The Organization of Federal Agency Lobbying

Peter Bils   Richard W. DiSalvo

Abstract

Federal agencies are important actors in the creation and implementation of policy, and thus are attractive forums for lobbying. Although the influence of groups and firms on agency policies is well-documented, less is known about the mechanisms by which lobbyists influence the bureaucracy. In this paper, we begin to study this question by exploring the role of lobbyists for agency lobbying. To do so, we explain why individual lobbyists lobby certain agencies. We start by examining the partisanship of lobbyists, finding that this does not appear to be a strong explanation. Instead, our analysis suggests that agency lobbying is primarily organized around lobbyists acting as issue area specialists. Our findings have important implications for how lobbying influences the bureaucratic policymaking process. Additionally, our results highlight the importance of distinguishing bureaucratic lobbying from legislative lobbying.

1 Introduction

Federal agencies are central actors in the development and implementation of policies. Furthermore, agencies act across a broad range of issue areas and, thus, impact a variety of stakeholders. Consequently, agencies attract a significant amount of attention from lobbyists and interest groups.[1] Indeed, researchers have shown that firms and interest groups can influence policy through the bureaucracy.[2] However, although the impact of lobbying on bureaucratic policymaking is well-documented, the process by which lobbying influences agency policy choices is less understood.

In this paper, we begin to unpack this process by studying the role played by the lobbyists themselves. Despite the prevalence of agency lobbying, previous work has focused on the role of lobbyists in lobbying legislators (e.g., Bertrand, Bombardini and Trebbi, 2014; Hirsch et al., 2021). Our approach is to study the factors that explain why individual lobbyists lobby certain agencies and not others. By explaining these agency-lobbyist connections we can understand the characteristics that make individual lobbyists valuable for interacting with agencies and, thus, provide insight into how lobbying shapes bureaucratic policymaking.

Previous research has shown in other contexts that partisanship is important for lobbyists and for lobbying. First, based on their campaign contributions, lobbyists appear highly partisan at an individual level (Koger and Victor, 2009). Second, partisan ties are valuable for lobbying firms (Furnas, Heaney and LaPira, 2019). Third, using data on foreign lobbying, researchers have shown that ideological similarity can explain con-

[1] You (2017) shows that most lobbying happens after a bill is passed, and approximately 50% of this activity is directed at administrative agencies.

[2] Previous work on lobbying of federal agencies has focused on issues of regulatory capture (Dal Bó, 2006; Johnson and Kwak, 2011), the points in the bureaucracy at which regulation and policy are open to influence (McKay, 2011; Haeder and Yackee, 2015), the best means and targets of lobbying (De Figueiredo and Tiller, 2001; Naoi and Krauss, 2009), and influence through the notice-and-comment process (Yackee and Yackee, 2006; Libgober, 2020). For example, firms that lobby the SEC and hire former SEC workers face less enforcement and lower penalties (Correia, 2014) and are cited more in final rules (Ban and You, 2019).

tacts between lobbyists and legislators (Hirsch et al., 2021; Liu, 2022). Consequently, partisanship may be an important driver of agency lobbying as well.

We begin by investigating this possibility. Our initial analysis studies the relationship between a lobbyist's partisanship and the average ideology of agencies the lobbyist lobbies. Although we find a positive and statistically significant relationship, the model has little explanatory power. Moreover, even using a more flexible model (random forest) we find that lobbyist partisanship is a very poor predictor of agency choice. We conclude that, overall, partisanship is not the primary lobbyist characteristic that drives agency lobbying.

Given the relatively weak role of partisanship, we next seek to provide a general description for agency-lobbyist connections. Rather than searching for more observables to explain agency venue choice, we use a unsupervised strategy (factor analysis) to construct quantitative measures that are more predictive from the data. Intuitively, when we observe an individual lobby two different agencies we use this link to group agencies together and use these groupings to interpret the role of the lobbyist.

This analysis yields a space of five factors. The agencies with the highest scores on each factor appear to work on similar policy issues. Looking at these agencies, we can name the five factors as: Finance, Military, Environment, Healthcare, and Trade. In particular, none of our factors look like ideology, and they all look like policy domains.

Consequently, our analysis suggests that agency lobbying is organized around policy domains. We conclude from this that lobbyists act as issue area specialists, rather than as partisan screening devices or purveyors of partisan political connections. In particular, agency lobbyists can broadly be characterized by how much they focus on issues of finance, military, environmental, healthcare, or trade. Additionally, a byproduct of our factor analysis is that we obtain measures of how much each lobbyist specializes on each of these policy domains.

2 Data and Background

Our study employs data on lobbyists, their political donations, and the agencies they contact while lobbying.

To obtain information on lobbyists and the agencies they contact we use data disclosed under the Lobbying Disclosure Act of 1995 and later amended in 2007. Disclosure policies that provide more detailed links between lobbyists and the federal agencies they contact (i.e., an expanded form L-2), as well as detail the political contributions made by lobbyists (i.e., form LD-203) went into effect starting on January 1st, 2008. For this reason, we use data from the start of 2008 through the end of 2020. See Table 1 for descriptive statistics.

Table 1: Lobbyist-Cycle-Agency Combination Data Summary Statistics

Variable	(1) All Data	(2) Analytic Sample
Number of Agencies	244	137
Number of Lobbyist-Cycles	77,567	59,084
% of Lobbyist-Cycles With Donations	44	45
% of Lobbyist-Cycles With Partisanship	41	42
% of Lobbyist-Cycles Lobbying Pres. or Cong.	99	98
% of Lobbyist-Cycles Lobbying at Least ...		
1 Agency	76	100
2 Agencies	62	87
5 Agencies	36	53
10 Agencies	17	26
20 Agencies	4	6

Notes: Column (1) displays summary statistics for all of the data, while Column (2) displays summary statistics for our analytic sample, described in the text.

For part of our study, we wish to identify the partisanship of the lobbyists in our data. Following previous work, we employ campaign finance data. Under the 2007 amendments, lobbyists must report their campaign finance donations directly on the LD-203 forms. Although they must disclose the name of the recipient, they do not report the recipient's party. We use OpenSecrets data for 2008 through 2020 and conduct a name match to link recipients to federal political candidates and their partisanship. Among the transactions in the LD-203 forms, we can link 86% to partisanship. Once we have a recipient name to party crosswalk, for each lobbyist-cycle, we take the average share of party-identified money donated to Republicans versus Democrats as our measure of lobbyist partisanship.

In our analysis we use a restricted sample. To construct this sample, we include only lobbyists who contact at least one federal agency. We additionally remove all agencies linked to fewer than 100 matched lobbyists and the five non-agency targets "House of Representatives," "Senate," "White House," "Vice President of the U.S.," "Executive Office of the President," "Undetermined," and "None."

As Table 1 shows, nearly all lobbyists in all years appear on reports of lobbying directly targeting the legislative or executive branches. Nevertheless, and importantly for our analysis, the rate at which lobbyists lobby federal agencies is high, about 76% of lobbyist-cycles in our sample. Indeed, among lobbyists in our analytic sample over half of lobbyist-cycles report lobbying at least 5 agencies.

3 Lobbyist Partisanship and Agency Choice

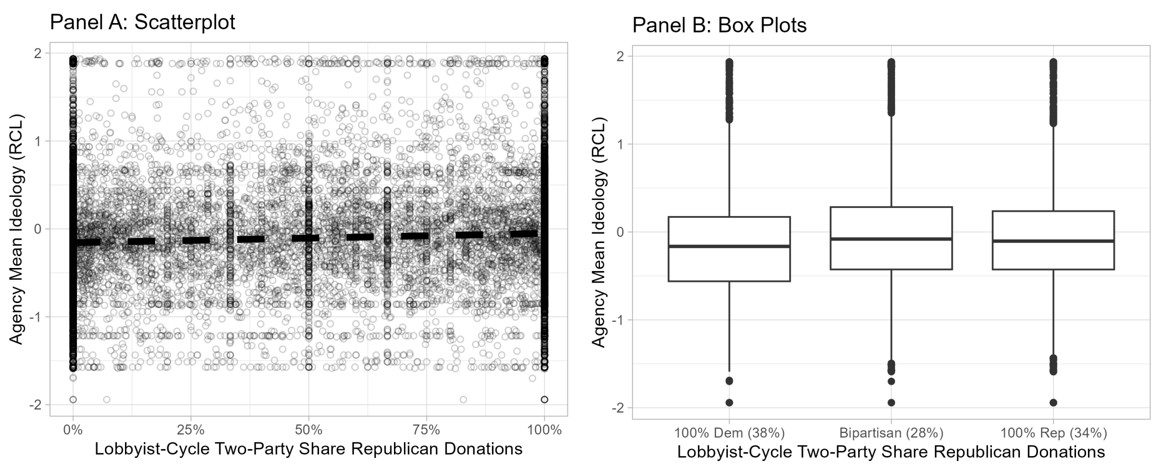
We now study the relationship between lobbyist partisanship and agency lobbying. We begin by presenting the bivariate relationship between the ideology of the agency lobbied by the lobbyist and the lobbyist's partisianship. To measure agency ideology, we use scores from Richardson, Clinton and Lewis (2018), while lobbyist partisanship is measured as previously described.[3]

Figure 1, Panel A, displays a scatterplot of these two variables, with a dashed line of best fit. There is a statistically significant positive relationship between these two variables, with a slope of about 0.11 (which is in units of standard deviations in the

[3]Richardson, Clinton and Lewis estimate perceived agency ideologies on a liberal-conservative spectrum using a survey of federal executives.

distribution of agencies). However, as the scatterplot makes clear, there is a large amount of scatter around the line, and the R^2 of the regression is only 0.005. This suggests lobbyist partisanship has relatively little explanatory power over mean agency partisanship. In Panel B, we repeat this descriptive analysis, using box plots instead because about two-thirds of lobbyist-cycles in our data have donations that are purely Democratic or purely Republican.

Figure 1: Predicting Average Partisanship of Agencies Lobbied Using Lobbyist-Cycle Campaign Contribution Partisanship



Notes: Figure displays plots relating the lobbied agency mean ideology based on Richardson, Clinton and Lewis (2018) scores (RCL; y-axis) to lobbyist-cycle partisanship based on two-party campaign finance donation shares (x-axis). Panel A displays this relationship using a scatterplot. The dashed line is a linear regression of best fit; the slope on this regression is 0.106 with a standard error (clustered on lobbyist) of 0.012 (significant at $p < 0.01$), but the R^2 is only 0.0054. Panel B studies this relationship using box plots for 100% Democratic donations (left most), a mixture (bipartisan) donations (middle), and 100% Republican donations (right most). The percents in parentheses are the share of lobbyist-cycles of each type (for example, 38% of lobbyist-cycles had a two-party share Democratic donation of 100%).

While suggestive, these preliminary results are inconclusive. First, agency ideology may be difficult to measure. If so, then part of the reason for a low R^2 is noise in the dependent variable. In addition, the relationship between partisanship and agency choice may not be linear. To circumvent this issue, we next take a machine learning approach and estimate vastly more flexible models. In these models, we explain whether a lobbyist-cycle lobbies each particular agency using a random forest model, fit for each of the 137 agencies in our data. By using these statistical models, we make very few assumptions about the sign or functional form of the relationship. These relationships could in theory be quite complicated; our goal is to see if there is a strong predictive relationship, regardless of the form of the relationship.

The outcomes we study in these random forests models are indicator variables: they are 1 if the lobbyist-cycle lobbies that agency, and 0 otherwise. A commonly used measure of goodness-of-fit for binary classification problems is the F1 Score (the harmonic mean of precision and recall). As such, we maximize the F1 Score when fitting these random forests.

The F1 score is on a scale from 0 to 1. The F1 measure varies across agencies, but its maximum value is about 0.10. To summarize this model fit measure across agencies we weight each agency by the number of lobbyist-cycles used in estimation. We find that this overall average F1 Score is very low, only 0.041.[4]

Next, we consider whether partisanship plays a more significant role for certain types of lobbyists. In particular, it may be that partisan considerations are more important for those who make more campaign donations or high income lobbyist. We re-run the above analysis, but restricted to the top 20% of lobbyists for each category. Table 2 presents the results of this analysis. The predictive power of partisanship improves, especially for lobbyists who report in the top 20% of income and expenditures. Indeed, the maximum F1 score across agencies is quite high for such lobbyists, suggesting that partisan connections may play an important role for lobbying some agencies. However, the average F1 score remains low. Thus, partisanship does not appear to broadly explain agency lobbying even for this subset of lobbyists.

The overall lack of a predictive power suggests that lobbyist partisanship is not the primary determinant of agency lobbying. One caveat is that only 42% of the lobbyists in

[4] If we weight agencies equally, we obtain an even smaller average F1 Score of 0.014.

Table 2: Predicting Agency Lobbying: Top 20%

	All	**Income**	**Campaign Spending**
Weighted average F1 score	0.041	0.1375	0.0945
Maximum F1 Score	0.1	0.7104	0.186

our data have a measure of partisanship based on campaign finance records (see Table 1). It is possible that partisanship drives decision-making more or less for the remaining 58%. We believe it likely drives it less for that subgroup because they may be less motivated by political connections or ideology, given they did not make donations.

4 The Organization of Agency Lobbying

Having found that partisanship does not adequately explain agency lobbying, we turn to an exploratory factor analysis to study the structure underlying the connections between lobbyists and agencies. In the Appendix we provide further details on the estimation process, the robustness of our results, and a microfoundation for using this particular approach.

The exploratory factor analysis model identifies the underlying features that determine whether a lobbyist lobbies a given agency and outputs these features as a list of K factors. Thus, by observing which agencies a lobbyist lobbies we are able make inferences about the lobbyist's latent characteristics. In this model, the *lobbyists* are the agents who have factors. In short, this means that individual lobbyists have specific characteristics, such as ideology or policy domain experience, and this leads them to lobby specific federal agencies. For each factor k and lobbyist i we estimate a score F_{ik}. As we find that the factors can be interpreted as policy domains, we interpret F_{ik} as how much lobbyist i specializes in issue area k.

As part of our analysis, we determine K, the number of factors. Using a standard criterion, we find that five factors is best, so we focus on $K = 5$. When we fit the model with $K = 5$, we explain about half of the variation. Increasing the number of factors beyond five adds little explanatory power (see Appendix Section A.2.3).

We next examine each factor, with the goal of naming it. To do so, we study the agencies that load the heaviest on each factor. Table 3 Panel A displays, for each of the five factors, the three agencies which have the highest factor loadings; the estimated factor loadings are also displayed with bootstrapped standard errors in parentheses.[5] The list of agencies in Panel A are the "purest" agencies, because they are those agencies whose lobbying links are most readily explained by the given factor. At the top of the table, we name the factors based on our understanding of these agencies. The five factors appear to be finance, military, environmental policy, healthhcare, and trade.[6]

Panel B instead lists what we call the "most influential agencies" for each factor. We measure influence using the probability that an agency is lobbied by a lobbyist who is highly specialized on that domain (see Appendix Section A.2.4 for details). Intuitively, this measure tends to recover the largest agencies. Additionally, now some agencies appear across multiple domains. For example, Commerce is primarily focused on the economy, however, it is important on an array of issues due to the scope of its mission.

[5] In the Appendix we expand to the top 5 agencies.

[6] We note that the factor loadings on the most pure agencies in Trade are typically much lower than those most pure agencies in the other factors, which suggests this factor is less cleanly separated in the model.

Table 3: Five Latent Factors Interpreted as Policy Domains

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Finance	**Military**	**Environment**	**Healthcare**	**Trade**
Panel A: Purest Agencies				
0.86 (0.03) - Federal Reserve System	0.82 (0.04) - U.S. Marines	0.72 (0.04) - Bureau of Land Management (BLM)	0.84 (0.03) - Health Resources & Services Administration (HRSA)	0.54 (0.47) - U.S. International Trade Commission (ITC)
0.86 (0.02) - Equal Employment Opportunity Commission (EEOC)	0.77 (0.04) - Navy	0.72 (0.04) - U.S. Fish & Wildlife Service (USFWS)	0.83 (0.02) - Agency for Healthcare Research & Quality (AHRQ)	0.53 (0.53) - Pipeline & Hazardous Materials Safety Administration
0.81 (0.06) - Office of the Comptroller of the Currency (OCC)	0.71 (0.05) - Air Force	0.70 (0.02) - Bureau of Reclamation	0.79 (0.02) - Centers For Medicare and Medicaid Services (CMS)	0.52 (0.54) - Surface Transportation Board (STB)
0.81 (0.02) - Securities & Exchange Commission (SEC)	0.70 (0.09) - U.S. Immigration & Customs Enforcement (ICE)	0.69 (0.03) - U.S. Forest Service	0.75 (0.02) - Medicare Payment Advisory Commission (MedPAC)	0.51 (0.43) - Defense Logistics Agency
0.80 (0.03) - Federal Deposit Insurance Corporation (FDIC)	0.69 (0.07) - Natl Security Agency (NSA)	0.65 (0.04) - Natl Park Service (NPS)	0.69 (0.03) - Substance Abuse & Mental Health Services Administration (SAMHSA)	0.51 (0.46) - Intl Trade Administration (ITA)
Panel B: Most Influential Agencies				
0.61 (0.03) - Treasury	0.37 (0.02) - Defense (DOD)	0.42 (0.03) - Environmental Protection Agency (EPA)	0.59 (0.01) - Health & Human Services (HHS)	0.27 (0.10) - U.S. Trade Representative (USTR)
0.43 (0.02) - Commerce (DOC)	0.31 (0.02) - State (DOS)	0.32 (0.01) - Transportation (DOT)	0.39 (0.01) - Centers For Medicare and Medicaid Services (CMS)	0.25 (0.10) - Environmental Protection Agency (EPA)
0.39 (0.02) - Health & Human Services (HHS)	0.30 (0.01) - Commerce (DOC)	0.30 (0.01) - Agriculture (USDA)	0.27 (0.01) - Office of Management & Budget (OMB)	0.25 (0.05) - Commerce (DOC)
0.38 (0.01) - Office of Management & Budget (OMB)	0.27 (0.03) - Homeland Security (DHS)	0.29 (0.01) - Energy	0.26 (0.01) - Food & Drug Administration (FDA)	0.18 (0.03) - State (DOS)
0.37 (0.02) - Homeland Security (DHS)	0.22 (0.01) - Health & Human Services (HHS)	0.26 (0.01) - Commerce (DOC)	0.17 (0.01) - Agriculture (USDA)	0.17 (0.04) - Energy
HHI = 0.0361 (0.0004)	0.0405 (0.0010)	0.0421 (0.0005)	0.0515 (0.0010)	0.0441 (0.0014)

Notes: Table 3 displays information related to the five factors obtained from the factor analysis. The top of each column names the factor. Each column lists the top agencies by the estimated factor loading. Throughout this table, estimates are shown with bootstrapped standard errors in parentheses. We find that about half (50.3%) of the variation (weighting agencies equally) is explained by five factors.

To give concrete examples of our "policy specialiaztion" interpretation, Table 4 provides the estimated factors for three of the lobbyists in our data. We see that each lobbyist score in one policy domain significantly larger than in the other four domains, suggesting these lobbyists tend to focus their efforts on specific issues. Looking into the backgrounds of these three lobbyists confirms this interpretation.

Michael Park is a lobbyist at Alston & Bird who "focuses his practice on representing health care providers, insurers, and manufacturers before Congress and the Administration on a wide range of health care legislative and regulatory issues."[7] Previous positions include being an advisor to Senator Grassley, and stints at the HHS and the Centers for Medicare & Medicaid Services. Lauren Bazel previously worked as a tax policy advisor for a number of U.S. senators and her earlier work as a lobbyist at Ernst & Young focused on federal corporate tax issues.[8] Finally, Katherine Hamilton is co-founder and Chair at 38 North Solutions "a public policy firm focused on clean energy and innovation."[9]

Table 4: Average Factor Scores for Selected Lobbyists

Lobbyist Name	Finance	Military	Environment	Healthcare	Trade
Katherine Hamilton	0.091	0.014	**0.181**	0.016	0.057
Lauren Bazel	**0.092**	-0.031	0.021	0.017	-0.002
Michael Park	0.095	-0.029	-0.047	**0.338**	0.006

Notes: Average factor scores are reported for the preferred five factor model (scores are averaged across all cycles the lobbyist appears in the data, weighting cycles equally). The largest factor score in each row is bolded. These average factor scores are not standardized.

Overall, the factors identified through this method closely align with policy domains. Moreover, when we regress measures of agency ideology on the agency loadings for each factor we only find consistent results for the military and healthcare domains, further suggesting these factors are not merely proxies for ideology (see Appendix Section A.4).

The factors give an estimate of how much a lobbyist specializes in a given domain.

[7] See: https://www.alston.com/en/professionals/p/park-michael-h.

[8] See: https://www.alpinegroup.com/bazel

[9] See: https://www.linkedin.com/in/katherine-hamilton-8728416.

Taking a lobbyist's maximum score, i.e., $\max\{F_{i1}, F_{i2}, F_{i3}, F_{i4}, F_{i5}\}$ across factors we can categorize lobbyists by their issue area specialization. Table 5 lists the percentage of lobbyists working in each policy area. Perhaps unsurprisingly, a significant portion of lobbyists appear to focus their practice on finance.

Table 5: Distribution of Lobbyist Specializations

Environment	**Finance**	**Healthcare**	**Military**	**Trade**
24%	35%	24%	14%	3%

We also use the factors to construct a simple measure of whether a lobbyist scores highly in one issue area or is relatively spread across the issue areas. Specifically, let lobbyist i's specialization be given by $\max_k F_{ik} - \min_k F_{ik}$. We then regress this measure on the lobbyist's reported income + expenditures (logged). We find a positive (estimate .91) and statistically significant effect (clustered s.e. of .041). Thus, more important lobbyists also tend to be more specialized.

Finally, we demonstrate that the factors derived from this analysis predict which agencies a lobbyist will lobby. To this end, we run random forest models identical to those in Section 3, predicting the indicator variable of agency choice for each agency using the lobbyist factor scores. The average F1 score in this case is 0.81, compared to the F1 score of 0.04 from the lobbyist partisanship model. We conclude that a five-factor model is a significant improvement over the lobbyist partisanship model, in terms of explaining lobbyist-agency connections. Given how the factors are estimated, it is not surprising that this prediction exercise would do better than that in Section 3. However, we perform this exercise in order to clearly show how much better the five factors we have identified using our model predict agency choice.

5 Discussion

Previous research has found that ideology plays an important role for Congressional lobbying (Hirsch et al., 2021; Liu, 2022) and that lobbyists themselves are partisan (Koger and Victor, 2009), whereas we find that lobbyist partisanship plays a minor role in bureaucratic lobbying compared to policy domain focus. Consequently, lobbyists may play a very different role in the bureaucratic policymaking process than in the legislative process. This suggests that lobbying responds to the different institutional settings. Consequently, caution is warranted when applying findings from legislative lobbying to the bureaucracy.

Our results help us better understand the role lobbyists play in the bureaucratic policymaking process. For instance, the most prominent perspectives argue lobbying is either a *quid pro quo* exchange (e.g., Grossman and Helpman, 1992) or informational, i.e., that lobbyists provide valuable information to policymakers (e.g., Potters and Van Winden, 1992). Although our findings do not adjudicate between these two theories, they have implications for each theory. If lobbying is direct exchange, then the market for agency policy appears to be segmented based on policy domains. In particular, issue area specialization is more important for gaining access to agencies and making such exchanges than ideological similarity. If lobbying is informational, then our results suggest that policy domain expertise is of primary importance for communicating to agencies. Moreover, if bias leads to strategic communication of information by lobbyists to agencies, then the source of the bias is likely from something other than partisanship.

Another important debate in the literature is whether lobbyists are compensated for expertise or connections (e.g., Bertrand, Bombardini and Trebbi, 2014). We caveat our results by noting that, although our results may *prima facie* suggest it is expertise and not connections that drive lobbyist behavior in this area, this need not be the case. It may be that bureaucrats and lobbyists working in the environment domain are rewarded due to having connections to environmental agencies and such lobbyists have little issue area expertise. However, although such reasoning may be plausible for elected officials, it is less likely to apply to bureaucrats promoted through the civil service based on arguably less political dealing. So while our findings remain consistent with only connections driving agency lobbying, when combined with other institutional details of the bureaucracy, they tend to run against this theory. More broadly, it may be that expertise is a necessary condition for lobbying agencies, but connections made while working in a policy domain are important for determining which lobbyists are most effective.

The results presented here improve our understanding of how lobbyists interact with federal agencies. Moving forward, it is important to try to further disentangle the mechanisms that drive agency lobbying. Additionally, continued work on both Congressional and agency lobbying is crucial for further exploring how lobbying operates across different branches of government, and how the organization of lobbying varies based on context.

References

Anderberg, Michael R. 1973. *Cluster analysis for applications.* Probability and mathematical statistics, 19 New York: Academic Press.

Ban, Pamela and Hye Young You. 2019. "Presence and influence in lobbying: Evidence from Dodd-Frank." *Business and Politics* 21(2):267–295.

Bertrand, Marianne, Matilde Bombardini and Francesco Trebbi. 2014. "Is It Whom You Know or What You Know? An Empirical Assessment of the Lobbying Process." *American Economic Review* 104(12):3885–3920.
URL: *www.jstor.org/stable/43495360*

Chen, Jowei and Tim Johnson. 2015. "Federal employee unionization and presidential control of the bureaucracy: Estimating and explaining ideological change in executive agencies." *Journal of Theoretical Politics* 27(1):151–174.

Correia, Maria M. 2014. "Political connections and SEC enforcement." *Journal of Accounting and Economics* 57(2-3):241–262.

Dal Bó, Ernesto. 2006. "Regulatory capture: A review." *Oxford Review of Economic Policy* 22(2):203–225.

De Figueiredo, John M and Emerson H Tiller. 2001. "The structure and conduct of corporate lobbying: How firms lobby the Federal Communications Commission." *Journal of Economics & Management Strategy* 10(1):91–122.

Drasgow, Fritz. 2004. "Polychoric and polyserial correlations." *Encyclopedia of statistical sciences* 9.

Furnas, Alexander C, Michael T Heaney and Timothy M LaPira. 2019. "The partisan ties of lobbying firms." *Research & Politics* 6(3):2053168019877039.

Grossman, Gene M and Elhanan Helpman. 1992. “Protection for sale.”.

Haeder, Simon F and Susan Webb Yackee. 2015. “Influence and the administrative process: Lobbying the US President’s Office of Management and Budget.” *American Political Science Review* 109(3):507–522.

Hirsch, Alexander V, Karam Kang, B Pablo Montagnes and Hye Young You. 2021. “Lobbyists as gatekeepers: Theory and evidence.”.

Johnson, Simon and James Kwak. 2011. *13 bankers: The Wall Street takeover and the next financial meltdown*. Vintage.

Kim, Jae-on and Charles Mueller. 1978. *Factor Analysis*. 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc.
URL: *http://methods.sagepub.com/book/factor-analysis*

Koger, Gregory and Jennifer Nicoll Victor. 2009. “Polarized Agents: Campaign Contributions by Lobbyists.” *PS: Political Science & Politics* 42(3):485–488. Publisher: Cambridge University Press.
URL: *https://www.cambridge.org/core/journals/ps-political-science-and-politics/article/polarized-agents-campaign-contributions-by-lobbyists/28130F9D36803AF22D0875DB5CF63E1E*

Libgober, Brian. 2020. “Meetings, Comments, and the Distributive Politics of Rulemaking.” *Quarterly Journal of Political Science* 15(4):449–481.

Liu, Huchen. 2022. “Campaign Contributions and Access to Congressional Offices: Patterns in Foreign Lobbying Data.” *Political Research Quarterly* 75(3):812–828.

McKay, Amy Melissa. 2011. “The decision to lobby bureaucrats.” *Public choice* 147(1-2):123–138.

Mulaik, Stanley A. 2010. *Foundations of factor analysis.* Statistics in the social and behavioral sciences series. second edition. ed. Boca Raton: CRC Press.
URL: *"https://www.taylorfrancis.com/books/9781420099812":["www.taylorfrancis.com"]*

Naoi, Megumi and Ellis Krauss. 2009. "Who lobbies whom? Special interest politics under alternative electoral systems." *American Journal of Political Science* 53(4):874–892.

Pearson, Karl. 1900. "I. Mathematical contributions to the theory of evolution. —VII. On the correlation of characters not quantitatively measurable." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 195(262-273):1–47.
URL: *https://royalsocietypublishing.org/doi/abs/10.1098/rsta.1900.0022*

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

Richardson, Mark D, Joshua D Clinton and David E Lewis. 2018. "Elite perceptions of agency ideology and workforce skill." *The Journal of Politics* 80(1):303–308.

Selin, Jennifer L. 2015. "What makes an agency independent?" *American Journal of Political Science* 59(4):971–987.

Warnes, Gregory R., Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz and Bill Venables. 2019. *gplots: Various R Programming Tools for Plotting Data.* R package version 3.0.1.1.
URL: *https://CRAN.R-project.org/package=gplots*

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

You, Hye Young. 2017. "Ex post lobbying." *The Journal of Politics* 79(4):1162–1176.

A Technical Details and Additional Figures

A.1 Predicting Agency Lobbying with Partisanship

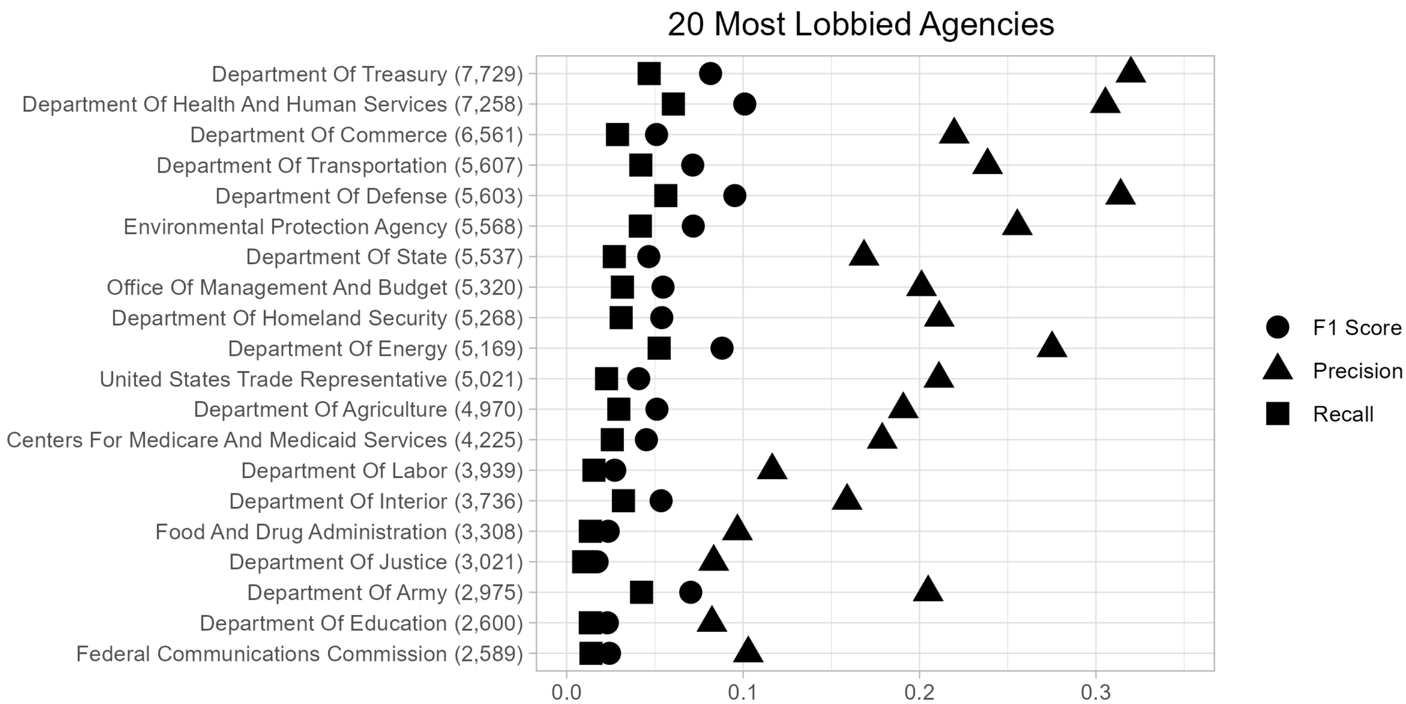
The outcomes we study in these random forests models are indicator variables: they are 1 if the lobbyist-cycle lobbies that agency, and 0 otherwise. Two measures of goodness-of-fit for binary classification problems are precision and recall. Precision is the share among cases that the model predicts the lobbyist should lobby this agency in which the lobbyist actually does lobby the agency, that is, Pr(Lobbies|Predicted to Lobby). Recall is the share among cases that the lobbyist-cycle does lobby that we predict the lobbyist will lobby, that is, Pr(Predicted to Lobby|Lobbies). These are in tension: if we predict very few lobbyists will lobby the agency, then we can generally get a high precision (since among those we do predict to lobby, the probability that they do can be made quite high); on the other hand, if we predict a lot of lobbyists will lobby the agency, we can make recall high (in fact, if we predict all lobbyist will lobby, recall will mechanically be 1). Because of this tension, we use the harmonic mean of these two measures as an omnibus measure of model fit. The harmonic mean of these measures is known as the F1 Score, and is commonly used in classification problems. When we fit these random forests, we maximize the F1 Score.

Table 6: F1 Scores from Predicting Agency Choice Using Partisanship

Case	Weighted average F1 Score	Unweighted average F1 Score	Maximum F1 Score
All lobbyist-cycles	0.041	0.014	0.100 (HHS)
Restricted to top 20% lobbyist-cycles by sum of report income + expenditure	0.138	0.038	0.710 (Treasury)
Restricted to top 20% of lobbyists-cycles by total lobbyist-cycle campaign spending	0.094	0.038	0.186 (DoD)

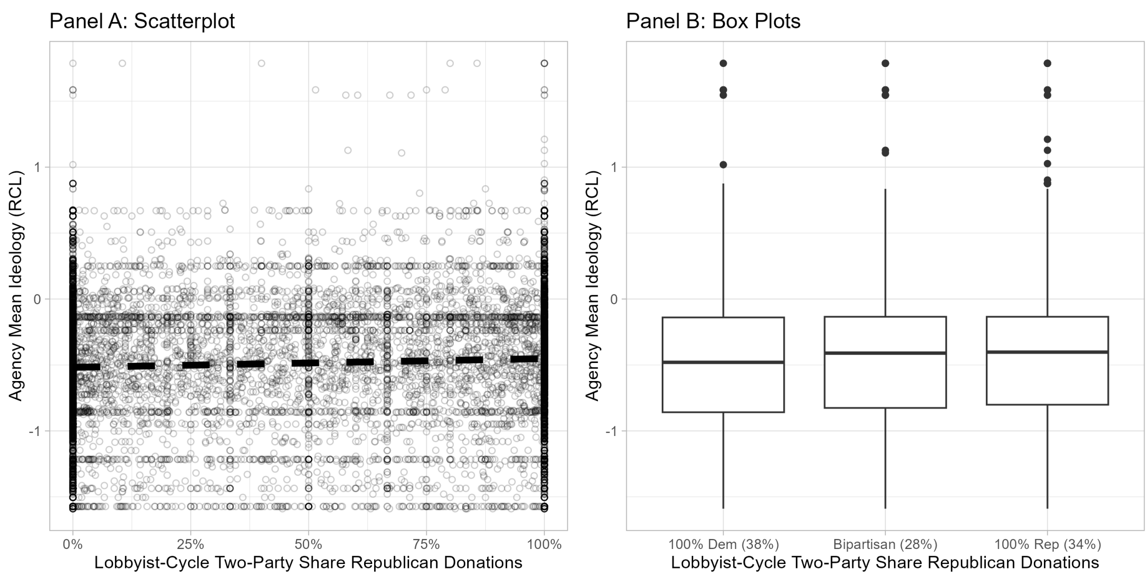
Notes: Table displays F1 scores for different lobbyist-cycle subsets. The weighted average F1 scores weights agencies by the number of lobbyist-cycles, while the unweighted average F1 score weights lobbyists equally. In the final column, the agency that achieves the maximum F1 score is in parentheses (HHS = Department of Health and Human Services; DoD = Department of Defense).

Figure 2: Predicting Whether a Lobbyist-Cycle Lobbies an Agency Using Lobbyist-Cycle Partisanship



Notes: The number of lobbyist-cycles in the analytic sample (including both training and testing subsamples) for each agency is displayed in parentheses after the agency name.

Figure 3: Predicting Average Partisanship of Agencies Lobbied Using Lobbyist-Cycle Campaign Contribution Partisanship for Only the 48 Least Independent Agencies



Notes: Figure is analogous to Figure 1 in the main text, except the data is restricted to the 48 agencies with below-median independence estimates based on the Selin (2015) "decision makers" (dimension 1) measure (the median is taken over the 96 matched agencies in our data). The dashed line is a linear regression of best fit; the slope on this regression is 0.066 with a standard error (clustered on lobbyist) of 0.01 (significant at $p < 0.01$), but the R^2 is only 0.0035.

A.2 Factor Analysis

A.2.1 Correlation Matrix

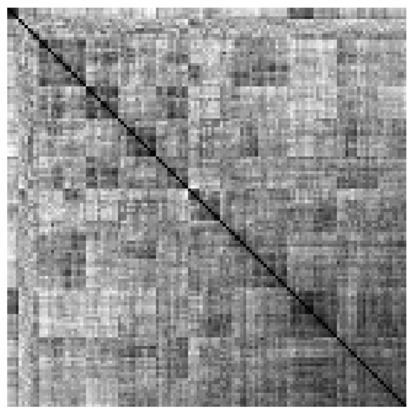
We begin our analysis by constructing a correlation matrix for the lobbyist-cycle-agency links in our data. Specifically, our data provides for each lobbyist-cycle-agency combination a dichotomous variable, "lobbied," that takes on the value one if and only if the lobbyist is on at least one report together with the agency in that cycle. Thus, we can consider the between-agency association between this lobbyist-cycle-level variable.

The usual Pearson correlations between the dichotomous variables are not ideal for our purposes, since these will be mechanically low whenever one or both of the agencies is rarely or commonly lobbied, and there is wide variation across agencies in the overall level of lobbying. The tetrachoric correlation (Pearson, 1900; Drasgow, 2004) overcomes this problem. In this approach, instead of correlating the binary variables of the two agencies directly, we assume that these variables are generated by a pair of latent random variables which are bivariate normal. Specifically, each variable takes a value of 1 if and only if its underlying normal random variable (a marginal distribution) exceeds a cut point. We then estimate and use the Pearson correlation for the latent bivariate normal distribution as a measure of association.

Note that the rows and columns of this correlation matrix are federal agencies. We thus now have a measure of "closeness" of federal agencies, as implied by lobbying connections. Specifically, agencies with a greater tetrachoric correlation are closer in the sense that lobbyists that lobby one agency are more likely to lobby the other (and vice versa). With this correlation matrix, we can provide preliminary evidence on how lobbyist-agency connections strategically cluster agencies, and therefore on how lobbyists are choosing among these agencies.

Figure 4 displays a heat map of the correlation matrix just described. The heat map is sorted using the complete-linkage hierarchical clustering algorithm, a simple clustering algorithm to be described shortly. It is clear from this exercise that the data has substantial tight clusters, which can be seen as darkened squares along the diagonal. These squares represent groups of federal agencies that are bound together by common lobbyists. A careful look at the sorted heat map suggests approximately five distinct agency clusters, or possibly four or six.

Figure 4: Heat Maps of the Tetrachoric Correlation Matrix for Lobbying of Federal Agencies



Notes: Figure displays heat map representations of the tetrachoric. Darker shades represent greater (more positive or less negative) correlations. The heat map is sorted using the complete-linkage hierarchical clustering algorithm; see the text for further description.

The clustering algorithm used in Figure 4 is complete-linkage hierarchical clustering, which is computationally simple but has disadvantages. Specifically, the clustering algorithm is used to construct a dendrogram (similar to a binary tree); then, rows are ordered in the heat map by the dendrogram, and within dendrogram levels, rows are ordered (arbitrarily) by row means. The complete linkage algorithm constructs the dendrogram as follows. First, the distance between every pair of agencies is calculated. The distance is defined to the Euclidean distance between the two agencies' rows of correlations in the matrix. Second, the agencies that are closest according to this distance are grouped. Once this is completed, distances between pairs of groups are calculated as the maximum distance between every possible pair of agencies, one taken from each group. The closest groups (allowing individual agencies to be groups as needed) are then grouped. The process continues until all agencies have been grouped. See Anderberg (1973) for a detailed discussion of hierarchical and other clustering methods, and the documentation for the R function heatmap.2 (Warnes et al., 2019) for further details.

The disadvantage of hierarchical clustering is that it is "greedy," not systematic: there could still be gains from rearranging rows (Anderberg, 1973); moreover, it does not quantify the tightness of the clusters. However, it provides an illustrative preliminary view of the correlation matrix we have constructed. In the following subsections, we motivate and implement an exploratory factor analysis model which has a stronger theoretical justification.

A.2.2 Microfoundation

We begin with a simple model of lobbyist behavior, which motivates our factor analysis methodology.

Consider a lobbyist i choosing among agencies a. A lobbyist may lobby as many agencies as he or she desires; his or her payoff (utility) from lobbying agency a is given by the number X_{ai}, and he lobbies a if this is positive. We assume there is a finite number K of factors that determine venue choice. Agency a is characterized by a length-K vector, $\boldsymbol{\beta}_a = (\beta_{a1}, \beta_{a2}, ..., \beta_{aK})$. Additionally, each agency a has a factor-independent utility term R_a, which captures the overall size of the agency.

The lobbyist's payoff X_{ia} is determined as follows. The lobbyist has a factor vector $\boldsymbol{F}_i = (F_{i1}, F_{i2}, ..., F_{iK})$ which describe the lobbyist's utility function. Then we assume the lobbyist's payoff from lobbying agency a is R_a, plus the dot product of the agency's characteristics and the lobbyist's utility vector, plus a normally distributed error,

$$X_{ia} = R_a + \boldsymbol{\beta}_a \cdot \boldsymbol{F}_i + E_{ia}$$

The orthogonal exploratory factor analysis we employ assumes that $Cor(F_{ij}, F_{ik}) = Cor(E_{ij}, E_{ik}) = 0$ for all $j \neq k$, and $Cor(F_{ij}, E_{ik}) = 0$ for all j and k (Kim and Mueller, 1978). From these assumptions, then the only common driver of lobbying different agencies are the dot products $\boldsymbol{\beta}_a \cdot \boldsymbol{F}_i$, so the model imposes that the only reason lobbying will be correlated across agencies is because agencies to some extent have weights β_{ak} on the same factors F_{ik}.

Note that because we will use the correlation matrix in order to estimate this model, we have implicitly scaled X_{ia} so that the mean of X_{ia} is R_a and the standard deviation of X_{ia} is one. Thus, all that identifies the policy domains β_{ak}, are the correlations in decisions, rather than agency sizes, that is, the number of unique lobbyists who have lobbied agency a is not directly relevant for determining β_{ak}. This facilitates interpretation of the β_{ak} as policy domains since then the "purest" agencies will be identified as having the highest loadings on particular factors, rather than being simply the largest agencies.

Moreover, estimation of this model proceeds by maximum likelihood, and assumes that the F_{ij} are multivariate normal, each with mean zero and standard deviation one; and the E_{ia} are multivariate normal, each with mean zero and variance (uniqueness) $\psi_a = 1 - \sum_k \beta_{ak}^2$. Putting this together with the previous paragraph, for each a, $X_{ia} - R_a$ is standard normal.

A.2.3 Estimation & Determining K

Beginning with the latent normally distributed variables $\boldsymbol{X}$ whose correlation matrix was obtained and analyzed in subsection A.2.1, we assume that for each agency a the standard normal latent variable $X_{ai} - R_a$, which varies over lobbyists i, is determined by K factors,

$$X_{ai} - R_a = \beta_{a1} F_{i1} + \beta_{a2} F_{i2} + \cdots + \beta_{aK} F_{iK} + E_{ai} \tag{1}$$

This *exploratory factor analysis* model relates the latent, normally distributed constructs that determine whether lobbyists lobby the given agency to a list of factors $\boldsymbol{F}_k$, which we will find represent policy domains. In this model, the *lobbyists* are the agents who have factors, while *agencies* have factor *loadings*, i.e. each agency has a vector of parameters $\boldsymbol{\beta}_a$.

Importantly, if a lobbyist scores more highly on one of these factors, that will have a potentially different effect on the probability that the lobbyist lobbies each individual agency. For example, if we compare two lobbyists, who are equal in every way except the first has a higher F_{i1}, then the chance the first lobbyist will lobby agency a will be greater than the second by some amount related to β_{a1} times the difference in F_{i1} between the two lobbyists (since the dependent variable is itself a latent construct, a normal distribution will relate this linear effect to the actual event probability). Thus, the group of agencies whose lobbying is described heavily by factor 1 are those agencies with large and positive β_{a1}, and similarly for the other factors. Note that it is possible for factor loadings to be negative, so that if a lobbyist exhibits certain characteristics then it actually decreases the chance that lobbyist is linked to certain agencies. For example, a lobbyist with healthcare expertise may be less likely to lobby an environmental agency, or a right-leaning lobbyist may be less likely to lobby a left-leaning agency. Also note that the solution to the model will not be unique, but will be unique up to a choice of rotation. We employ the commonly used *varimax* rotation, which maintains the orthogonality of the F_{ik} while, roughly-speaking, maximizing of the ability of the factors to distinguish agencies (Mulaik, 2010).

This model has two latent, unobserved constructs: the dependent variables $X_{ai} - R_a$, and the factors, $F_{ik}, k = 1, ..., K$. While we never observe $X_{ai} - R_a$, in subsection A.2.1 we described and implemented the tetrachoric correlation method, which recovered the correlation matrix for the X_{ai} (i.e. the correlations between X_{ai}, X_{bi}, for each pair of agencies a, b). This correlation matrix is all that is needed to conduct the factor analysis.

Note that we perform a factor analysis, rather than a principal component analysis, since the structure of a factor analysis follows more closely the decision model we presented in the previous subsection. Specifically, a factor analysis views behavior X_{ai} as being caused by the factors F_{ik}; in the model, behavior is the choice of a lobbyist to lobby a particular agency, while the factors are the agency's unobserved policy domains, and it is the latter which causes the former. By contrast, principal component analysis would view the behaviors as causing the factors, in the sense that the X_{ai} would be on the right-hand side of the equation, rather than the left. Kim and Mueller (1978), chapter 2, contains a useful discussion of the differences between these methods.

Now that we have presented the general framework, we discuss the determination of K. In the literature, a commonly used plot for determining K is a *scree plot* (Mulaik, 2010). The idea of a scree plot is to show the marginal contribution of adding an additional factor to explanatory power, then pick the number of factors so that the marginal factor is discontinuously better than the remaining choices.

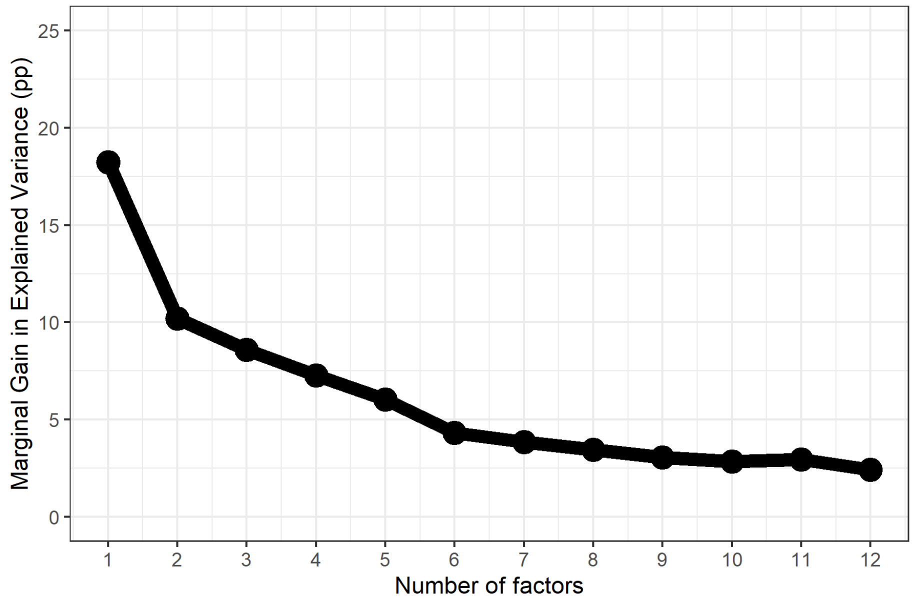
The scree plot for our factor analysis model is displayed in Figure 5. A rule-of-thumb is to pick the number of factors just before the marginal contributions flatten out and decline linearly. By our reading, five factors appears to satisfy this criterion, which was similar to what we obtained through hierarchical clustering, so we focus on $K = 5$.

When we fit the model in equation (1) with $K = 5$, we explain about half of the variation (but note that "variation" in these models weights agencies equally). Consistent with the scree plot and the heat map, increasing factors beyond five adds little explanatory power. However, we also consider the cases $K = 4$ and $K = 6$ in the next section.

A.2.4 Naming Factors

Having estimated the factor model described by equation (1), we examine each factor, with the goal of naming it. If the factor model fits the data in a consistent way, then by examining the agencies that load heavily on each factor, we expect to be able to name each factor.

Figure 5: Scree Plot for the Lobbying of Federal Agencies Correlation Matrix



Notes: Figure displays the marginal gain in the percent of variation explained (marginal gains in percentage points) by the number of factors included in the model.

It is also important to also give a sense of the precision of our estimates. In order to calculate standard errors for the factor loadings (the β_a in equation 1), we use a bootstrap method as follows. In each bootstrap iteration, we resample lobbying-agency links with replacement. We then run the factor analysis model on this resampled data. However, the factors that are obtained from the resampled data may not align with the factors in our main model fit, and especially the order of the factor names will vary randomly. We thus name the factors that come out of this method by first finding the factor from the bootstrap run whose factor loadings (the β_{ak}) have the highest correlation with those of our first factor from our main model (fit on the actual data). We call that factor the (bootstrap) "factor one." We then find, among the remaining bootstrapped factors, the one most correlated with our main factor two. We call that factor two. We then continue in this way until all factors in the bootstrap run are named. After running one thousand bootstrap replications through this approach, we calculate standard errors as the standard deviation of the bootstrapped (renamed) factor loadings.

Panel A of Table 7 displays, for each of the five factors, five agencies which have the highest factor loadings; the estimated factor loadings (the β_{ak} from equation 1) are also displayed with bootstrapped standard errors in parentheses. This is analogous to Table 3 in the main text.

The information in Panel A captures the purest agencies on each factor. While discarding information on agency size is useful for uncovering the latent factors, it is also important to understand the relative influence agencies have in these areas. Influence is likely inclusive of agency size. We leverage our model of lobbyist venue choice to create a measure of agency factor-specific influence, as follows. To begin, we define a *unit lobbyist* i on factor k as a lobbyist such that $F_{ik} = 1$ for some factor k and $F_{ik'} = 0$ for all factors $k' \neq k$, e.g., $(0, 1, 0, 0, ..., 0)$. Thus, the payoff to unit lobbyist i for lobbying agency a only depends on the agency's focus on dimension k and R_a, i.e.,

$$X_{ia} = R_a + \beta_k + E_{ia}.$$

We define the influence of agency a on factor k as the probability that a unit lobbyist on factor k lobbies agency a. Deriving this term yields the following measure for the influence of agency a on factor k:

$$I_{ak} = \Phi\Big(\frac{R_a + \beta_{ak}}{\sqrt{\psi_a}}\Big)$$

where Φ is the standard normal cumulative distribution function, and ψ_a is the variance of the error term E_{ia}. In factor analysis, ψ_a is called the "uniqueness" for agency a lobbying, and represents the variation in lobbying agency a that cannot be explained by the common factors. Moreover, since $X_{ia} - R_a$ is standard normal, and lobbying occurs when $X_{ia} > 0$, the (unconditional) chance that agency a is lobbied is $\Phi(R_a)$. Thus, we can solve for $R_a = \Phi^{-1}(S_a)$, where S_a is the percent of all lobbyists who lobby agency a. Thus I_{ak} is readily computed. Panel B of Table 7 lists the most influential agency on each factor.

When registrants file lobbying reports, they record relevant issue areas, providing further insight into their lobbying intent. Here, we use these reports as a check on our conclusion: if we are indeed recovering policy domains through our method and naming them correctly, then we would expect issue areas to line up in an intuitive way with each factor. To this end, Panel C of Table 7 lists the most common issues linked to the agencies listed in Panel A of the same column, with the number of times that each issue is linked to an agency in parentheses. By and large these results reinforce our naming of each factor. It is interesting to note that the "Budget/Appropriations" issue shows up among top linked issues for four out of five of the domains, and the "Taxation" issue shows up in every domain. We emphasize that in no part of our analysis have we used these issue areas directly; the fact that they appear to line up with the policy domains we have named is a direct consequence of lobbyist venue choices lining up in this fashion.

Tables 8 and 9 repeat Table 7 for two other choices of the number of factors, $K \in \{4, 6\}$. To summarize briefly, with four factors, we lose the Trade factor; with six factors, we gain a factor we call "Education."

Table 7: Five Latent Factors Interpreted as Policy Domains

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Finance	**Military**	**Environment**	**Healthcare**	**Trade**
Panel A: Purest Agencies (Ranked by the Factor Model Coefficients i.e. Coefficients on Lobbyist Utility)				
0.86 (0.03) - Federal Reserve System	0.82 (0.04) - U.S. Marines	0.72 (0.04) - Bureau of Land Management (BLM)	0.84 (0.03) - Health Resources & Services Administration (HRSA)	0.54 (0.47) - U.S. International Trade Commission (ITC)
0.86 (0.02) - Equal Employment Opportunity Commission (EEOC)	0.77 (0.04) - Navy	0.72 (0.04) - U.S. Fish & Wildlife Service (USFWS)	0.83 (0.02) - Agency for Healthcare Research & Quality (AHRQ)	0.53 (0.53) - Pipeline & Hazardous Materials Safety Administration
0.81 (0.06) - Office of the Comptroller of the Currency (OCC)	0.71 (0.05) - Air Force	0.70 (0.02) - Bureau of Reclamation	0.79 (0.02) - Centers For Medicare and Medicaid Services (CMS)	0.52 (0.54) - Surface Transportation Board (STB)
0.81 (0.02) - Securities & Exchange Commission (SEC)	0.70 (0.09) - U.S. Immigration & Customs Enforcement (ICE)	0.69 (0.03) - U.S. Forest Service	0.75 (0.02) - Medicare Payment Advisory Commission (MedPAC)	0.51 (0.43) - Defense Logistics Agency
0.80 (0.03) - Federal Deposit Insurance Corporation (FDIC)	0.69 (0.07) - Natl Security Agency (NSA)	0.65 (0.04) - Natl Park Service (NPS)	0.69 (0.03) - Substance Abuse & Mental Health Services Administration (SAMHSA)	0.51 (0.46) - Intl Trade Administration (ITA)
Panel B: Most Influential Agencies (Ranked by the Share of Unit Lobbyists Predicted to Lobby)				
0.61 (0.03) - Treasury	0.37 (0.02) - Defense (DOD)	0.42 (0.03) - Environmental Protection Agency (EPA)	0.59 (0.01) - Health & Human Services (HHS)	0.27 (0.10) - U.S. Trade Representative (USTR)
0.43 (0.02) - Commerce (DOC)	0.31 (0.02) - State (DOS)	0.32 (0.01) - Transportation (DOT)	0.39 (0.01) - Centers For Medicare and Medicaid Services (CMS)	0.25 (0.10) - Environmental Protection Agency (EPA)
0.39 (0.02) - Health & Human Services (HHS)	0.30 (0.01) - Commerce (DOC)	0.30 (0.01) - Agriculture (USDA)	0.27 (0.01) - Office of Management & Budget (OMB)	0.25 (0.05) - Commerce (DOC)
0.38 (0.01) - Office of Management & Budget (OMB)	0.27 (0.03) - Homeland Security (DHS)	0.29 (0.01) - Energy	0.26 (0.01) - Food & Drug Administration (FDA)	0.18 (0.03) - State (DOS)
0.37 (0.02) - Homeland Security (DHS)	0.22 (0.01) - Health & Human Services (HHS)	0.26 (0.01) - Commerce (DOC)	0.17 (0.01) - Agriculture (USDA)	0.17 (0.04) - Energy
HHI = 0.0361 (0.0004)	0.0405 (0.0010)	0.0421 (0.0005)	0.0515 (0.0010)	0.0441 (0.0014)
Panel C: Top Linked Issues				
12,005 - Financial Institutions/Investments/Securities	7,782 - Defense	6,053 - Natural Resources	23,506 - Health Issues	3,087 - Trade (domestic/foreign)
10,564 - Taxation	4,719 - Budget/Appropriations	5,816 - Budget/Appropriations	23,113 - Medicare/Medicaid	2,351 - Taxation
7,496 - Banking	2,930 - Homeland Security	3,002 - Clean Air and Water	10,938 - Budget/Appropriations	1,658 - Transportation
5,695 - Trade (domestic/foreign)	2,098 - Taxation	2,704 - Energy/Nuclear	5,510 - Taxation	1,651 - Budget/Appropriations
4,817 - Health Issues	1,909 - Health Issues	2,573 - Taxation	3,539 - Pharmacy	1,594 - Energy/Nuclear

Notes: Table displays information related to the five factors obtained using methods discussed in the text. At the top of each column, we write what we call the factor, which is based on our assessment of the information provided in panel A. Panel A lists the top agencies by the estimated factor loading, which we call the "purest" agencies. Throughout this table, estimates are shown with bootstrapped standard errors in parentheses. Panel B verifies our analysis by listing the most common issues linked to the agencies listed in panel A; specifically, we rank issues by the number of issue-agency combinations in all reports and list the highest ranked issues (we report the number of issue-agency combinations). We find that about half (50.3%) of the variation (weighting agencies equally) is explained by five factors.

Table 8: Four Latent Factors Interpreted as Policy Domains

Factor 1	Factor 2	Factor 3	Factor 4
Finance	**Healthcare**	**Environment**	**Military**
Panel A: Purest Agencies (Ranked by the Factor Model Coefficients i.e. Coefficients on Lobbyist Utility)			
0.85 (0.01) - Equal Employment Opportunity Commission (EEOC)	0.85 (0.02) - Health Resources & Services Administration (HRSA)	0.73 (0.04) - Bureau of Justice Assistance	0.73 (0.05) - Navy
0.83 (0.01) - Natl Labor Relations Board (NLRB)	0.78 (0.03) - Social Security Administration (SSA)	0.72 (0.03) - Natl Park Service (NPS)	0.68 (0.03) - U.S. Coast Guard (USCG)
0.81 (0.02) - Federal Reserve System	0.74 (0.05) - Agency for Healthcare Research & Quality (AHRQ)	0.69 (0.03) - U.S. Forest Service	0.67 (0.07) - U.S. Marines
0.78 (0.01) - Securities & Exchange Commission (SEC)	0.73 (0.04) - Centers For Medicare and Medicaid Services (CMS)	0.68 (0.03) - Bureau of Reclamation	0.67 (0.06) - Air Force
0.77 (0.01) - U.S. Copyright Office	0.72 (0.03) - Administration on Aging	0.67 (0.06) - Bureau of Land Management (BLM)	0.66 (0.05) - Office of the Secretary of Defense
Panel B: Most Influential Agencies (Ranked by the Share of Unit Lobbyists Predicted to Lobby)			
0.59 (0.01) - Treasury	0.57 (0.02) - Health & Human Services (HHS)	0.35 (0.03) - Environmental Protection Agency (EPA)	0.31 (0.02) - Defense (DOD)
0.47 (0.01) - Commerce (DOC)	0.36 (0.02) - Centers For Medicare and Medicaid Services (CMS)	0.29 (0.01) - Agriculture (USDA)	0.29 (0.01) - Commerce (DOC)
0.40 (0.01) - Homeland Security (DHS)	0.24 (0.01) - Office of Management & Budget (OMB)	0.28 (0.01) - Transportation (DOT)	0.28 (0.01) - State (DOS)
0.40 (0.01) - Office of Management & Budget (OMB)	0.24 (0.01) - Food & Drug Administration (FDA)	0.25 (0.02) - Energy	0.26 (0.02) - Health & Human Services (HHS)
0.39 (0.01) - Health & Human Services (HHS)	0.22 (0.01) - Treasury	0.22 (0.01) - Interior (DOI)	0.22 (0.01) - Office of Management & Budget (OMB)
HHI = 0.0331 (0.0002)	0.0457 (0.0013)	0.0367 (0.0008)	0.0358 (0.0005)
Panel C: Top Linked Issues			
9,616 - Financial Institutions/Investments/Securities	22,097 - Health Issues	4,945 - Natural Resources	8,573 - Defense
9,502 - Taxation	21,767 - Medicare/Medicaid	4,633 - Budget/Appropriations	4,928 - Budget/Appropriations
6,669 - Trade (domestic/foreign)	10,687 - Budget/Appropriations	2,336 - Clean Air and Water	3,018 - Homeland Security
5,151 - Labor Issues/Antitrust/Workplace	5,665 - Taxation	1,955 - Energy/Nuclear	2,371 - Taxation
4,962 - Banking	3,472 - Pharmacy	1,840 - Taxation	1,750 - Health Issues

Notes: See notes to Table 7. This table repeats the analysis using four factors rather than five. We find that 43.0% of the variance is explained by four factors.

Table 9: Six Latent Factors Interpreted as Policy Domains

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Finance	**Environment**	**Healthcare**	**Military**	**Trade**	**Education**
Panel A: Purest Agencies (Ranked by the Factor Model Coefficients i.e. Coefficients on Lobbyist Utility)					
0.89 (0.07) - Federal Reserve System	0.72 (0.03) - Bureau of Land Management (BLM)	0.84 (0.02) - Health Resources & Services Administration (HRSA)	0.90 (0.05) - Navy	0.76 (0.31) - Intl Trade Administration (ITA)	0.71 (0.48) - Office of Juvenile Justice & Delinquency Prevention
0.85 (0.07) - Office of the Comptroller of the Currency (OCC)	0.70 (0.04) - U.S. Fish & Wildlife Service (USFWS)	0.82 (0.02) - Agency for Healthcare Research & Quality (AHRQ)	0.89 (0.07) - Air Force	0.67 (0.25) - U.S. International Trade Commission (ITC)	0.61 (0.52) - Institute of Museum and Library Services (IMLS)
0.84 (0.03) - Equal Employment Opportunity Commission (EEOC)	0.69 (0.04) - Bureau of Reclamation	0.79 (0.02) - Centers For Medicare and Medicaid Services (CMS)	0.88 (0.05) - U.S. Marines	0.62 (0.25) - Patent & Trademark Office (PTO)	0.61 (0.50) - Natl Endowment for the Humanities
0.82 (0.06) - Securities & Exchange Commission (SEC)	0.68 (0.03) - U.S. Forest Service	0.73 (0.02) - Medicare Payment Advisory Commission (MedPAC)	0.77 (0.09) - Office of the Secretary of Defense	0.60 (0.20) - U.S. Trade Representative (USTR)	0.58 (0.40) - Natl Archives & Records Administration (NARA)
0.82 (0.06) - Federal Deposit Insurance Corporation (FDIC)	0.65 (0.05) - Bureau of Justice Assistance	0.73 (0.02) - Substance Abuse & Mental Health Services Administration (SAMHSA)	0.74 (0.03) - U.S. Coast Guard (USCG)	0.54 (0.26) - Library of Congress (LOC)	0.53 (0.40) - Corporation for Natl & Community Service
Panel B: Most Influential Agencies (Ranked by the Share of Unit Lobbyists Predicted to Lobby)					
0.61 (0.03) - Treasury	0.41 (0.02) - Environmental Protection Agency (EPA)	0.60 (0.01) - Health & Human Services (HHS)	0.39 (0.02) - Defense (DOD)	0.37 (0.08) - Commerce (DOC)	0.22 (0.04) - Health & Human Services (HHS)
0.40 (0.04) - Commerce (DOC)	0.33 (0.01) - Transportation (DOT)	0.39 (0.01) - Centers For Medicare and Medicaid Services (CMS)	0.24 (0.02) - Homeland Security (DHS)	0.35 (0.08) - U.S. Trade Representative (USTR)	0.22 (0.07) - State (DOS)
0.39 (0.02) - Health & Human Services (HHS)	0.29 (0.01) - Energy	0.27 (0.01) - Office of Management & Budget (OMB)	0.23 (0.02) - Commerce (DOC)	0.30 (0.07) - State (DOS)	0.20 (0.06) - Treasury
0.37 (0.02) - Office of Management & Budget (OMB)	0.29 (0.01) - Agriculture (USDA)	0.25 (0.01) - Food & Drug Administration (FDA)	0.22 (0.03) - State (DOS)	0.30 (0.06) - Environmental Protection Agency (EPA)	0.20 (0.06) - Commerce (DOC)
0.36 (0.02) - Labor (DOL)	0.25 (0.01) - Interior (DOI)	0.17 (0.01) - Agriculture (USDA)	0.21 (0.01) - Office of Management & Budget (OMB)	0.22 (0.04) - Health & Human Services (HHS)	0.18 (0.03) - Agriculture (USDA)
HHI = 0.0394 (0.0008)	0.0451 (0.0008)	0.0570 (0.0014)	0.0463 (0.0012)	0.0494 (0.0018)	0.0438 (0.0017)
Panel C: Top Linked Issues					
12,005 - Financial Institutions/Investments/Securities	5,563 - Natural Resources	23,506 - Health Issues	8,573 - Defense	18,363 - Trade (domestic/foreign)	735 - Budget/Appropriations
10,564 - Taxation	5,364 - Budget/Appropriations	23,113 - Medicare/Medicaid	4,928 - Budget/Appropriations	10,072 - Taxation	450 - Education
7,496 - Banking	2,925 - Clean Air and Water	10,938 - Budget/Appropriations	3,018 - Homeland Security	6,721 - Copyright/Patent/Trademark	217 - Civil Rights/Civil Liberties
5,695 - Trade (domestic/foreign)	2,494 - Energy/Nuclear	5,510 - Taxation	2,371 - Taxation	6,597 - Budget/Appropriations	165 - Health Issues
4,817 - Health Issues	2,313 - Taxation	3,539 - Pharmacy	1,750 - Health Issues	5,545 - Health Issues	164 - Taxation

Notes: See notes to Table 7. This table repeats the analysis using six factors rather than five. We find that 54.7% of the variance is explained by six factors.

A.3 Predicting Agency Lobbying with Factors

With the estimated factor loadings and the tetrachoric correlation matrix (both described previously), we calculate predicted factors for lobbyist-cycles i using the method of Thomson (1951) (see also the detailed discussion in Grice, 2001),

$$\hat{F}_{ik} = \sum_a w_{ak} X_{ai}$$

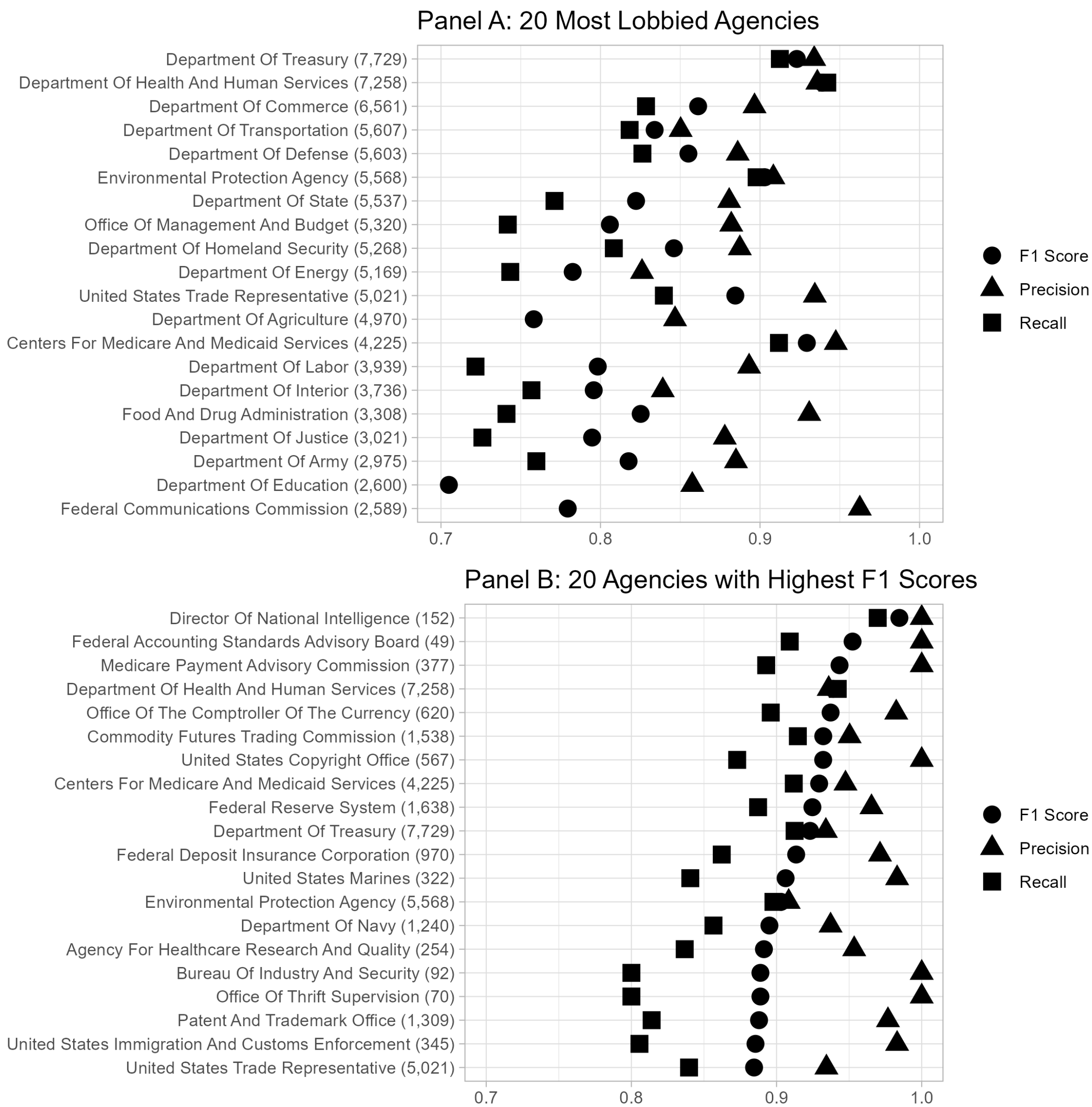
where the w_{ak} are regression coefficients ("weights") calculated as the product of the (transposed) factor loadings matrix and the inverse of the correlation matrix for observed factors (i.e., the latter is the tetrachoric correlation matrix that is the raw input to the factor analysis model). The interpretation is that $\hat{F}_{ik}$ is the predicted value of factor k for lobbyist-cycle i, obtained as-if we ran a regression of those factors on the observable lobbyist choices and then took the predicted values (the weights, w_{ak}, can thus be interpreted as regression coefficients).

We then use these lobbyist-cycle factor scores (predicted factors) $\hat{F}_{ik}$ to predict agency choices, for each agency a. To this end, we run random forest models identical to those in Section 3, predicting the indicator variable of agency choice for each agency using the factor scores, and reporting the F1 scores for this prediction. We focus on our preferred $K = 5$ factor model for this exercise. In order to ensure that measured differences in predictive accuracy are driven by differences in methods, and not differences in samples, we restrict the analysis in this subsection to the lobbyist-cycles with partisanship measures based on campaign finance records, i.e., the exact same lobbyist-cycles that were used in Section 3.

Figure 6 displays the results from this analysis. This figure is analogous to Figure 2, except using the five factor model lobbyist-cycle-level factor scores to predict agency choice, rather than lobbyist-cycle level partisanship. Panel A displays measures of predictive quality for the 20 most lobbied agencies, which are all quite high (above 0.80). Panel

B displays the agencies with the highest F1 scores. Overall, the average F1 score in this case, weighting agencies by their number of lobbyist-cycles, is 0.81. (Weighting agencies equally, it is 0.73.) We can compare these to F1 score of 0.04 (weighted; 0.01 unweighted) from the lobbyist partisanship model (in Section 3 of the main text). We conclude that a five-factor model is an astronomical improvement over the lobbyist partisanship model, in terms of explaining lobbyist agency choice. Since, as we have argued, the factors we identified look very much like policy domains, we argue that it is most likely substantive policy areas, and not partisanship, that drive agency venue choice.

Figure 6: Predicting Whether a Lobbyist-Cycle Lobbies an Agency Using Lobbyist-Cycle Factor Scores (Five Factor Model)



Notes: The number of lobbyist-cycles in the analytic sample (including both training and testing subsamples) for each agency is displayed in parentheses after the agency name.

A.4 Factors and Agency Ideology

Table 10: Associations Between Agency Liberal-Conservative Ideal Points and Lobbying-Inferred Policy Domains

	(1)	(2)	(3)	(4)
Dep. Var. =	CJ, Purity	CJ, Influence	RCL, Purity	RCL, Influence
Finance	-0.001	-0.020	-0.122	-0.010
	(0.020)	(0.034)	(0.092)	(0.193)
Military	0.051^{**}	0.104^{***}	0.389^{***}	0.565^{***}
	(0.024)	(0.025)	(0.083)	(0.170)
Environment	0.035^{**}	0.061^{*}	-0.276^{***}	-0.322^{**}
	(0.018)	(0.032)	(0.085)	(0.131)
Healthcare	-0.027	-0.030^{**}	-0.368^{***}	-0.328^{***}
	(0.017)	(0.012)	(0.079)	(0.069)
Trade	-0.009	-0.105^{***}	-0.018	-0.073
	(0.017)	(0.034)	(0.076)	(0.152)
Constant	-0.211^{***}	-0.211^{***}	-0.018	-0.018
	(0.016)	(0.017)	(0.087)	(0.096)
Observations	47	47	96	96
R^2	0.309	0.254	0.294	0.137
Adjusted R^2	0.225	0.163	0.255	0.089

Notes: Coefficients are reported from linear regression models relating agency ideal points to our lobbying-inferred policy domains. Columns (1) and (2) use ideal points from Chen and Johnson (2015), while columns (3) and (4) use ideal points from Richardson, Clinton and Lewis (2018). Columns (1) and (3) measure the independent variables using our purity measures of agency policy domains, while columns (2) and (4) use our influence measures instead. Note that the independent variables are demeaned in these models, so that the constant is interpretable as the grand mean of the dependent variable in the sample. Heteroskedasticity-consistent standard errors reported in parentheses; *p<0.1; **p<0.05; ***p<0.01.

B Lobbying Forms

Figure 7: Example LD-203 Lobbying Form

Filer Type and Name

Type:
☐ Organization ☑ Lobbyist

Lobbyist Name:
Ms. Katherine Heil Hamilton

Employer Name:
38 North Solutions, LLC

Identification Numbers

House Registrant ID:
42071

Senate Registrant ID:
400944730

Reporting Period

Year:
2016
☐ Mid-Year (January 1 - June 30)
☑ Year-End (July 1 - December 31)
☐ Amendment

Political Action Committee Names

Contributions

☐ No Contributions

#1.

Contribution Type:	Contributor Name:	Amount:	Date:
FECA	Self	$1,000.00	09/06/2016
Payee:	**Honoree:**		
Katie McGinty for Senate	Katie McGinty for Senate		

#2.

Contribution Type:	Contributor Name:	Amount:	Date:
FECA	Self	$1,000.00	09/16/2016
Payee:	**Honoree:**		
Hillary Victory Fund	Hillary Victory Fund		

Comments

Figure 8: Example L-2 Lobbying Form

Lobbying Disclosure Act of 1995 (Section 5) - **All Filers Are Required to Complete This Page**

1. Registrant Name ☑ Organization/Lobbying Firm ☐ Self Employed Individual
38 North Solutions, LLC

2. Address
Address1 1133 15th St, NW   Address2 12th floor
City Washington   State DC   Zip Code 20005   Country USA

3. Principal place of business (if different than line 2)
City ______ State ______ Zip Code ______ Country ______

4a. Contact Name Ms. Katherine Hamilton   b. Telephone Number 2025248832   c. E-mail katherine@38northsolutions.com   5. Senate ID# 400944730-176

7. Client Name ☐ *Self* ☐ *Check if client is a state or local government or instrumentality*
Tesla Motors, Inc.   6. House ID# 420710017

TYPE OF REPORT   8. Year 2016   Q1 (1/1 - 3/31) ☐   Q2 (4/1 - 6/30) ☐   Q3 (7/1 - 9/30) ☑   Q4 (10/1 - 12/31) ☐

9. Check if this filing amends a previously filed version of this report ☐

10. Check if this is a Termination Report ☐   Termination Date ______   11. No Lobbying Issue Activity ☐

INCOME OR EXPENSES - YOU MUST complete either Line 12 or Line 13

12. Lobbying	**13. Organizations**
INCOME relating to lobbying activities for this reporting period was:	**EXPENSE** relating to lobbying activities for this reporting period were:
Less than $5,000 ☐	Less than $5,000 ☐
$5,000 or more ☑ $ 25,000.00	$5,000 or more ☐ $ ______
Provide a good faith estimate, rounded to the nearest $10,000, of all lobbying related income for the client (including all payments to the registrant by any other entity for lobbying activities on behalf of the client).	**14. REPORTING** Check box to indicate expense accounting method. See instructions for description of options. ☐ **Method A.** Reporting amounts using LDA definitions only ☐ **Method B.** Reporting amounts under section 6033(b)(8) of the Internal Revenue Code ☐ **Method C.** Reporting amounts under section 162(e) of the Internal Revenue Code

Signature Digitally Signed By: Katherine Hamilton   **Date** 10/19/2016 1:27:53 PM

LOBBYING ACTIVITY. Select as many codes as necessary to reflect the general issue areas in which the registrant engaged in lobbying on behalf of the client during the reporting period. Using a separate page for each code, provide information as requested. Add additional page(s) as needed.

15. General issue area code ENG

16. Specific lobbying issues

Energy storage provisions in House and Senate energy bills, Clean Power Plan. S. 2012, H.R. 8

17. House(s) of Congress and Federal agencies ☐ Check if None

U.S. SENATE, U.S. HOUSE OF REPRESENTATIVES, Environmental Protection Agency (EPA)

18. Name of each individual who acted as a lobbyist in this issue area

First Name	Last Name	Suffix	Covered Official Position (if applicable)	New
Katherine	Hamilton			☐
Isaac	Brown		Legislative Director, Rep. Jan Schakowsky	☐

19. Interest of each foreign entity in the specific issues listed on line 16 above ☑ Check if None

LOBBYING ACTIVITY. Select as many codes as necessary to reflect the general issue areas in which the registrant engaged in lobbying on behalf of the client during the reporting period. Using a separate page for each code, provide information as requested. Add additional page(s) as needed.

15. General issue area code TAX

16. Specific lobbying issues

Investment tax credit for energy storage (H.R. 5350); EV tax credit

17. House(s) of Congress and Federal agencies ☐ Check if None

U.S. SENATE, U.S. HOUSE OF REPRESENTATIVES

18. Name of each individual who acted as a lobbyist in this issue area

First Name	Last Name	Suffix	Covered Official Position (if applicable)	New
Katherine	Hamilton			☐
Isaac	Brown		LD Rep. Schakowsky	☐