnitude – its value, -0.213, is almost identical to that of the direct effect of lagged grants, indicating a zero correlation between the receipt of a grant and co-commenting for research-focused organizations. In column (3) we disaggregate *Research* into universities versus all others and, while neither coefficient is statistically significant, we find that the two are both negative (though the university research interaction is more negative). Finally, column (4) looks at differential behavioral for *Advocacy* organizations. The interaction term is negative and, while not significant, very large in magnitude, more than double the size of the direct effect of grant receipt. Recall that overall advocacy organizations are relatively frequent commenters; one possible interpretation of this result is a "hush money" effect, with firms paying advocacy firms to stifle would-be comments. We explore this issue further in section 6.

Table G.11: Heterogeneity in the grant-co-comment relationship by non-profit type

Dependent variable	Firm f and grantee g commented on the same rule in year $t\times100$			
	(1)	(2)	(3)	(4)
Grantee g received donation from firm f at t or $t-1$	0.966*** (0.043)	0.220*** (0.045)	0.220*** (0.045)	0.190*** (0.042)
log(Income)	0.026*** (0.000)			
Research	0.007*** (0.002)			
Advocacy	0.142*** (0.005)			
Grantee g received donation $\times$Research		-0.213* (0.118)		
Grantee g received donation $\times$Research University			-0.226* (0.125)	
Grantee g received donation $\times$Research non Uni.			-0.136 (0.304)	
Grantee g received donation $\times$Advocacy				-0.418 (0.288)
Fixed Effects				
Year	Y	Y	Y	Y
Firm-Grantee		Y	Y	Y
Observations	65,733,360	75,163,302	75,163,302	75,163,302
R-Squared	0.004	0.131	0.131	0.131

Notes: The dependent variable is equal to 100 if grantee g and firm f comment on the same rule in year t. The independent variable is equal to one if grantee g received a donation from firm f at year t or $t-1$. The dummy variables Research, Research University, Research non Uni. and Advocacy are set equal to one when the grantee belongs to one of those categories. Standard errors are clustered at the firm-grantee pair level in all columns. *** p<0.01, ** p<0.05, * p<0.1