Does Access Mean Success? Connection to Policy-Makers and Lobbying Success of Political Actors*

Rosanne Logeart[†]

Preliminary Draft - Most recent version here

December 2024

Abstract

This article investigates the relationship between access to policymakers and lobbying success. I collect unique, large-scale textual data to capture the content of lobbying activities and subsequent changes in EU regulations. I identify lobbying success with both plagiarism detection and a large language model (GPT-4). I measure direct access to policymakers based on meetings between policymakers and interest group representatives. To address potential endogeneity in access, I build a matched sample using propensity scores and a balanced sample using Mahalanobis distance. I find that access to policymakers is associated with a larger likelihood of lobbying success, consistent across success measures and samples. This effect is stronger for organizations with more meetings and those with access to higher-level officials. Finally, I exploit turnover in policymakers across mandates to explore the underlying mechanisms and find evidence that political connections primarily drive the effect of access on success, surpassing the influence of information transmission, institutional knowledge, or intrinsic quality of the organization in driving lobbying success.

Keywords: Political economy, Lobbying, Advocacy, Interest groups, European Union

JEL codes: D72, P48

^{*}Email: rosanne.logeart@psemail.eu. I thank Mireille Chiroleu-Assouline, Matilde Bombardini, Frederico Finan, Teresa Esteban-Casanelles, Roberto Galbiati, Brad Hackinen, Emeric Henry, Florian Hollenbach, Katrin Millock, Fabien Prieur, Cailin Slattery, Francesco Trebbi and participants of seminars at UC Berkeley, Sciences Po, PSE, Dauphine, Nanterre University and Monash/Warwick, participants of the 2024 EAYE, the 2024 DevPEC, the 2024 EPCS, the CESifo Workshop on Political Economy, EuroWEPS, the 2023 FAERE conference, the PSE-CEPR policy forum, and the CBS Money in Politics conference for helpful comments and suggestions. This work has been funded by a French government subsidy managed by the Agence Nationale de la Recherche under the framework of the Investissements d'avenir programme reference ANR-17-EURE-001.

[†]Copenhagen Business School, Paris School of Economics & Paris 1 Panthéon-Sorbonne University.

1 Introduction

Optimal policies are extensively studied in the literature and yet not necessarily implemented. Understanding the functioning of the policy-making process is crucial for the implementation of optimal policies, ranging from climate to trade or urban policies. The policy-making process is inherently marked by an asymmetry of information, as policy-makers often lack the necessary information to design relevant policies. This deficit is typically addressed through interactions with interest groups, where interest groups exchange information against access to the policy-making process (Chalmers, 2013). This process is known as informational lobbying. However, access to policy-makers is costly, and bias in the information collected can arise. Acknowledging this selection bias, institutions have called attention to "the risk of the policy-makers just listening to one side of the argument or of particular groups getting privileged access on the basis of sectoral interests or nationality, which is a clear weakness with the current method of ad hoc consultations" (European Commission, 2001).

To address accountability issues and mitigate corruption concerns, institutions have introduced measures to open up the policy-making process. Online platforms have been established to collect input from organizations and citizens, aimed at enhancing transparency and accountability. This additional layer for information gathering offers less costly means of influence, reducing selection bias by providing entry points for new actors and potentially bringing new information. However, one could argue that if policy-makers are efficient at selecting who they meet with and gather information from, information from online consultations is not relevant to them, given that lobbying also happens at other stages. Furthermore, assuming that aggregating received information incurs costs for policy-makers, it is not clear whether policy-makers would account for online comments and change policy content. They might want to streamline the process by prioritizing comments from entities they are connected to and trust, whether it is because these entities have better information to offer, or because they are favored by their political connections and reputation. This paper analyzes these dynamics between access to policy-makers and lobbying success, defined as the ability to influence policy through these online comments.

In this paper, I study the relationship between access to policy-makers and lobbying success. I provide a novel database and new evidence on lobbying activities directed at the European Commission, specifically focusing on Commission regulations, delegated regulations, and implementing regulations. The choice of Commission regulations as the focal point for this study is strategic for several reasons. Firstly, these regulations are subject to public comments, providing an opportunity to observe the substance of lobbying activities. Secondly, the mandatory disclosure of an internally validated complete draft facilitates a

detailed observation of the evolution of the text. By comparing the draft and the adopted text, I can observe the position of the policymaker at two points in time - before and after receiving comments. This dual perspective serves as a reliable measure of policy changes. This distinguishing feature of systematic draft publication sets Commission regulations apart from other policy-making processes involving comments, such as those in the United States rule-making. The combination of these two features enables the identification of lobbying success, defined by instances where policy changes align with the requests made in the comments. Moreover, unlike the United States, where there is a legal obligation for agencies to respond to received comments, the European Commission faces no such obligation. This absence of risk of a lawsuit for not addressing a comment provides a unique lens to uncover the nuanced mechanisms influencing the decision to account for a comment or not. Additionally, the regulations I study are adopted by the European Commission, without going through the European Parliament or the Council of the EU. This feature allows me to observe the entire regulatory process, including meetings between members of the European Commission and interest representatives. Lastly, EU regulations carry significant weight by regulating critical areas that touch the lives of every citizen, ranging from determining CO2 emissions allowed per type of vehicle to dictating permissible chemicals and their quantities, as well as establishing energy efficiency requirements for household appliances.

To explore the relationship between access to policy-makers and lobbying success, I provide a new database on the Commission regulations falling under the draft publication and comment feature of the European Union (European Commission, 2015). It means that the internally-approved draft of these regulations must undergo public disclosure, and be subject to a 4-week commenting period. During this period, individuals and entities have the opportunity to submit online comments on this draft. Comments can be written in any of the 24 official EU languages, reducing further entry costs. The text may be modified after evaluation of these comments, and then the Commission adopts the final text. I compile information on 482 such Commission initiatives. For each of them, I gather the draft regulation along with its annexes, the received comments together with their respective attachments and information on respondents, and the adopted regulation and its annexes. The extensive dataset encompasses a total of 129,153 comments. To identify substantive policy changes, I isolate the legally binding parts of texts and analyze modifications between the draft and the adopted version. I then identify instances of lobbying success by comparing these policy changes with the received comments, using a method based on text reuse. Text reuse allows for a transparent and tractable assessment of success while offering little flexibility. I complement this measure with a more flexible but black-box method, using a large language model (GPT-4). I also assemble data on meetings conducted between European Commission members or Directors-General and interest representatives to measure access to policy-makers. My comprehensive dataset spanning the years 2014 to 2023 includes details on 36,101 meetings. Leveraging the high temporal precision of this data (day-level), I construct a measure of access to policy-makers based on the frequency and timing of these meetings. I combine the data on success together with the data on access using a name-matching algorithm. It enables me to explore the relationship between direct access to policy-makers and the ability of lobbying efforts to initiate policy changes. I use propensity scores and Mahalanis distance to create respectively a matched and a balanced samples of comments written by an entity having or not having access to policy-makers. Balancing is done on comment quality and type of entity writing, within each regulation. I use these samples to estimate the effect of direct access to policy-makers on lobbying success probability. Lastly, I exploit a change of Commission to distinguish four possible channels explaining this effect: information-transmission, political connection, insider knowledge, and better intrinsic quality.

I establish four main sets of results. First, I provide new insights into the landscape of lobbying activities targeting the European Commission. The business sector emerges as the predominant contributor during these commenting periods, in addition to constituting over 50% of direct meetings with Commission members. Following business associations and companies, NGOs emerge as the third major actor, while the other entities play a minor role.

Second, a limited number of organizations engaging in commenting activities also have access to policy-makers. Conversely, 14% of the organizations having access to policy-makers also write comments. These organizations have on average three times more meetings than the whole sample. While one may thing that comments are only used by organizations not finding another way to influence policy-makers, I find that organizations having the most access to the policy-making process use this feature.

Third, I find that among similar comments, comments written by an entity having access to policy-makers are 25% more likely to be successful than comments written by entities without such access. This effect is consistent across both the plagiarism detection and the large language model measures. This effect is driven by access to highest level of the Commission hierarchy, the Presidency; or the person in charge of specific policy areas, the Directors-General. Additionally, entities with more access, characterized by numerous meetings with policy-makers, emerge as the most likely to influence and shape regulatory text changes.

Fourth, I provide suggestive evidence that active political connections play an important role in lobbying successfully, surpassing the effects of information transmission, institutional knowledge and intrinsic quality of the organization writing the comment.

This paper first contributes to the literature on the influence of lobbying on policies (Baumgartner et al., 2009; Klüver, 2012; Mahoney, 2007). In the EU, Anger et al. (2015,

2016); Burghaus et al. (2019) find that lobbying expenditures affect the energy prices a sector faces and the allocation of free allowances under the EU Emission Trading Scheme the different sectors obtain. In the US, Meng and Rode (2019) find an asymmetry of efficiency of lobbying: interest groups lobbying for a pro-environmental measure are less efficient than interest groups lobbying against it. Bertrand et al. (2021) use a similar feature of comments in the US rule-making process and find that comments of corporations giving money to nonprofits commenting on the same text are closer to the adopted regulation. The US setting does not offer systematic access to a drafted regulation before comments, which prevents them from controlling for the proximity to the initial position of the policy-maker.¹ To my knowledge, this paper is the first to measure systematically the evolution of policy-makers' positions. This is done thanks to the unique setting offering draft versions of regulations and the use of text as data. It is also one of the few papers measuring lobbying by its content rather than its related expenditure.

Second, I contribute to the literature on political connections (Brown and Huang, 2020; Bertrand et al., 2014; Ferguson and Voth, 2008; Fisman, 2001). I measure access through direct meetings with policy-makers, and I exploit the temporal granularity of the data (day-level) to distinguish access over time.

Third, this paper closely relates to the literature on the determinants of lobbying success in EU consultations (Klüver, 2013). Bunea (2013) shows business groups have a higher degree of lobbying success relatively to other stakeholders in five environmental consultations. Studying nine environmental consultations, Hermansson (2016) highlights the importance of expertise and finds business organizations are more likely to be successful even in the absence of expertise on a specific policy issue. In contrast, Dür et al. (2015) finds that citizen groups are more likely to be successful. Lee et al. (2024) analyze the role of lobbying in Canadian front-of-pack labelling regulations. In this paper, I study the Commission feedback procedure on 482 initiatives and I show that the relatively higher success rate of business groups is driven by business groups having direct access to policy-makers.

In the following section, I describe in more detail the context of this study and the data used. I then analyze the relationship between chances of lobbying success and access to policy-makers in Section 3. The mechanisms and channels at play are studied in Section 4. Section 5 concludes.

¹See Bombardini and Trebbi (2020) for a more complete review of empirical studies on lobbying.

2 Context and Data

2.1 EU Policy Making

The European Union policy-making process provides a context in which one can observe the position of the policymaker at several points in time, as well as the presence and content of lobbying from different entities in between.

The draft publication and comment process I study was introduced in 2015, as a part of the European Commission's strategy to restore citizens' and businesses' confidence in its "ability to deliver" (European Commission, 2015). It applies to the acts adopted by the Commission without going through the vote of the European Parliament or of the Council of the EU. The Commission is the executive body of the EU, it is proposed and appointed by the European Council, and validated by the Parliament. It is not directly accountable to voters (Besley and Coate, 2003) and can be seen as not legitimate. The Commission seeks to increase transparency, legitimacy, and accountability of its work by inviting inputs from civil society and "by the active engagement of civil society" (European Commission, 2017). This process of draft publication and comments became mandatory in November 2021 (European Commission, 2021b). It applies to delegated acts, implementing acts, and legislative acts under the regulatory procedure with scrutiny (RPS). Acts can be decisions, directives, or regulations. In this paper, I focus on regulations as they are legal acts of general application, and are binding automatically and uniformly to all EU countries. It makes the European Commission a relevant target for lobbying on these issues.

The draft publication and comments process works as follows. Drafts are made publicly accessible online. The draft publication can only be done once per act and is irreversible. It thus has to go through interservice consultation first and be approved by the hierarchy (European Commission, 2021a). Given the level of validation and the public and official nature of these drafts, they act as counterfactual of the text that would have been adopted without the commenting process. Drafts are then open to comments for four weeks. Any individual or entity can write a comment, in any of the 24 official EU languages. Comments are limited to 4,000 characters and can include one attachment. The comments are assessed and the final text is adopted after potential modifications.

²Delegated acts are acts that add or amend aspects of existing laws.

³Implementing acts set out rules to make sure Member States implement EU legislation in the same way, such as measurement standards.

⁴RPS existed between 2006 and 2009 for adopting EU secondary legislation. It is aimed to be phased out but is still used today as hundreds of basic legislative acts still provide for this procedure.

2.2 Data

Drafts, Comments and Adopted Texts. I collect delegated regulations, implementing regulations and regulations under the regulatory procedure with scrutiny of the European Commission. I restrict my sample to the period 2015-2023 and to initiatives that are subject to the draft publication and comments process, which amounts to 482 initiatives. I gather the official draft, the comments received, and the adopted text for each of initiative. In total, there are 129,153 comments. I translate all comments and related attachments into English. I also gather information on which Directorate-General is in charge of each initiative, the exact date of each comment, and information on entities writing comments.

Information on entities includes the user type and country of origin, as well as the organization name and size when relevant.⁵ The type is selected by the user among "academic/research institution", "anonymous", "business association", "company/business organisation", "consumer organisation", "environmental organisation", "EU citizen", "non-EU citizen", "non-governmental organisation", "public authority", "trade union", and "other". I group the EU citizen and the non-EU citizen categories into a "citizen" category and I create an EU indicator variable based on the declared country of origin.

Meetings with Commission members. I collect the universe of meetings held between interest representatives and members of the European Commission or Directors-General. I gather 39,098 meetings for the 2014-2023 period, with information on attendees and the exact date of the meeting. Several organizations sometimes have meetings together. Considering meetings at the organization level, I obtain 45,047 observations. I use Transparency Register IDs and name matching to match organizations having meetings and organizations writing comments. The matching procedure of comments and meeting data is described in Appendix C.

I find that 14% of the organizations having meetings with the European Commission also write comments on the regulation I study (1,023 out of 7,525). Only a minority of entities having access to policy-makers also write comments. While this could indicate that organizations having direct access to policy-makers do not need to write comment to lobby as they have other means of lobbying, I find that entities commenting have on average more meetings with the European Commission. Taking all organizations having meetings between 2014 and 2023, I find an average of 6 meetings per organization over the period, with a median at 2 meetings. Restricting the sample to organizations also writing comments on

⁵I do not use the size variable as it is based on the number of employees, and is ill-suited to measure the size of business associations, consumer organisations or NGOs. The number of employees measure leads to classify major actors such as BusinessEurope, the Bureau Européen des Unions de Consommateurs (BEUC) or WWF European Policy Programme as "Small (10 to 49 employees)".

the regulations I study, they have an average of 20 meetings over the period, with a median at 7 meetings per organization (see Table 1). Entities with the highest number of meetings also write comments: the business association Business Europe, the companies Google and Airbus, and the consumer organization Bureau Européen des Unions de Consommateurs all have more than 250 meetings over the whole period and write comments on the regulations I study. This suggests that having meetings and writing comments are complements rather than substitutes, at least for entities lobbying the most.

Table 1: Distribution of meetings

	Organizations	Number of meeting	
		Mean	Median
Meetings	7,525	6.49	2
Meetings & Comments	1,023	19.88	7

Notes. Number of organization, average and median number of meetings with the European Commission per organizations. The second row restricts the sample to organizations having meetings and writing comments on the regulations studied.

Descriptive statistics. My sample consists of 129,153 comments on regulation drafts of the European Commission during the 2015-2023 period. The majority of comments are written by citizens (119,593 comments), followed by the business sector (2,835 from companies and 2,580 by business associations), anonymous entities (1,241), and NGOs (1,141). The full distribution is presented in Table 2. Citizens do not have access to direct meetings with policy-makers, while organizations do. Overall, 2.06% of comments are written by an entity with access to policy-makers. This share increases to 32.3% when removing citizens and anonymous entities from the sample, as I cannot assess access for them.

Business associations writing comments have proportionally more access to policy-makers compared to other actors writing comments. 46.40% of comments written by business associations are written by a business association having access to policy-makers at least once before or contemporaneously to writing this comment. Business associations are also historically closer to the European Commission. The Commission sought legitimacy through connection with industries and supported the creation of several business associations in the 1960s. Aiming at gaining further legitimacy and responding to industry-bias critiques, the Commission started to connect with NGOs in the 1980s (Laurens, 2015). While the Commission sought and created connections with business associations, it started to meet NGOs under public pressure. It seems to translate today into relatively more access for business

associations than for NGOs.

Figure 1 displays the distribution of the meetings held between organizations and policy-makers, as well as the monthly distribution of regulations studied. Meetings show a cyclical pattern, with less activity in August and December every year. In addition to these cycles, a downward trend is observed throughout the Juncker Commission mandate. Data cover the entire Juncker Commission (Nov 2014 - Nov 2019) and part of the Von der Leyen Commission (Dec 2019 - Aug 2023). The von der Leyen Commission also displays a spike of meetings in the first months of its mandate. On the contrary, more regulations are adopted towards the end of a mandate, and less at the beginning of a mandate as it takes time. This is coherent with interest groups wanting to meet more with a newly formed Commission to establish a connection, and with the Commission needing time at the beginning of a mandate to elaborate new regulations, and wrapping up ongoing initiatives at the end of a mandate.

Figure 1: Monthly distribution of meetings with policy-makers and regulations studied

Notes. The vertical red dashed line represents the change of Commission. The Juncker Commission mandate lasts from November 2014 to November 2019. The first Von der Leyen Commission starts in December 2019. Black dots and line represent the number of meetings held per month (left-axis). Each red and gray horizontal line represents a regulation studied, they start at the draft publication date and finish at the date of adoption. The red part represents the 4-week commenting period. The right-axis refers to the number of cumulated regulations.

I define lobbying success as the inclusion of a comment into the policy changes happening between the draft and the adoption of a regulation. The main measure I use is based on text reuse. In this case, y_{ir} is a 0/1 categorical variable constructed as follows. Consider first the set of word sequences of a regulatory text r that have been modified between the draft and the adopted text. Consider then the set of word sequences of a comment from entity i on regulation r. If there is any overlap between the set of modified word sequences from the regulatory text r and the set of word sequences from i's comment on r, I set the variable y_{ir} equal to 1. This measure presents the advantage of being transparent and easy to understand, although it offers little flexibility. Alternatively, I use a measure based on the large language model (LLM) provided by OpenAI: GPT-4. The LLM measure is more flexible but less transparent. Using a large language model enables me to get a more refined measure of success. A rapidly growing literature uses GPT in social sciences to deal with text data and overcome the limitations of less nuanced natural language processing tools (Djourelova et al., 2023; Gilardi et al., 2023; Törnberg, 2023). Appendix B.1 details the construction of these two measures.

Table 2 presents the share of successful comments based on these two measures. I find more successful comments when using the plagiarism-detection algorithm: 29% of the comments are identified as successful when excluding citizens and anonymous from the sample, while only 13% comments are successful when measuring with the LLM-based measure. Shared verbatim between a comment and the modified text is more frequent than success assessed by GPT-4. Comparing the three main actors (companies, business associations and NGOs), comments writen by companies are on average less successful than business associations or NGOs' comments. The measure provided by GPT-4 also presents smaller variations between business associations and NGOs success rates.

I read a random sample of successful comments as assessed by GPT-4 to understand the meaning of these changes. Due to the nature of regulations, comments are highly technical. I observe that a large number of comments that are successful are successful in adding exemptions related to their specific sector, especially from companies. Other comments are successful in changing the implementation date of a regulation, adding up to 10 years before having to align with an environmental regulation for instance.

On average, 10.69% of the legally binding part of the draft is modified after the comments. 62 regulations are not modified at all between the draft and the adoption. Among these modifications, I map an average of 10.72% to specific comments with the text reuse method. Appendix Table 10 displays the number of initiatives per Directorate-General, together with the average share of modified text, the average number of comments received, and the share of modifications that can be mapped to at least one specific comment.

Table 2: Distribution of Success and Access per Actor Type

Type	N	Access $(\%)$	Success	(%)
			Plagiarism	LLM
Academic/research Institution	251	7.97	26.29	5.98
Anonymous	1,241	-	15.07	-
Business association	2,580	46.40	35.35	15.04
Citizen	119,593	0	0.06	-
Company/business org.	2,835	27.20	27.90	12.24
Consumer organization	78	35.90	12.82	8.97
Environmental organization	160	25.62	23.75	20.63
NGO	1,141	36.99	24.36	13.50
Other	709	18.19	20.59	8.32
Public authority	452	2.43	33.85	15.49
Trade union	113	38.05	14.16	14.16
Total	129,153	2.06	2.06	-
w/o citizen and anonymous	8,319	32.30	28.97	13.09

Notes. This table displays the number of comments written per type of actor, their access rate, and success rate. Access equals one when the entity writing the comment met with the European Commission at least once before the adoption of the corresponding regulation. Success is based on the plagiarism-detection measure in Column (4) and on GPT-4 in Column (5).

3 Lobbying Success and Access to Policy-Makers

This section investigates the relationship between lobbying success and access to policy-makers. In the first stage, I analyze the determinants of access to account for its endogeneity. I create two balanced samples, the first one is based on propensity score matching and the second relies on Mahalanobis distance matching. I then relate overall access to the European Commission to lobbying success on these balanced samples. To further understand this relationship, I distinguish access to different Commission hierarchical levels, and different levels of intensity of access, as measured by the number of meetings an entity has.

Access to policy-makers. Having direct access to policy-makers is likely related to the political implication of an entity, which in turn may relate to the entity's comment likelihood of success. The simple regression of access on success may thus be upward biased. There may be a selection into having access to policy-makers based on observable characteristics of a comment. I report results from a balancing test in Panel A of Table 4. It shows that entities having access to policy-makers at least once before the adoption of the regulation they write

on tend to (1) write comments that are on average longer, (2) write more complex comments, (3) be more likely to write their comment in one of the main EU languages (English, French or German), and (4) to be more likely to be from the European Union. To alleviate the concerns about the comparability of comments written by entities having access to policy-makers and comments written by entities without access to policy-makers, I use nonparametric balancing methods to create samples of more comparable comments: propensity score matching (PSM) and Mahalanobis distance matching (MDM).

I estimate a logit model at the comment level, with a dummy for the commenting entity having access to policy-makers before the adoption of regulation r. $Access_{ir}$ equals one if entity i writing a comment on regulation r has access to the policy-makers at least once before the adoption of r. The model writes

$$\mathbb{P}(Access_{ir}) = \alpha + \eta X_{ir} + \sum_{t} \gamma_t type_i^t + \delta_r + \varepsilon_{ir}$$
(1)

where $type_i^t$ represents the type of entity i among academic/research institution, business association, company/business organisation, consumer organisation, environmental organisation, non-governmental organisation, other, public authority, and trade union. The type "company/business organisation" is set as the baseline as it is the type of entity writing the most comments (see Table 2). X_{ir} is a set of four control variables. It includes two measures of the comment quality: the comment length, computed as the log of the number of words; and the comment complexity, computed as the average length of words.⁶ It also includes a dummy variable indicating comments written in one of the main EU languages: English, French or German, and a dummy equals to one when the entity is from the EU.⁷ Lastly, δ_r indicates regulation fixed effects. Standard errors are clustered by entity.

Table 3 displays the results from the estimation of the logit model in Equation 1. It shows that longer comments, comments written in English, French or German, and comments written by an EU organization are more likely to be written by an organization having access to policy-makers. Additionally, comments written by business associations are more likely to be written while having access to policy-makers, compared to comments written by companies.

The logit model estimated from Equation 1 correctly classifies 75.53% of the observations. I use the results from this regression to compute the predicted probabilities of having access to policy-makers for each observation, also called propensity scores (Rosenbaum and Rubin, 1983). I use a nearest neighbor matching strategy with a 0.0613 caliper radius without

⁶Appendix B.2 describes the construction of these two measures of comment quality in detail.

⁷The UK is considered within the EU for comments written until January 31st, 2020.

Table 3: Access results

Dependent variable:	Access before adoption
Comment length	.380***
<u> </u>	(.0364)
Comment complexity	.363
- *	(.208)
Main EU languages	1.174***
	(.284)
EU origin	.930***
	(.155)
Type	
Academic/research Institution	-1.686***
	(.348)
Business association	.566***
	(.109)
Consumer organisation	.220
	(.654)
Environmental organisation	165
	(.530)
Non-governmental organisation (NGO)	.315*
	(.154)
Other	0791
	(.177)
Public authority	-3.343***
	(.557)
Trade union	.642*
	(.324)
Regulation FE	Yes
Observations	7,970
Pseudo \mathbb{R}^2	.246

Notes. Results from the first stage logit estimation. Standard errors clustered by entity. The baseline type is "Company/business organisation". Observations at the comment level: Access before adoption equals one when the organization writing a comment on regulation r has access to policy-makers before the adoption of r. * p < 0.05, ** p < 0.01, *** p < 0.001.

replacement and including ties, which for each comment by an entity having access to policy-makers finds the closest comparable comment by an entity not having access to policy-makers within the 0.0613 radius in terms of propensity score.⁸

⁸I calculate the caliper as a quarter of one standard deviation of the propensity score, which is estimated to remove about 90 percent of the bias due to the measured confounders (Rosenbaum and Rubin, 1985). Eliminating more bias would involve the risk of facing the PSM paradox and creating higher imbalances

I report the results from a balancing test performed on the matched sample in Panel B in Table 4. The sample obtained with propensity score matching is balanced over the covariates observed: no statistical difference is left between comments written by entities having or not having access to policy-makers. This matched sample is composed of 3,802 observations, out of the initial 8,319.

Propensity score matching produces a balanced sample, at the cost of reducing the sample size. Other balancing methods overcome this bias by weighting observations, rather than discarding observations. I use a second matching method, the Mahalanobis distance matching (MDM), to create a balanced sample while preserving the richness of the initial sample. Using the Kmatch algorithm by Jann (2017), I match each observation to its five nearest neighbors on the type of entity, the length and complexity of the comment, a dummy for writing in a main EU language and a dummy for being an EU entity. I match within each regulation, with standard errors clustered at the entity level.

Panel C of Table 4 reports the results from a balancing test performed on the MDM sample. The sample obtained is more balanced on the observable covariates than the whole sample (panel A), the observed differences are reduced, although they remain statistically significant. The MDM sample is larger than the PSM sample, but less balanced on the observable covariates. This directly stems from the fact that the PSM sample is a matched sample while the MDM is a balanced one. I present below the analysis performed on both samples.

Access and Success. I relate the lobbying success of a comment written by entity i on regulation r to the overall access measure described in Section 2.2. The variable $Access_{ir}$ equals 1 if i had direct meetings with policy-makers before the adoption of r. I estimate the following model.

$$\mathbb{P}(y_{ir}) = \alpha + \beta Access_{ir} + \sum_{t} \gamma_t type_i^t + \eta X_{ir} + \delta_r + \varepsilon_{ir}$$
 (2)

 y_{ir} is an indicator variable for the success of i's comment on regulation r. $type_i^t$ represents the type of i among academic/research institution, business association, company/business organisation, consumer organisation, environmental organisation, non-governmental organisation, other, public authority, and trade union. "Company/business organisation" is set as the baseline type as it is the type writing the most comments (see Table 2). X_{ir} is a set of four control variables that may affect the likelihood of a comment's success. It includes two measures of the comment quality: the comment length computed as the log of the number

(King and Nielsen, 2019). Computations are done using PSMATCH2 by Leuven and Sianesi (2018).

Table 4: Balancing tests

5,637 (67.76%)	2,682 (32.24%)	
5.459 (1.267)	6.185(1.148)	$.726^{***}(.0289)$
7.136 (.658)	7.296 (.545)	.160*** (.0146)
.874 (.332)	.977 (.149)	.104*** (.00672)
.867 (.340)	.940 (.237)	.0734*** (.00728)
.041 (.198)	.007 (.086)	0335***(.00400)
` ,	` /	.203*** (.0106)
.365 (.481)	` ,	0737*** (.0111)
	` ,	.00157 (.00226)
, ,	` ,	00582(.00322)
` '	` ,	.0320*** (.00806)
` '		0526***(.00653)
		$0741^{***} (.00526)$
` '	` /	.00361 (.00272)
ore matching	, ,	,
1,888 (49.66%)	1,914 (50.34%)	
,		0236 (.0360)
		.0114 (.0162)
` '	\ /	000148 (.00532)
` '	` /	.00630 (.00858)
` /	` ,	.00251 (.00301)
` ,	` /	$0169\ (.0158)$
` '	` ,	.0156 (.0151)
` '	` ,	00224 (.00323)
` '	, ,	$000281\ (.00460)$
` '	, ,	.000825 (.0122)
` '	` ,	00134 (.00766)
` ,	` ,	000602 (.00240)
.013 (.112)	.015 (.122)	.00244 (.00380)
distance	, ,	,
	2,682 (50.00%)	
		.441*** (.0354)
	` ,	.0975*** (.0144)
\ /	\ /	.0691*** (.00709)
(/	` ,	.0414*** (.0102)
` /	` ,	0178*** (.00511)
` ,	` /	.128*** (.0169)
` /	` ,	0657^{***} (.0153)
\ /	` ,	00186 (.00167)
` /	` ,	00343 (.00332)
` '	` ,	.0371** (.0138)
` '	` ,	$0318^{***} (.00783)$
		$0453^{***} (.00668)$
		.000709 (.00292)
	7.136 (.658) .874 (.332) .867 (.340) .041 (.198) .245 (.430) .365 (.481) .009 (.094) .021 (.144) .127 (.333) .102 (.303) .078 (.269) .012 (.111) ore matching 1,888 (49.66%) 6.063 (1.133) 7.266 (.404) .972 (.164) .921 (.270) .007 (.086) .400 (.490) .312 (.463) .011 (.105) .021 (.142) .170 (.376) .060 (.237) .006 (.076)	7.136 (.658)

Notes. Comparison of whole and balanced samples for comments written from entities without or with direct access to policy-makers: number of observations, mean of observed comment characteristics, and their standard deviation in parenthesis. Comment length is measured as the log of the number of words in a comment. Comment complexity is measured as the average length of words in a comment, after removing stop words. Main EU languages equals one when a comment is written in English, French or German. The third column reports the statistical difference in means, with standard error in parenthesis. * p < 0.05, *** p < 0.01, **** p < 0.001

of words, and the comment complexity computed as the average length of words. It also includes a variable indicating whether the comment has been written in one of the main EU languages. As comments can be written in any of the 24 official EU languages and the Commission staff is left in charge of translating it, one can assume that the original language of the comment being English, French, or German might increase its probability of being accounted for. One could also consider that the European Commission prioritizes suggestions made by EU entities, I thus include an EU origin indicator variable. Lastly, δ_r indicates regulation fixed effects, which control for regulation-specific changes in practices or ways of drafting the regulation and including comments. Standard errors are clustered by entity.

Table 5 presents the results of this model. Columns (1) and (2) report results for the outcome measure based on plagiarism-detection. Columns (3) and (4) report results for the LLM-based measure. I report estimations of the model on the PSM sample in Columns (1) and (3) and on the MDM sample in Columns (2) and (4).

I find a robust and significant association between direct access to policy-makers and the likelihood of lobbying success.¹¹ The estimations performed on the sample matched with propensity score matching (PSM) reveal a 28.79%, or 28.66%, increase in the odds of lobbying success for comments from entities having a direct access to policy-makers, respectively with the plagiarism-detection measure and the large language model (LLM) success measure.¹² The two coefficients are not statistically different from each other at any confidence level (p-value of 0.992). However, the results are more noisy with the LLM measure, with a p-value of 0.017 against 0.008 for the plagiarism-detection measure. This indicates a greater heterogeneity of the effect when using a more flexible way of detecting lobbying success. It suggests that this more flexible measure captures more heterogeneous success cases, compared to plagiarism-detection. I investigate this heterogeneity further below.

I find similar results when estimating this model on the Mahalanobis distance matching sample (MDM). I find an increase in the odds of a comment's success when the entity writing has access to policy-makers of 24.98% when measuring success with plagiarism-detection, and of 22.14% when measuring with LLM. These two coefficients are not statistically different from each other, with a p-value of 0.870. Results are again less precise when estimating with the LLM success measure, with a p-value of 0.063 against 0.022 for the plagiarism-detection measure. The two estimations on the MDM sample are less precise than their respective estimations on the more balanced PSM sample.

⁹Appendix B.2 describes the construction of these two comment quality measures in detail.

¹⁰The UK is considered within the EU for comments written until January 31st, 2020.

¹¹The highest p-value across the four coefficient is 0.063.

¹²From the logit estimated coefficient β , a percent-increase interpretation requires to apply an $\exp(\beta) - 1$ transformation. Note that interpretations are on odds: the ratio of the probability of success and probability of failure.

Table 5: Baseline results

Dependent variable:	Lobbying success $(Y = 1)$				
	Plagiarisn	n-detection	L	$_{ m LLM}$	
Sample balancing:	PSM	MDM	PSM	MDM	
	(1)	(2)	(3)	(4)	
Access before adoption	.253**	.223*	.252*	.200	
A d:-/	(.0958)	(.0971)	(.105)	(.107)	
Academic/research Institution	390 (.523)	239 (.380)	-1.821 (1.067)	-1.210** (.371)	
Business association	.0331	.0709	.0501	.0353	
	(.117)	(.117)	(.141)	(.157)	
Consumer organisation	-2.123**	-2.124***	-1.321	873	
-	(.665)	(.508)	(.859)	(.486)	
Environmental organisation	903**	636	.534	.465*	
N	(.322)	(.372)	(.289)	(.220)	
Non-governmental organisation	370*	323*	0580	.0459	
Other	(.163) 0989	(.158) $.208$	(.180) 121	(.172)	
Other	(.234)	(.268)	(.270)	455 (.298)	
Public authority	(.2 3 4) 925	212	.673	.189	
1 done authority	(.594)	(.433)	(.722)	(.224)	
Trade union	686	0624	1.029	.894	
Trade union	(.439)	(.358)	(.532)	(.528)	
Comment length	1.227***	1.226***	.211***	.206***	
	(.0655)	(.0608)	(.0605)	(.0609)	
Comment complexity	.0843	.147	0464	134	
_ *	(.179)	(.161)	(.124)	(.127)	
Main EU languages	.970*	.726*	0147	.420	
	(.454)	(.295)	(.339)	(.237)	
EU origin	190	.00412	0770	178	
	(.185)	(.166)	(.206)	(.231)	
Regulation FE	Yes	Yes	Yes	Yes	
Observations	3,120	6,958	2,870	6,491	
Pseudo \mathbb{R}^2	.259	.305	.154	.194	

Notes. Estimations done with a logit model. The dependent variable equals 1 if entity i lobbied successfully on r. In Columns (1) to (2), success is measured through plagiarism-detection. In Columns (3) to (4), success is measured with a large language model (LLM). Balanced samples created using propensity score matching (PSM) or Mahalanobis distance matching (MDM). The independent variable $Access_{ir}$ equals 1 if i had direct meetings with policy-makers before the adoption of r. The baseline type of entity is "Company/business organization". Standard errors clustered by entity are in parenthesis. *p<.05; **p<.01; ***p<.001.

17

Overall, I obtain similar results across outcome measures and samples used, both qualitatively and quantitatively: a comment written by an entity having access to policy-makers is 22 to 29% more likely to influence the final version of a regulation, compared to a comment written by an entity not having such access to policy-makers.

Access to different hierarchical levels. I find that access to policy-makers leads to a higher likelihood of lobbying success, pooling access to all hierarchical levels together. I here decompose this access in access to four different hierarchical levels: the Presidency, the Directors-General, the Commissioners and the Cabinet Members. Commissioners are politicians nominated by Member States and validated by the European Parliament. Among them, I isolate the President, the Vice-Presidents and the High Representative to classify them as "the Presidency". The Directors-General are the highest level bureaucrats in charge of specific policy area. They report to Commissioners. Note that these four variables of access are correlated with each other, with coefficients of correlation ranging from 0.4847 for Commissioners and the Presidency to 0.6923 for Commissioners and Cabinet Members.

I re-estimate Equation 2 for each of these hierarchical levels and report the results in Columns (1) to (4) of Table 6. I also estimate a model with all hierarchical levels and report these results in Column (5) of Table 6. Estimations presented in Table 6 are done using the lobbying success measure based on LLM and the PSM sample. I also present and discuss below the results of the estimations from other balanced samples.

The results show that access to cabinet members is less valuable for lobbying success than access to the Presidency or to a Director-General (DG). The coefficient for access to cabinet members reported in Column (4) is significantly lower than the coefficients for Presidency or DG access, with a p-value of 0.027 and 0.038 respectively. Access to the Presidency of the Commission or to a DG drives the results observed before. This is also observed in Column (5), although the results are less precise because of the correlation between the variables. In this last model, only the difference between the coefficients for Presidency access and cabinet member access can be considered to be statistically significant, with a p-value of 0.062.

These estimations are done using the large language model success measure and the PSM sample. Appendix Table 11 presents the results of the same estimations, performed on the MDM sample. Similarly to Table 6, it shows a stronger effect of access when considering access to the Presidency or Directors-General, compared to access to Commissioners or cabinet members (with larger coefficients for Presidency access). The effect of access to the Presidency is statistically greater than the effect of access to Commissioners and cabinet members, both in the separate regressions, displayed in Columns (1) to (4), and in the model

Table 6: Results by hierarchical level

Dependent variable:		Lobbying	g success ((Y = 1)	
	(1)	(2)	(3)	(4)	(5)
Access to					
Presidency	.489*** (.128)				.329* (.162)
Directors-General	()	.421*** (.114)			.300
Commissioners		,	.341** (.111)		.0801 (.191)
Cabinet Members			()	.233* (.105)	143 (.177)
Types	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y
Regulation FE	Y	Y	Y	Y	Y
Observations	2,870	2,870	2,870	2,870	2,870
Pseudo R^2	.157	.157	.155	.154	.159

Notes. Estimations done with a logit model. The dependent variable is a measure of lobbying success based on a large language model. Balanced sample created using propensity score matching. Standard errors clustered by entity in parenthesis. * p < 0.05, ** p < 0.01, *** p < 0.001

with all access variables shown in Column (5).¹³

Entities may select differently into access to the Presidency than into general access to the European Commission. I thus re-estimate the model for Presidency access on a balanced sample created using propensity score matching on Presidency access to check the robustness of this result. Appendix Table 12 reports the results for the estimations on the Presidency access PSM sample. The large positive effect of having access to the Presidency on lobbying success is robust to the different matching methods used.

Access to the highest levels of the Commissions' politicians or bureaucrats hierarchy brings larger benefits in terms of lobbying success probability than access to lower levels, this type of access drives the effect found earlier.

Access intensity. The indicator variable of access to policy-makers pools together comments written by entities with a single meeting, and up to 311 meetings with the Commission

 $^{^{13}}$ p-values of the statistical differences are respectively 0.021 and 0.013 for the separate regressions, and 0.006 and 0.003 for the common model.

before the adoption of regulation they comment on.¹⁴ Comments written by an entity with more numerous meetings may be more likely to be successful than a comment written by an entity with few meetings. I here consider access intensity, measured by the number of meetings an entity has before the adoption of the regulation it comments on.

To capture the nuanced effects by intensity of access, I estimate a model with quartiles of the number of meetings:

$$\mathbb{P}(y_{ir}) = \alpha + \sum_{q \in \{1,2,3,4\}} \beta_q \, Qq_{ir} + \sum_t \gamma_t \, type_i^t + \eta \, X_{ir} + \delta_r + \varepsilon_{ir}$$
(3)

where Q1, Q2, Q3, and Q4 represent quartiles of the number of meetings associated with a comment, *i.e.* the number of meetings the entity writing the comment has with policy-makers before the adoption of the relevant regulation. Quartiles are calculated excluding zeros. The rest of the notations are the same as defined for Equation 2. Standard errors are clustered by entity.

Coefficients β_1 , β_2 , β_3 and β_4 indicate the additional probability of success a comment in the respective quartiles faces, relative to a comment written by an entity without access to policy-makers. The results from these estimations, using the LLM success measure and the PSM sample, are presented in Table 7. I find that the effect of access to policy-makers on lobbying success is driven by comments written by entities having more meetings with policy-makers, especially for comments in the third and fourth quartiles, who experience a significantly larger effect than comments in the first quartile (coefficients are statistically different, with a p-value of respectively 0.043 and 0.063). Compared to comments written by entities not having access to policy-makers, comments written by entities having more intense access to policy-makers are the most likely to be successful and influence regulatory text changes, while having little access (1 to 3 meetings) does not have an effect.

Direct access to policy-makers substantially increases the odds of an entity being able to influence the final version of a regulation through a comment written in the European Commission commenting process. This effect is heterogeneous across different types of access. It is stronger for access to higher levels of the Commission hierarchy (the Presidency, and to a smaller extent Directors-General). Moreover, the effect is driven by comments written by

¹⁴The number of meetings associated with a comment is the number of meetings the entity writing this comment had between the beginning of the meeting records and the adoption of the associated regulation. By construction of the variable, recent comments can be associated with more numerous past meetings of the entity writing this comment. The 14 comments with 200 meetings or more before the adoption of the associated regulation have all been written since 2020. They come from major companies, business associations, consumer organisations and NGOs: Google (2 comments), BusinessEurope, Bureau Européen des Unions de Consommateurs (5 comments), Transport and Environment (3), Airbus (2), and WWF.

Table 7: Results by quartiles of the number of meetings

Dependent variable:	Lobbying success $(Y = 1)$
Access: number of meetings	
Quartile 1 [1, 3]	.0116 (.164)
Quartile 2 [4, 8]	.159 (.180)
Quartile 3 [9, 26]	.420** (.158)
Quartile 4 [27, 311]	.397* (.163)
Type	Y
Controls	Y
Regulation FE	Y
Observations	2,870
Pseudo R^2	.156

Notes. Logit model. Balanced sample created using propensity score matching. Quartiles of the number of meetings computed on the balanced sample. The baseline is comments written by an entity not having access to policy-makers. Standard errors clustered by entity in parenthesis. *p<.01; **p<.05; ***p<.001.

entities with more numerous meetings with policy-makers.

4 Mechanisms

One open question in the political economy literature on informational lobbying is distinguishing the information and connection channels (Bombardini and Trebbi, 2020). In my analysis, the fact that organizations having access to policy-makers have a higher probability to lobby successfully may come from four main channels. First, organizations bring credible information to the policy-makers during meetings to convince policy-makers of the relevance of their requests, and their comments are thus taken into account by policy-makers. Second, policy-makers may be more likely to account for comments written by organizations they are connected with, favoring them irrespectively of the value of the information they bring. Third, organizations may build institutional knowledge through their meetings with the European Commission and gain insider knowledge, which can be instrumental in lobbying successfully. Fourth, these organizations may have other unobserved characteristics

that explain both their access to policy-makers and their higher likelihood of success. Such characteristics may be credibility, reputation, intrinsic quality of the organization and information it produces. They may produce the best and most relevant information, which makes their comment more likely to be taken into account. Assuming policy-makers are efficient at selecting who they meet with, these organizations also have access to policy-makers. In the analysis shown above, I use balanced samples to study comments of comparable characteristics to account for and minimize this fourth channel. However, this channel may still explain part of the result. In this section, I isolate the different channels to analyze them separately.

I exploit the time dimension and the change in Commission mandate that the data offers to isolate the different channels and shed new light on this question. The first channel is the information-transmission channel, which is time- and topic-specific. Second, the political connection channel is person-specific, it is associated to the members of the Commission the organization meets with. Third, the institutional knowledge channel is institution-specific. It does not depend on a specific topic, period, or person. Fourth, the intrinsic quality of the organization is not specific to a time or a person met. I assume this intrinsic quality to be time-invariant.

I leverage the Commission change that took place in December 2019 between the Juncker Commission and the first Von der Leyen Commission to distinguish different timing of access.

First, I consider access to policy-makers contemporaneously to the elaboration of a regulation on which an entity writes a comment. I consider an entity to have contemporaneous access when it has at least one meeting in a time window sufficiently close to the comment, and before the adoption of the regulation. Some organizations might be able to access drafts before their official publication, I thus consider a time window starting before the draft publication. Figure 2 presents the distribution of the time between an organization's comment and its last and next meeting with the European Commission, for organizations that both write comments and have access to policy-makers. It shows that organizations tend to meet with policy-makers close to the day they write a comment, starting approximately three months before. I choose three-month before the draft publication as the cut-off to consider a meeting contemporaneous to a comment.¹⁵ Meetings during this time period are likely related to the regulation the organization comments. Organizations might use such meetings to bring relevant information and convince policymakers to take their comment into account when modifying the draft regulation. Additionally, such meetings may as well be a way of building political connection or institutional knowledge, or be a signal of the organization's intrinsic quality. I assume all channels to be potentially at play in these meetings (see Table

 $^{^{15}}$ This corresponds to about four months before a comment, as more than 90% of the comments are written 21 to 35 days after the draft publication.

8 for a recapitulation of the timing of access and channels potentially at play).

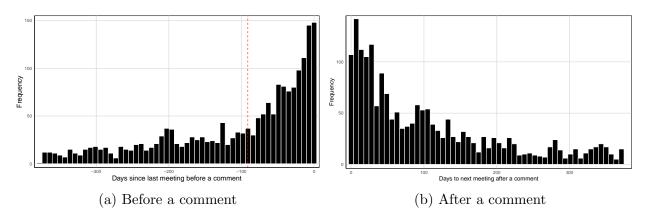


Figure 2: Time between a comment and its closest meetings

Notes. The red dashed line represents three months before the comment.

Second, I consider entities having access to the Commission more than three months before the draft publication. Such access may contribute to the organization building political connections and institutional knowledge, or be a signal of the organization's quality. However, I assume that organizations do not transmit information on or request potential draft modifications at this stage, as the draft would be published in more than three months. Although these meetings may be related to the topic of the regulation, and even influence the draft itself, I assume that the information-transmission channel is not at play in meetings more than three months before the draft publication.

Third, I leverage the change of Commission mandate and the associated turnover of Commission members to differentiate between political connections and institutional knowledge building. Political connections are specific to the members of the Commission an organization meets with. These connections only survive the change of Commission if the connected members stay in the Commission after the change of mandate. Otherwise, the connections are not active anymore, and the effect of past access to members of the Commission who left cannot be explained by the political connection channel. However, the institutional knowledge built through these meetings remains, as well as the intrinsic quality of the organization.

Fourth, I consider future meetings, happening after the regulation is adopted. These meetings cannot affect the likelihood of a comment being accounted for in the adopted regulation, as they happen later. Such meetings cannot be used as a mean of information transmission, nor to build political connections or institutional knowledge that could be useful for the given comment. However, the effect of this type of access on the success probability of the comment represents the fourth channel, which is the intrinsic quality of the organization.

I estimate a model with these four types of access. Using the notation defined for Equation

Table 8: Timing of access & potential benefits

	Channels				
Access:	Information transmission	Political connection	Institutional knowledge	Intrinsic quality	
Contemporaneous	Yes	Yes	Yes	Yes	
Prior / Past Com. (stayed)	No	Yes	Yes	Yes	
Past Commission (left)	No	No	Yes	Yes	
Post adoption	No	No	No	Yes	

2, the model writes

$$\mathbb{P}(y_{ir}) = \alpha + \beta_I Contemp A_{ir} + \beta_P Prior A_{ir} + \beta_S Past Com S_{ir}
+ \beta_L Past Com L_{ir} + \beta_F Post Adoption_{ir}
+ \sum_t \gamma_t type_i^t + \eta X_{ir} + \delta_r + \varepsilon_{ir}$$
(4)

where $ContempA_{ir}$ is the indicator variable for i having access to policy-makers contemporaneously to its comment on regulation r. $PriorA_{ir}$ is the indicator variable for i having access to policy-makers of the current Commission more than three months before the draft publication of regulation r. $PastComS_{ir}$ is the indicator variable for i having with members of the past Commission who stayed. $PastComL_{ir}$ is the indicator variable for i having access to policy-makers from the past Commission who left. $PostAdoption_{ir}$ equals one when organization i has access to policy-makers after the adoption of regulation r.

 β_I represents the effect of the combination of all four channels. β_P and β_S represent the combined effect of political connection, institutional knowledge, and intrinsic quality. β_L is the effect of institutional knowledge and intrinsic quality, while β_F solely estimates the latter. It follows that the effect of institutional knowledge can be retrieved by computing the difference between β_F and β_L . The same backward logic applies to the other channels.

To estimate this model, I restrict the sample to comments written during the von der Leyen Commission to be able to build the variables $PastComS_{ir}$ and $PastComL_{ir}$ based on entities' past meetings with the Juncker Commission.

I report the results from this estimation in Table 9. It shows that when controlling for all types of access, having access to policy-makers contemporaneously to commenting on a regulation does not increase the comment's success probability. Access to members of the past Commission who left and access after the adoption of the regulation do not have an effect

either. It suggests that when accounting for all possible channels, institutional knowledge and intrinsic quality of the organization are not the main drivers of the positive effect of access on success probability. On the contrary, having access to members of the previous Commission who stayed after the Commission changed, and thus having potential active and longer-term connections, has a positive effect on the probability of lobbying success, controlling for all other type of access.

Table 9: Mechanisms results

Dependent variable:	Lobbying success $(Y = 1)$		
	(1)		
Access			
contemporaneously	205		
(up to the regulation adoption)	(.183)		
more than 3 months before	.104		
draft (same Commission)	(.189)		
to past Commission - stayed	.468*		
-	(.193)		
to past Commission - left	127		
	(.192)		
post adoption	.0841		
•	(.182)		
Types	Y		
Controls	Y		
Regulation FE	Y		
Observations	4,414		
Pseudo R^2	.168		

Notes. Results from logit estimations. Sample restricted to comments written during the von der Leyen Commission. LLM-based success measure. Standard errors clustered by entity are in parentheses. *p<.05; **p<.01; ***p<.001.

Together, these results indicate that the increased lobbying success probability associated with access to policy-makers cannot be entirely tied to the information this organization provides to the policy-makers, or the intrinsic quality of this organization. The results suggest that political connections play a more important role than the other channels.

5 Conclusion

In this paper, I study the relationship between direct access to policy-makers and lobbying success, in the context of online commenting process implemented by policy-makers to increase accountability, legitimacy and transparency of the policy-making process.

To analyze this, I create a novel and unique dataset that combines textual data on comments, draft and adopted regulations to measure lobbying success and temporal meetings data to assess access to policy-makers over time. Lobbying success is measured with two complementary methods, one based on plagiarism-detection and one based on a large language model (GPT-4).

The paper provides direct evidence on the role of access to policy-makers in lobbying. I find that access to policy-makers substantially enhances the probability of lobbying success when comparing similar comments. This effect is driven by access to the highest level politicians and highest level bureaucrats. Exploiting a change of Commission, I present suggestive evidence that having active and longer-term political connections seems to explain this effect to a greater extent that information-transmission, insider knowledge accumulation, or initial better characteristics.

References

- Anger, N., Asane-Otoo, E., Böhringer, C., and Oberndorfer, U. (2016). Public interest versus interest groups: a political economy analysis of allowance allocation under the EU emissions trading scheme. *International Environmental Agreements: Politics, Law and Economics*, 16(5):621–638.
- Anger, N., Böhringer, C., and Lange, A. (2015). The political economy of energy tax differentiation across industries: theory and empirical evidence. *Journal of Regulatory Economics*, 47(1):78–98.
- Baumgartner, F. R., Berry, J. M., Hojnacki, M., Leech, B. L., and Kimball, D. C. (2009). Lobbying and policy change: Who wins, who loses, and why. University of Chicago Press.
- Bertrand, M., Bombardini, M., Fisman, R., Hackinen, B., and Trebbi, F. (2021). Hall of mirrors: Corporate philanthropy and strategic advocacy. *The Quarterly Journal of Economics*, 136(4):2413–2465.
- Bertrand, M., Bombardini, M., and Trebbi, F. (2014). Is it whom you know or what you know? an empirical assessment of the lobbying process. *American Economic Review*, 104(12):3885–3920.

- Besley, T. and Coate, S. (2003). Elected versus appointed regulators: Theory and evidence. Journal of the European Economic Association, 1(5):1176–1206.
- Bombardini, M. and Trebbi, F. (2020). Empirical models of lobbying. *Annual Review of Economics*, 12:391–413.
- Brown, J. R. and Huang, J. (2020). All the president's friends: Political access and firm value. *Journal of Financial Economics*, 138(2):415–431.
- Bunea, A. (2013). Issues, preferences and ties: determinants of interest groups' preference attainment in the eu environmental policy. *Journal of European Public Policy*, 20(4):552–570.
- Burghaus, K., Koch, N., Bauer, J., and Edenhofer, E. (2019). Lobbying, relocation risk and allocation of free allowances in the EU ETS. World Bank, Washington, DC Coalition, Carbon Pricing Leadership.
- Casas, A., Denny, M. J., and Wilkerson, J. (2020). More effective than we thought: Accounting for legislative hitchhikers reveals a more inclusive and productive lawmaking process.

 American Journal of Political Science, 64(1):5–18.
- Chalmers, A. W. (2013). Trading information for access: informational lobbying strategies and interest group access to the European Union. *Journal of European Public Policy*, 20(1):39–58.
- Djourelova, M. (2023). Persuasion through slanted language: Evidence from the media coverage of immigration. *American Economic Review*, 113(3):800–835.
- Djourelova, M., Durante, R., Motte, E., and Patacchini, E. (2023). Experience, narratives, and climate change beliefs. *Available at SSRN 4680430*.
- Dür, A., Bernhagen, P., and Marshall, D. (2015). Interest group success in the european union: When (and why) does business lose? *Comparative Political Studies*, 48(8):951–983.
- European Commission (2001). European Governance White Paper. COM/2001/0428.
- European Commission (2015). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Better regulation for better results An EU agenda, COM/2015/0215 final.
- European Commission (2017). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of

- the Regions. Completing the Better Regulation Agenda: Better solutions for better results, COM/2017/0651 final.
- European Commission (2021a). Better Regulation Toolbox.
- European Commission (2021b). Commission Staff Working Document. Better Regulation Guidelines, SWD/2021/0305 final.
- Ferguson, T. and Voth, H.-J. (2008). Betting on hitler—the value of political connections in nazi germany. *The Quarterly Journal of Economics*, 123(1):101–137.
- Fisman, R. (2001). Estimating the value of political connections. *American Economic Review*, 91(4):1095–1102.
- Gilardi, F., Alizadeh, M., and Kubli, M. (2023). Chatgpt outperforms crowd-workers for text-annotation tasks. arXiv preprint arXiv:2303.15056.
- Hermansson, H. (2016). The European Commission's environmental stakeholder consultations: Is lobbying success based on what you know, what you own or who you know? *Interest Groups & Advocacy*, 5:177–199.
- Jann, B. (2017). KMATCH: Stata module module for multivariate-distance and propensity-score matching, including entropy balancing, inverse probability weighting, (coarsened) exact matching, and regression adjustment. Statistical Software Components, Boston College Department of Economics.
- King, G. and Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political analysis*, 27(4):435–454.
- Klüver, H. (2012). Informational lobbying in the European Union: The effect of organisational characteristics. West European Politics, 35(3):491–510.
- Klüver, H. (2013). Lobbying as a collective enterprise: winners and losers of policy formulation in the European Union. *Journal of European Public Policy*, 20(1):59–76.
- Laurens, S. (2015). Les courtiers du capitalisme. Milieux d'affaires et bureaucrates à Bruxelles. Agone.
- Lee, J. J., Ziraldo, E. R., Jeong, H., and L'Abbé, M. R. (2024). Examining the role of industry lobbying on canadian front-of-pack labelling regulations. *Canadian Journal of Public Health*, pages 1–7.

- Leuven, E. and Sianesi, B. (2018). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing.
- Mahoney, C. (2007). Lobbying success in the United States and the European Union. *Journal of Public Policy*, 27(1):35–56.
- Meng, K. C. and Rode, A. (2019). The social cost of lobbying over climate policy. *Nature Climate Change*, 9(6):472–476.
- Morelli, M., Giommoni, T., and Paserman, D. (2023). Does position taking crowd out policy making? Congressional capacity and the quality of laws.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.
- Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38.
- Törnberg, P. (2023). Chatgpt-4 outperforms experts and crowd workers in annotating political twitter messages with zero-shot learning. arXiv preprint arXiv:2304.06588.

Appendices

A Appendix: Additional Tables

Table 10: Modifications and Comments per Directorate-General

Directorate-General	N	Del.	New	Comments	Found
Competition	1	.38	.48	29.00	.42
Energy	27	.28	.34	$2,\!586.07$.14
Justice and Consumers	2	.16	.16	30.00	.22
Mobility and Transport	44	.13	.18	17.16	.13
Health and Food Safety	132	.13	.15	55.21	.08
Climate Action	24	.13	.20	12.62	.10
Neighbourhood and Enlargement Negotiations	1	.11	.15	4.00	.05
Environment	21	.11	.12	54.52	.10
FISMA	25	.10	.12	1,882.88	.23
GROW	54	.09	.15	14.56	.13
Secretariat-General	1	.09	.12	6.00	.00
Connect	3	.09	.11	59.00	.31
Regional and Urban Policy	1	.07	.05	1.00	.00
Economic and Financial Affairs	2	.06	.04	6.50	.02
Migration and Home Affairs	2	.04	.09	5.00	.07
Agriculture and Rural Development	64	.03	.04	22.45	.08
Maritime Affairs and Fisheries	7	.03	.03	3.86	.16
Taxation and Customs Union	11	.02	.03	8.64	.03
Eurostat	21	.01	.02	2.43	.00
Total	443	.11	.14	291.38	.11

Notes. This table displays the number of initiatives studied per Directorate-General, the average share of deleted or modified text from the draft (Del.), the average share of new text in the adopted regulation (New), the average number of comments per initiative, and the average share of modified text mapped to at least one specific comment with the text reuse method. DG FISMA stands for Financial Stability, Financial Services and Capital Markets Union. DG GROW is the Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs. DG Connect is the Directorate-General for Communications Networks, Content and Technology.

Table 11: Results by hierarchical level - MDM sample

Dependent variable:		Lobbyin	g success	(Y=1)	
	(1)	(2)	(3)	(4)	(5)
Access to					
Presidency	.622*** (.168)				.644** (.207)
Directors-General	()	.386** (.137)			.208 (.181)
Commissioners		(/	.272* (.135)		152 (.218)
Cabinet Members			(.100)	.204 (.110)	0928 (.188)
Types	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y
Regulation FE	Y	Y	Y	Y	Y
Observations Pseudo R^2	6,491 $.200$	6,491 .196	6,491 .194	6,491 .194	6,491 .201

Notes. Estimations done with a logit model. The dependent variable is a measure of lobbying success based on a large language model. Balanced sample created using Mahalanobis distance matching. Standard errors clustered by entity in parenthesis. * p < 0.05, ** p < 0.01, *** p < 0.001

Table 12: Results by hierarchical level - Presidency access PSM sample

Dependent variable:	Lobbying success $(Y = 1)$
	(1)
Access to	
Presidency	.488*
Directors-General	(.237) .408
Commissioners	(.240) 0306
Cabinet Members	(.279) 195
	(.276)
Types Controls Regulation FE	Y Y Y
Observations Pseudo \mathbb{R}^2	1,454 .171

Notes. Estimations done with a logit model. The dependent variable is a measure of lobbying success based on a large language model. Balanced sample created using propensity score matching (PSM) for access to the Presidency. Standard errors clustered by entity in parenthesis. * p < 0.05, ** p < 0.01, *** p < 0.001

B Appendix: Text analysis methods

B.1 Measures of lobbying success

B.1.1 Text reuse approach

In this section, I start by describing the text pre-processing and the construction of the measure of lobbying success I use. I then present examples of lobbying success.

I am interested in policy changes only, the first step is thus to pre-process drafts and adopted regulations to consider only the regulatory text itself. I remove the explanatory memorandum of delegated regulations as it is not legally binding. I remove headers, footers, signatures, dates, locations, as well as footnotes, which are citations of other legal acts. I also remove all procedural texts, such as "This Regulation shall be binding in its entirety and directly applicable in all Member States". I do consider the annex, which stipulates key definitions, exemptions, thresholds, or criteria for the application of regulations.

For comments, I consider together the comment and its potential enclosure, after having translated them to English.

The pre-processing of all text –comments, drafts, and regulations– ends by transforming it to lowercase, removing punctuation, symbols, and stop-words, stemming and tokenizing it to n-grams.

To keep only the changing parts of the regulation, I isolate n-grams that are deleted, added, or modified between the draft and the adoption. I label a comment as being successful if its n-grams overlap with these changes. I consider 5-grams as it enables some context without being too long (Casas et al., 2020; Djourelova, 2023).

Using this procedure, 2,666 cases of lobbying success are found. Appendix Figure 3 displays two of them: one from a business association that states clearly what they want by directly suggesting edits to the draft and one from a citizen whose comment is not in English and does not include direct edit suggestions. The first example is a comment by the European Agricultural Machinery Industry Association (CEMA) on the draft regulation on ecodesign requirements for lighting products. The enclosed document of this comment shows clearly, in bold, what changes they want to see in the regulation (Figure 3c). The regulation changes exactly in this direction (Figures 3a and 3b). The second example is a comment written in Spanish by a citizen. Figure 3f shows part of this comment on the draft regulation setting out scientific criteria for the determination of endocrine-disrupting properties. After translation of the comment, the algorithm finds that the regulation on criteria to identify endocrine disruptors changes and now accounts for this comment. These policy changes are visible in Figures 3d and 3e. In both cases, lobbying is successful as the comment is taken

into account in the changes of policy observed between the draft stage and the adoption stage of a regulation.

B.1.2 LLM-based measure

I also create a more flexible measure of lobbying success using GPT-4, a large language model (LLM) provided by OpenAI.

In practice, I start by removing procedural text as explained above. I then isolate each article or annex to compare them individually between the draft and the adopted regulation. I discard articles or annex that remain identical. For articles or annex that are modified, I ask GPT to enumerate the policy-relevant modifications with the following prompt:

You are a policy-maker. You will be provided with an article of a draft regulation first, followed by its final version. Your goal is to identify and extract only policy-relevant changes between the draft and the final version. Ignore purely stylistic or wording changes that do not impact the regulatory obligations, scope, or intent. If the draft is an empty string, it means that the article is new. Explain the additions. If the final version is an empty string, it means that the article has been deleted. Explain what has been discarded from the regulation.

For each policy-relevant change, provide a 1-sentence explanation, clearly explaining the impact of the change on the regulation's implementation or stakeholders.

For each comment, GPT has to enumerate the requested changes. The prompt is:

You are a policy-maker. You will be provided with a comment on a draft regulation. Your goal is to extract requests from comments, as well as a 1-sentence summary for these requests.

I then combine the modifications of each article or annex of a regulation to compare them with comments written on this regulation. Finally, for each comment, I ask GPT whether the requested changes have been incorporated in the modifications observed.

You are a policy-maker tasked with evaluating stakeholder influence on regulatory changes. You will first be provided with article-by-article policy changes between the draft and final versions of a regulation. Next, you will receive requests made by a stakeholder, indicating how they wanted the draft to evolve. Your goal is to identify whether the stakeholder's requests successfully led to changes in the final regulation. Respond with 'Yes' or 'No', followed by a brief explanation. Indicate whether the stakeholder's requests have been 'entirely successful', 'partially successful', or 'not successful'.

I consider a comment to be successful as soon as one of its requests is accounted for in the regulation modifications.

B.2 Measures of comment length and complexity

To capture the different levels of complexity of comments sent to policy-makers, I use the metrics of the number of words, as well as the average length of words. ¹⁶ I preprocess comments by removing numbers, symbols, URLs, stopwords, and single-character words to consider actual words only. The length of a word is the number of characters it contains. 70 comments of the sample do not contain any word after pre-processing the data, I set their average word length to zero.

C Appendix: Matching comments and meetings data

Meeting data is provided by Eurostat, with the Transparency Register identification number of each organization attending a meeting. On the other hand, comments identifying information are filled out by the organizations themselves. It follows that they do not always provide their Transparency Register identification number when they have one, or that there are mistakes in the way it is entered.

I start by cleaning the Transparency Register identification numbers of the comment database by removing special characters and letters. For observations that do not correspond to the usual format (8 to 13 numbers, hyphen, 2 numbers), I proceed to manual correction of typographical errors. I discard other identifying numbers such as ZVRs for Austrian organizations or SIREN for French organizations that have been entered by mistake instead of the Transparency Register identification number (TR ID).

I then identify the different names and IDs one organization can have (e.g. name, acronym, and name with a typographical error). The TR ID of an entity can also change throughout the period when an organization leaves the Transparency Register and registers back later on.

These steps enable me to match the comments and meetings data, based on TR IDs and organization's names.

¹⁶See for instance Morelli et al. (2023).

ANNEX III Exemptions

This Regulation shall not apply to light sources and separate control gears specifically tested and approved to operate

- (a) in potentially explosive atmospheres, as defined in Directive 2014/34/EU of the European Parliament and of the Council 3;
- (b) for emergency use, as set out in Directive 2014/35/EU of the European Parliament and of the Council 4
- (c) in radiological and nuclear medicine installations, as defined in Article 3 of Directive 2009/71/EURATOM 5;
- in or on military or civil defence establishments, equipment, ground vehicles, marine equipment or aircraft, as set out in Member States' regulations or in documents issued by the European Defence Agency;
- (e) in or on motor vehicles, their trailers and systems, interchangeable towed equipment, components and separate technical units as set out in Regulation (EC) No 661/2009 of the European Parliament and of the Council ⁵, Regulation (EU) No 167/2013 of the European Parliament and of the Council ⁷ and Regulation (EU) No 168/2013 of the European Parliament and of the Council ⁵;
- (f) in or on non-road mobile machinery as set out in Regulation (EU) 2016/1628 of the European Parliament and of the Council *;
- (g) in or on civil aviation aircrafts, as set out in Commission Regulation (EU) No 748/2012 10;
- (h) in railway vehicle lighting, as set out in Directive 2008/57/EC of the European Parliament and of the Council 11 ;
- (i) in marine equipment, as set out in Directive 2014/90/EU of the European Parliament and of the Council 12;
- in medical devices, as set out in Council Directive 93/42/EEC ¹³ and in vitro medical devices as set out in Directive 98/79/EC of the European Parliament and of the Council ¹⁶.

(a) Draft regulation on lighting products

ANNEX III Exemptions

This Regulation shall not apply to light sources and separate control gears specifically tested and approved to operate:

- (a) in potentially explosive atmospheres, as defined in Directive 2014/34/EU 3;
- (b) for emergency use, as set out in Directive 2014/35/EU 4;
- (c) in radiological and nuclear medicine installations, as defined in Article 3 of Directive 2009/71/EURATOM 5;
- (d) in or on military or civil defence establishments, equipment, ground vehicles, marine equipment or aircraft, as set out in Member States' regulations or in documents issued by the European Defence Agency;
- (e) in or on motor vehicles, their trailers and systems, interchangeable towed equipment, components and separate technical units as set out in Regulation (EC) No 661/2009 ⁸, Regulation (EU) No 168/2013 ³ and Regulation (EU) No 168/2013 ³;
- f) in or on non-road mobile machinery as set out in Regulation (EU) 2016/1628 ⁹ and in or on their trailers;
- (g) in or on interchangeable equipment as set out in Directive 2006/42/EC ¹⁰ intended to be towed or to be mounted and fully raised from the ground or that cannot articulate around a vertical axis when the vehicle to which it is attached is in use on a road by vehicles as set out in Regulation (EU) pol 167/2013 ¹¹;
- (h) in or on civil aviation aircraft, as set out in Commission Regulation (EU) No 748/2012 12;
- (i) in railway vehicle lighting, as set out in Directive 2008/57/EC 13;
- (j) in marine equipment, as set out in Directive 2014/90/EU 14 ;
- (k) in medical devices, as set out in Council Directive 93/42/EEC ¹⁵ or Regulation (EU) 2017/745 ¹⁶ and in vitro medical devices as set out in Directive 98/79/EC ¹⁷.

(b) Adopted regulation on lighting products

To define this type of equipment CEMA suggests to add the following exemption as part of Annex III of the regulation:

Annex III Exemptions

...

(f) In or on non-road mobile machinery as set out in Regulation (EU) 2016/1628 of the European Parliament and their trailers

(..) In or on interchangeable equipment as set out in Directive 2006/42/EC intended to be towed or to be mounted and fully raised from the ground or that cannot articulate around a vertical axis when the vehicle to which it is attached is in use on a road by vehicles as set out in Regulation (EU) No 167/2013

(c) Comment of the CEMA - p.2

Section A - Endocrine disrupting properties with respect to humans

 An active substance shall be identified as having endocrine disrupting properties with respect to humans if it is a substance that meets all of the following criteria:

Section A - Endocrine disrupting properties with respect to humans

- (1) it is known to cause an adverse effect relevant for human health, which is a change in the morphology, physiology, growth, development, reproduction, or, life span of an organism, system, or (sub)population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other influences:
- (1) A substance shall be considered as having endocrine disrupting properties that may cause adverse effect in humans if, based on points (a) to (d) of point (2), it is a substance that meets all of the following criteria, unless there is evidence demonstrating that the adverse effects identified are not relevant to humans:
 - (a) it shows an adverse effect in an intact organism or its progeny, which is a change in the morphology, physiology, growth, development, reproduction or life span of an organism, system or (sub)population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress or an increase in susceptibility to other influences:

(d) Draft regulation on endocrine disruptors

(e) Adopted regulation on endocrine disruptors

1) La Comisión propone identificar sólo los EDCs que provoquen efectos adversos "conocidos" en humanos y vida silvestre. La expresión "conocido" significa tener pruebas. El hecho de pedir esas pruebas debilita la legislación actual que exige regular aquellas sustancias que "puedan" causar daño (como para los cancerígenos se utiliza "que puedan causar cáncer"). Ese umbral de prueba tan elevado se opone a la opinión de los expertos sobre la probabilidad de un efecto y es inaceptable, ya que puede generar daños en humanos y medio ambiente antes de que se actúe, en contra del Principio de Precaución consagrado en los tratados de la UE. También choca con el actual enfoque de identificación y clasificación de sustancias cancerígenas y tóxicas para la reproducción, en base al nivel de evidencia. Las sustancias deben identificarse como EDCs cuando se conoce o presume que tienen efectos adversos. Sólo este enfoque es consistente y coherente con la ley de biocidas que dice que no se debe permitir en el mercado biocidas si "se considera que tienen propiedades de disrupción endocrina que puedan causar efectos adversos en humanos". La Comisión debe actuar con el conocimiento científico existente y la mejor opción para la salud es un enfoque con 3 categorías de acuerdo al nivel de evidencia

(f) Comment of a Spanish citizen

Figure 3: Examples of successful comments