Do Public Pressure Campaigns Influence Bureaucratic Policymaking?

Devin Judge-Lord

University of Wisconsin-Madison

December 9, 2019

I assess whether public pressure campaigns increase lobbying success in agency rulemaking. To measure lobbying success, I develop computational methods to identify lobbying coalitions and estimate lobbying success for all rules posted for comment on regulations.gov. These methods are validated against a random sample of 100 rules with a mass-comment campaign and 100 rules without no mass comment campaign that I hand-code for whether each coalition got the policy outcome they sought. I then assess potential mechanisms by which mass public engagement may affect policy. Each mechanism involves a distinct type of information revealed to decisionmakers. Of primary interest is the extent to which public pressure campaigns affect agency decisionmakers directly or indirectly through their effects on elected officials' oversight behaviors. I test this by assessing congressional oversight as a causal mediator using a subset of rules where I collect and code correspondence from Member of Congress to agencies about proposed agency rules.

1 DATA

I create an original dataset that combines several sources of data on U.S. federal agency rulemaking. The core data are the texts of draft and final rules and public comments on these proposed rules. This includes all 16 thousand proposed rules from 144 agencies (as defined by regulations.gov) that were open for comment on regulations.gov between 2005 and 2018, that received at least one comment from an organization, and that saw a final agency action between 2005 and 2019. There are over 50 million comments on this set of rules. I scrape draft and final rule texts from federalregister.gov and comments submitted as attachments

or by mail from regulations.gov. I retrieve comments submitted directly on regulations.gov and metadata on rules and comments (such as the dates that the proposed rule was open for comment and whether the agency identified the organization submitting the comment) from the regulations.gov API. I add additional metadata on rules (such as whether the rule was considered "significant") from the Unified Agenda published by the Office of Information and Regulatory Affairs (reginfo.gov).

Finally, to better capture positions expressed by Members of Congress on proposed rules, I supplement congressional comments posted on regulations.gov with Freedom of Information Act Requests for all communication from Members of Congress to each agency on proposed rules from 2007 to 2019.[1]

The combined dataset has over 50 million observations of one public or legislator comment on a proposed rule. I attempt to identify the organization(s) that submitted or mobilized each comment by extracting all organization names from the comment text. For comments that do not reference an organization, I am often able to identify organizations with an internet search using the comment's text. I then identify lobbying coalitions by clustering comments that use similar phrases or word frequencies. Because I attribute comment text to each signatory of a comment, co-sign comments are always assigned to the same coalition. Likewise, form-letter comments are always assigned to the same coalition[2]

Because my hypotheses are about the influence of organizations and coalitions, for analysis, I collapse these data to either one observation per organization or coalition per proposed rule and identify the main substantive comment submitted by each organization's staff or lawyers, which are usually much longer than supporting comments like form letters. For hand-coding, I first select a random sample of 100 proposed rules with a mass-comment campaign and then selecting a matched sample of 100 proposed rules without a mass comment campaign. Matching prioritizes, presidential administration, policy area (following Policy

---

[1]Many agencies provided records of their congressional correspondence going back to 2005 or earlier.

[2]For more on how I identify organizations and coalitions, see Chapter 2, "Why Do Agencies (Sometimes) Get So Much Mail?".

Agendas Project coding), rule significance, department, agency, subagency, and proposed rule length, respectively.[3] This hand-coded sample is several times larger than leading studies using hand-coding and includes rules with very large and small numbers of comments that previous studies exclude. The full sample is four hundred times larger.[4]

2   Methods

The most direct way to assess the hypothesis that mass engagement increases lobbying success is to assess the magnitude of the relationship between the number of comments that a coalition mobilizes and its lobbying success. However, public pressure campaigns may only be effective under certain conditions. Thus, I first assess the main relationship and then assess evidence for or against different potential causal pathways of influence.

2.1   The Dependent Variable: Lobbying Success

The dependent variable is the extent to which a lobbying coalition got the policy outcome they sought, which I measure in three ways.

First, on a sample of rules, I hand-code lobbying success for each lobbying coalition, comparing the change between the draft and final rule to each organization's demands on a five-point scale from "mostly as requested" to "significantly different/opposite than requested." To do this, I first identify organizational comments. For each organization, I identify the main overall demand and the top three specific demands and the corresponding parts of the draft and final rule texts.[5]

I then code overall lobbying success and lobbying success on each specific demand for each organization and coalition. Both the overall score and average score across specific demands

---

[3]For more on policy area coding, see "Trends in U.S. Executive-branch Policymaking 1980-2016" in Chapter 1.

[4]Except for comment texts and attachments, these data are available separately and combined on github.com/judgelord/rulemaking.

[5]This does not capture rule changes on which an organization did not comment. The codebook is available here. See examples of coded cases here.

both fall on the interval from -1 ("significantly different") to 1 ("mostly as requested").

Second, I use methods similar to automated plagiarism detection algorithms to identify changes between a draft and final rule that were suggested in a comment. Specifically, I count the number of words in phrases of at least ten words that appear in the comment and final rule, but not the draft rule. To do this, I first identify new or changed text in the final rule by removing all 10-word or longer phrases retained from the draft rule. I then search each comment for any 10-word or longer phrases shared with the new rule text and count the total number of shared words in these shared phrases. Finally, I normalize this count of "copied" words across shorter and longer comments by dividing it by the total number of words in the comment. This measure falls between 0 (zero percent of words from the comment added to the final rule) and 1 (100 percent of words from the comment added to the final rule). As a robustness check, I also use the non-normalized version of this variable, i.e. the raw number of "copied" words.

Third, I capture a broader dimension of lobbying success by modeling the similarity in word frequency distributions between comments and changes to the rule. New or changed text is identified as described above, except that I also include the rule's preamble and the agency's responses to comments. Agencies write lengthy justifications of their decisions in response to some comments but not others. By including preambles and responses to comments, this measure captures attention to a comment's demands and the extent to which the agency adopts a comment's discursive framing (i.e. the distribution of words it uses). I us cosign similarity to scale the word frequencies used by each comment relative to those in changes between draft and final rule.[6] This measure falls between 0 (no common words) and 1 (exactly the same distribution of words).

---

[6] For the subset of rules with 5 or more organizational comments, I create a more sophisticated measure of word frequency similarity by averaging the absolute value of differences in topic proportions θ between the comment and new rule text across 45 LDA models of all organizational comments estimated with 5 through 50 topics, normalized by the number of topics k_n and the number of models such that y_i falls between 0 (completely different estimated topic proportions) and 1 (exactly the same topic proportions), $y_i = \sum_{5}^{n=50} (\frac{\sum |\theta_{rule\ change_i|k=n} - \theta_{comment_i|k=n}|}{n}) * \frac{1}{45}$. For more on these methods of measuring textual similarity, see "Measuring Change and Influence in Budget Texts".

To assess the performance of these automated methods (text-reuse and word-frequency similarity), I calculate the correlation between these scores and my hand-coded 5-point scale for rules in the hand-coded sample where a final rule was published. As the automated methods apply at the organization-level, coalition scores are those from the lead organization–by default, the organization(s) with the longest comment. At the coaltion level, the correlation between hand-coded influence is __ with the text-reuse method and __ with the word-frequency method.

2.2 The Main Predictor Variable

The number of supportive comments generated by a public pressure campaign (the main variable of interest) is a tally of all comments mobilized by an organization or coalition that ran a mass-comment campaign on a proposed rule. This does not include the main substantive comments submitted by an organization's staff or lawyers. If an organization mobilizes more than 1000 comments or 100 identical comments on a proposed rule, I code that organization, its coalition, and the proposed rule as having a mass comment campaign. Because the marginal impact of additional comments likely diminishes, the number of comments is logged. Where organizational comments are not supported by a mass comment campaign *log mass comments* takes a value of 0.

2.3 Explanatory variables

Other predictors of lobbying success in the models below are the length of the (lead) organization's comment, whether the coalition lobbies unopposed, the size of the lobbying coalition, and whether the coalition is business-led. *Comment length* is normalized by dividing the number of words in the comment by the number of words in the proposed rule, thus capturing the complexity of the comment relative to the complexity of the proposed rule. The number and type(s) of organization(s) is an attribute of each coalition (e.g. a *business-led* coalition with N organizational members). Coalition *size* is the number of distinct commenting

5

organizations in the coalition (including those that co-sign a comment). For organizations lobbying alone, coalition *size* is 1. A coalition is *unopposed* when no opposing organizations comments. I code a coalition as *business-led* if the majority of commenting organizations are for-profit businesses, or if upon investigation, I find it to be primarily led or sponsored by for-profit businesses.[7]

2.4 LIMITATIONS

The two main limitations of this design both bias estimates of public pressure campaign influence toward zero.

First, lobbying success may take forms other than changes in policy texts. Agencies may speed up or delay finalizing a rule, extend the comment period, or delay the date at which the rule goes into effect. Indeed, commentors often request speedy or delayed rule finalization, comment period extensions, or delayed effective dates. I capture these potential outcomes in my hand-coding but not in the two automated methods, which apply only to observations with a final rule text. Likewise, there there is no change between draft and final rule, both automated methods necessarily record lobbying success as 0, even if a comment asks an agency to publish a rule without change.[8]

Second, bureaucrats may anticipate public pressure campaigns when writing draft rules, muting the observed relationship between public pressure and rule change at the final rule stage of the policy process.

2.5 MODELING THE DIRECT RELATIONSHIP

For all three measures of lobbying success, I assess the relationship between lobbying success and mass comments by modeling coalition i's lobbying success, y_i as a combination of the

---

[7]For more on how I identify types of organizations and coalitions, see Chapter 2, "Why Do Agencies (Sometimes) Get So Much Mail?".

[8]Time allowing, I will hand-code a larger sample of proposed rules where no final rule was published so that these cases can be included in the analysis. Likewise, I will identify cases where organizations requested rules to be published as-is and recode these cases by hand.

relative length of the (lead) organizations comment, whether the coalition is unopposed, the coalition's size, whether it is a business coalition, and the logged number of mass comments. I estimate OLS[9] regression:

$$y_i = \beta_0 + \beta_1 log(comments_i) + \beta_2 length_i + \beta_3 unopposed_i + \beta_4 size_i + \beta_5 business_i + \epsilon_i$$

2.6 MODELING MEDIATED RELATIONSHIPS

To estimate mediated effects, I estimate the average conditional marginal effect (ACME) and the proportion of the total effect attributed to mediation through congressional support (comments or other communication from Members of Congress supporting the coalition's position on the proposed rule). As developed by Imai et al. (2010), this involves first estimating a model of the proposed mediator as a combination of covariates, X (*length*, *unopposed*, *size*, and *business*) and then the outcome as a combination of the mediator, *congressional support*, and covariates, X.

Mediator model:

$$congressional\ support_i = \beta_0 + \beta_1 log(comments_i) + \beta_{2-n} X_i + \epsilon_i$$

Outcome model:

$$y_i = \beta_0 + \beta_1 log(comments_i) + \beta_2 congressional\ support_i + \beta_{3-n} X_i + \epsilon_i$$

---

[9]See OLS model estimates with simulated data. I also estimate hand-coded lobbying success with beta regression and ordered logit, which are more appropriate but less interpretable. For the automated measures of lobbying success, I estimate beta regression models with the same variables.