City Research Online

City, University of London Institutional Repository

Exploring Information Exchange among Interest Groups: A Text-Reuse Approach

Stefano Pagliari
City, University of London
stefano.pagliari@city.ac.uk

Kevin L. Young
UMass Amherst
keviny@umass.edu

Abstract

The presence of communication channels plays a key role in how interest groups engage in EU policymaking. However, the capacity of researchers to explain these patterns is constrained by the informal and opaque nature of such interaction. In this paper we develop a novel text-as-data approach which maps the informal patterns of information exchange among the stakeholders that engage with EU policymaking by detecting instances of text reuse among the comment letters submitted by these groups to the same policy proposal. We use this approach to analyse a novel dataset of publicly available comments to a wide a range of EU policies. We find that there are significant differences between the structure of information exchange networks and more formal lobbying coalitions in the EU, as well as between the groups that engage in these forms of coordination.

1 Introduction

The extent to which the decision-making process is open to the participation of a broad range of voices and stakeholders has often been presented as an important component of the legitimacy of the regulatory governance in the EU (Judge and Thomson 2019). As a result, recent years have witnessed a sustained push to introduce a variety of consultative instruments such as open online consultations, stakeholder consultations or meetings with selected groups (Binderkrantz,

Blom-Hansen, and Senninger 2020) designed to provide opportunities for organized groups and citizens to weigh in on the policy initiatives originated in Brussels.

An important recent body of scholarship has investigated how EU regulatory bodies involve stakeholders through consultative processes (Arras and Braun 2018; Beyers and Arras 2019; Binderkrantz, Blom-Hansen, and Senninger 2020; Quittkat 2011), what are the factors driving the use of consultations by European institutions, (Van Ballaert 2017), and the impact that these mechanisms have over the design of policies within the EU (Bunea and Thomson 2015).

For all we do know about European regulatory governance, there is nevertheless still a lot we do not know about how stakeholders engage with these consultative mechanisms and the origins of information that is fed to policymakers through them. A number of studies have studied 'who' answers the call for public comments from EU authorities and investigated the number, type, and diversity of the groups that are involved in the EU policymaking process through various consultative mechanisms (Berkhout, Hanegraaff, and Braun 2017; Beyers and Arras 2019; Quittkat 2011; Rasmussen and Carroll 2013). One limitation with existing scholarship in this area is that it when it is not looking at population aggregates, it tends to examine interest groups individually and separately. Yet stakeholders' engagement in the design of public policy - European or otherwise - is rarely an individual endeavour. The process through which stakeholders select which consultations to engage with and develop a response is often the result of the interactions with other groups. As Mahoney (2007, 366) puts it, groups frequently 'share tidbits of information after committee meetings, forward emails with talking points, hold conference calls to get allies up-to-speed on policy-making developments, send joint letters to policy-makers'.

In this paper we develop a novel text-as-data approach to investigate information exchange among the stakeholders which relies on detecting instances of text reuse among the comment letters submitted by different groups to the same policy proposal. Ideas and position papers float around interest group communities that are engaged in a policy space, and text reuse allows us to observe which groups is sharing text with whom.

We deploy this approach to analyse the characteristics and determinants of the communication networks among the groups that engage with EU consultative mechanisms. To do so, we build a new dataset comprising all the publicly available comments submitted online from interest groups to a wide a range of different policy initiatives by the European Commission. By applying the text reuse approach to this corpus of online comments, we are able to analyse informal communication networks that emerge among stakeholders engaged with the EU-level

policymaking process and to answer the following questions: How do information flow across groups that responds to EU-level policy consultations? Which groups are engaged in the exchange of information regarding EU policies? And how do these informal communication networks differ from more formal lobbying coalitions?

Our analysis reveals that the way information flows across groups that engage with the EU-level policymaking process is in many ways distinctive from more formalized instances of coordination among the same actors. First, we show how information exchange networks connecting stakeholders in the EU tend to link fewer groups and have less staying power than formal lobbying coalitions but are also more likely to connect groups based in different countries. Then, we analyse statistically the conditions under which information exchange networks across stakeholders engaging with EU consultative processes are likely to emerge and compare them with formal lobbying coalitions. We find that the sharing of text is particularly associated with groups outside of the business community, although we don't find support for the notion that this is a 'weapon of the weak'. The likelihood of engaging in text reuse is also associated with the number of associational ties that a group has, thus providing support for the notion that trade associations and other groups act as important channels for the sharing of information among stakeholders. We conclude by suggesting further ways in which text reuse could be deployed to investigate the relationship among interest groups in the policymaking process.

2 Literature Review

An established body of literature has long recognized how a key way in which interest groups engage in the policymaking process is by working together. While this literature has primarily focused on the US context (Baumgartner et al. 2009; Hojnacki 1998), recent work has detailed how 'coalitional activities' have become a prominent feature of the way interest groups engage in advocacy vis-à-vis EU institutions as well (Beyers and Braun 2014; Beyers and De Bruycker 2018; Klüver 2013; Mahoney 2008; Pijnenburg 1998; Sorurbakhsh 2016). This literature has found that, in the world of EU policymaking, the ties that groups establish with others play a key role in in determining their preferences towards EU policies (Bunea 2014), the information provided to policymakers (Chalmers 2013), and the access that they have to EU institutions (Beyers and Braun 2014).

While research to date has generated important insights into the way interest groups work together in the policymaking process, one of the main difficulties associated with studying this issue is the fact that such coordination often varies significantly in the level of formality and openness (Heaney and Strickland 2016). Most scholarship to date has often focused on formal lobbying coalitions. These are easy to identify since they involve public declarations of the alliance among different groups and in some cases 'highly coordinated enterprises with logos, letterheads and secretariat' (Mahoney 2007, 366).

These are, however, only one way in which interest groups work together and coordinate their advocacy initiatives. Interaction among groups in the policy process often remain 'informal and loose' (Mahoney 2007, 366), based on the occasional information sharing and private communications with other groups that mobilize around the same issues. Numerous scholars have argued that information regarding a given policy frequently flows within 'communication networks that emerge from the interaction of interest group representatives', either in person or through telephones or e-mail (Heaney 2014, 68).

The existing literature has explored the reasons why groups engage in this type of exchange. Responding to an open consultation requires citizens, business groups, NGOs and other stakeholder to devote significant resources to monitor the policy agenda, understand the implications of proposed policy, and formulate a response By connecting with other groups and exchanging information, groups are able to lower the costs of responding to a consultation by drawing upon the expertise and resources of other groups (Baumgartner et al. 2009; Hojnacki 1998; Junk 2019a). Moreover, while the pursuit of formal lobbying coalitions requires groups to compromise on a common text endorsed by all the coalition partners (Hojnacki 1998; Holyoke 2009), coordination based on the informal exchange of information allows groups to retain greater autonomy to pursue the preferred policy position by selectively using information from other groups (Mahoney 2007).

The existing literature has established that informal communication networks play a key role in the way groups engage with the policymaking process in the EU and elsewhere and called for a 'better theoretical integration of the formal sides of collaboration (e.g., coalitions) with the informal sides (e.g., communication, trust)' (Heaney and Strickland 2016). For instance, we still lack clear answers to key questions such as: How do informal communication networks differ from more formal lobbying networks? Which groups are more likely to engage in information exchanges within informal networks rather than creating formal coalitions?

One of the main limitations in approaching these questions are the empirical mearures that scholars deploy. Three main empirical approaches are frequently found in this literature. First, different studies have mapped coalitional activities among interest groups by collecting information on existing formal memberships within different associations (Bunea 2014; Chalmers 2018; Chalmers and Young 2020). As Bunea (2014, 1232) acknowledges, however, 'this estimation technique suffers… from the obvious neglect of informal communication or co-operation ties that most probably characterized the interactions of interest groups at the time of the event'. The mere presence of common associational ties between distinct groups does not necessarily imply that these will be 'active' and that information will flow from the association to its members or between different members.

Second, a more direct way to capture instances when groups have worked together on an issue relies on the analysis of the patterns of co-signing of the same document by different groups. For instance Box-Steffensmeier and Christenson (2014) explore alliances by mapping when interest groups co-sign amicus curiae briefs before the US Supreme Court, while Pagliari (2018) maps co-signing of response letters to policy consultations by US financial regulatory agencies. According to Box-Steffensmeier and Christenson (2014) the advantage of this approach is that it is 'culled from the actual, purposive and coordinated work of interest groups' and 'it also comes close to a complete network of the population of interest'. Also, in this case, however, the focus on mapping public linkages between interest groups captured through the co-signing of documents misses informal and less visible information exchanges among different groups.

Third, the most common way in which existing studies of have tracked the existence of informal networking and coordination among groups in the EU (Beyers and De Bruycker 2018; Chalmers 2013; Mahoney 2007; Sorurbakhsh 2016) and elsewhere (Hojnacki 1998; Beyers and Braun 2014; Junk 2019b)relies on surveys or interviews of representatives from interest groups active on a given issue. This approach has the advantage of allowing researchers to capture actual channels of communication among groups on a given issue, including those that remain informal. However, surveying interest groups remains highly labour intensive (Junk 2019b), forcing researchers to focus on a single or a small number of policy issues while trying to generalize from those. It is also arguably subject to limitations such as 'memory fatigue' due to the distance in time (Holyoke 2009), 'staff turnover' within the organization and low response rates among the interview population (Bunea 2014, 123).

In the next section we outline a novel text-as-data approach that aims to capture informal channels of coordination across the entire population of interest groups active on a given policy space.

3 Method

This section outlines an approach to map the patterns of information exchange among groups by mapping instances of text reuse between the responses submitted by interest groups to the same policy consultation. This is certainly not the first study to rely on this type of textual data to study the relation among groups in the policymaking process. For instance, different studies have hand-coded the text of the responses of different groups to policy consultations of these responses against a coding scheme, for instance to identify those groups calling for a more stringent or weaker policy response (Yackee and Yackee 2006) . The resource intensive nature of manual coding and issues of reliability have also pushed scholars to explore other computer-aided approaches to content analysis. In her studies of coalitions in European politics, Klüver (2013) extracts the policy position of different interest groups based on the relative frequency of words used within and across the texts produced by different groups in response to a given policy consultation (Slapin and Proksch 2008). As Klüver (2013, 54) acknowledges, for groups to share the same policy position, 'it is not necessary that interest groups formally cooperate or exchange information'. In other words, as Junk (2019b, 660) argues, existing approaches studying focusing on the text produced by different groups have captured the presence of 'camps' or 'sides' to an issue, but 'they typically overlook active cooperation between these actors'.

In order to identify instances of 'active' cooperation (Junk 2019b) whereas groups interact and exchange information in the pursuit of a common policy objective, we turn to a different approach and investigate patterns of text reuse across different submissions. As interest groups work together behind the scenes and share information regarding a policy or develop a coordinated response, we would expect some of the same text to be reused verbatim or with only limited changes in their respective responses to the same consultation. For instance, Figure 1 shows the first page of the letters submitted two letters submitted by two distinct groups in response to same policy.
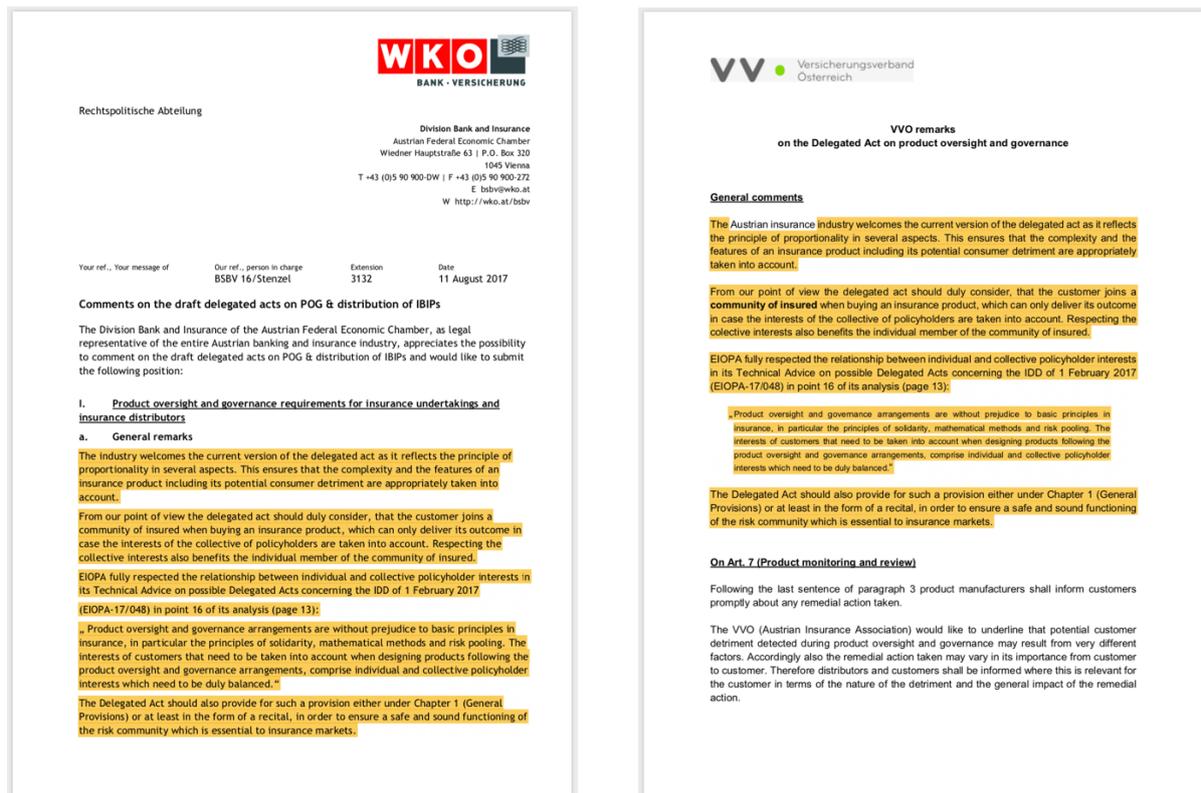
Figure 1: Letters submitted by the Austrian Federal Economic Chamber (on the left) and the Austrian Insurance Industry Association (on the right) on the 'Delegated Act on conduct of business rules for the distribution of insurance-based investment products' published by the European Commission in 2017. Sentences including text reuse were highlighted by the authors to facilitate visualization.

The two letters are submitted independently from two different business associations and they differ both in the overall length as well as in the content for large part of the text. At the same time, as the highlighted sections illustrate, a number of paragraphs are reused verbatim or with only limited alteration across the two documents For instance, two otherwise identical sentences starts with 'The Austrian insurance industry welcomes…' in one document but only 'The industry welcomes' in the other one.

In order to extract from the text of the responses to the same consultatioin instances where similar sequences of text are reused, we turn to a quantitative text analysis approach known as 'substring matching'. In contrast with so-called 'bag of words' approaches whereby distribution of word frequencies in the texts produced by different groups are analysed without taking into account the location of these words in the text (Klüver 2013; Slapin and Proksch 2008), 'substring matching' approaches take the ordering of text into account when comparing different documents, thus enabling us to capture shared sequences of words across documents (Acree et al. 2016). A common commercial application of this approach can be found in the plagiarism detection pieces of software.

This is an increasingly popular tool for 'understanding political influence by examining to what degree one elite actor's ideas are copied by another' (Acree et al. 2016). For instances, recent studies have adopted this approach to map the flow of ideas and to identify changes among different policy initiatives, such as between bills introduced within Congress (Wilkerson, Smith, and Stramp 2015), between the policies introduced in different state legislatures (Linder et al. 2018), between the proposed and final rules by US regulatory agencies (Haeder and Yackee 2015), between interest groups and the Supreme Court (Collins, Corley, and Hamner 2015), and between interest group lobbyists and sitting Members of Congress (McKay 2018).

In order to identify these instances of text reuse we compare all pairwise combinations of sentences from different responses submitted to the same policy consultation by using the Smith-Waterman algorithm through the 'textreuse' package in R (Mullen 2016). Originally introduced in molecular biology to find similar sub-sequences of genes within longer strings (Smith and Waterman 1981), this algorithm has in recent years found use within the literature to find matching sequences of text within two documents (for a recent application of the same algorithm in political science see Linder et al. 2018; Wilkerson, Smith, and Stramp 2015).

An important advantage of this approach is that it allows us to identify similar passages of text between two documents that are 'imperfect' matches; that is, sequences of text that are similar but not completely overlapping. This is important to identify instances where groups formulated their position starting from the same common text but adapted in part. Moreover, this approach does not require us to limit the analysis to the documents in a single language. Instead, we are able to capture instances where responses written in any language share text with others written in the same language. This is particularly important when it comes to analysing patterns of coordination in a multi-lingual setting such as the EU policymaking context. At the same time, this approach would not be able to identify instances where a group translates the text produced by a different group into a different language. A detailed discussion of the operationalization of this approach can be found in the Appendix.

The presence of text-reuse among different letters signed by separate groups does not allow us to identify the specific role that each group played in the information exchange. Text reuse could be the product of different groups equally contributing to the formulation of a shared policy position, but it could also be the result of one group feeding information to others in order to solicit their mobilization and gain allies. Moreover, the presence of text re-use could also be the result of two groups both citing a contribution by a third-party, such as in the case two groups including in their

individual submission language provided by a trade association they are both members of. While it is not possible to pinpoint the specific type of engagement between two groups, we consider the presence of text reuse across their submissions to the same consultation as evidence that their policy responses have not been developed independently but rather they were informed by the use of the same information.

It is important to note that our approach only allows us to analyse the position of those groups that submitting a comment letter to a given policy. As a result, we are not able to capture those instances where the coordination and information exchange among two different groups is not followed by the mobilization of both groups, such as in cases when a peak association speaks on behalf of its members (see Chalmers 2018). As a result, our approach may underestimate the level of information exchange among stakeholders around a certain policy issue.

A simple search for text reuse among two responses to the same consultation could also significantly overestimate the level of coordination and information exchange among the two groups. We take a number of steps to identify 'false positives'. such as instances where different respondents use common expressions that cannot be attributed to information exchange or cite the same text from the consultation document or another document. For instance, in the example in Figure 1, both letters cite verbatim a passage from an EU regulatory authority (EIOPA). A detailed breakdown of the steps taken identify and exclude these false positives and to validate the results is provided in the Appendix.

4 Data

We use this approach to analyse a novel dataset comprising publicly available comments provided by interest groups in response to many policy initiatives by the European Commission. In constructing this corpus, we take advantage of a recent innovation in the EU policymaking: the '2015 Better Regulation Agenda' (Golberg 2018). This initiative has expanded the number of policy decisions open to input from stakeholders, thus allowing us to capture the mobilization of interest groups across a broad range of topics, as well as the entire policy-cycle, from roadmaps outlining new policies to secondary legislation and the evaluations of existing policies (a breakdown of the policies included in our corpus is in Appendix. The feedback from stakeholders are submitted and made publicly available through a common webpage called *'Have Your Say'*. We used webscraping tools to collect all the comments submitted on this platform in the period between July 2016 and March 2019 across 642 policy initiatives. The number of responses vary

significantly across policy initiatives, ranging from only 1 comment for many policy initiatives to a maximum of 421 (the average number of responses per policy is 12.4). In the cases when respondents also upload a separate document, we also downloaded this document and added its text to that of the online comment submitted through the platform .

In order to explore the patterns of information exchange across organized groups, we focus on a subset of these comments. More specifically, we first exclude the responses from those actors that self-identified as 'individuals' (45.5% of the comments) and only considers only those comments from organizations. Second, we only include those groups we could identify in the 'Transparency Register'. The Transparency Register has been described as the most comprehensive register available of organisations seeking to directly or indirectly influence the formulation or implementation of EU policies (Berkhout et al. 2018) While the registration to the Transparency Register is voluntary, in practice this has become a 'de facto mandatory' (Bunea 2018) tool as access to Commission and European Parliament decision makers is conditional upon joining the Register and it has reached a satisfactory level of coverage and quality of entries (Greenwood and Dreger 2013). Limiting our analysis to the groups registered in the Transparency Register allow us to take advantage of a rich body of demographic information groups submit when registering, although this comes at the cost of excluding from the analysis some of the groups that do not engage with EU policymaking frequently enough to justify registering on this platform.

This data collection resulted in the creation of a corpus comprising 4222 different comments submitted by 1796 distinct groups. These represent 12.4% of all the groups that have registered in the Transparency Register between January 2015 and January 2019. The majority of the groups that comprise our dataset are firms and business associations, in line with the findings of existing studies that have detailed the presence of a strong 'business bias' in responses to EU online consultations (Rasmussen and Carroll 2013). A summary of the characteristics of the respondents included in our dataset is in Appendix.

5 Descriptive Statistics

What do the patterns of information exchange among stakeholders engaging with EU policies look like? We first tackle this question by reviewing some descriptive statistics concerning the patterns of text reuse detected by our approach.

Our data collection strategy identified text reuse across the letters submitted by 302 different groups. This represents 17.18% of the 1758 submitting a comment which we could identify within

the Transparency Register. Text reuse with other submissions to the same consultation is found in 10.02% of all the letters, thus suggesting that while this practice is a regular occurrence in the way stakeholders engage with EU consultations, it is detected only in a minority of the submissions.

To provide further insights into how this form of coordination differs from more public and formal lobbying coalitions, we identify those the letters submitted in response to the same policies that were co-signed by multiple groups (Box-Steffensmeier and Christenson 2014). This comparison reveals how the number of groups who engage in informal coordination as measured via text re-use (302 groups) is comparable with the number of those that co-sign one or more letters (280 groups). At the same time, the configuration of these ties varies significantly across the two forms of coordination along different dimensions, summarized in Figure 2.
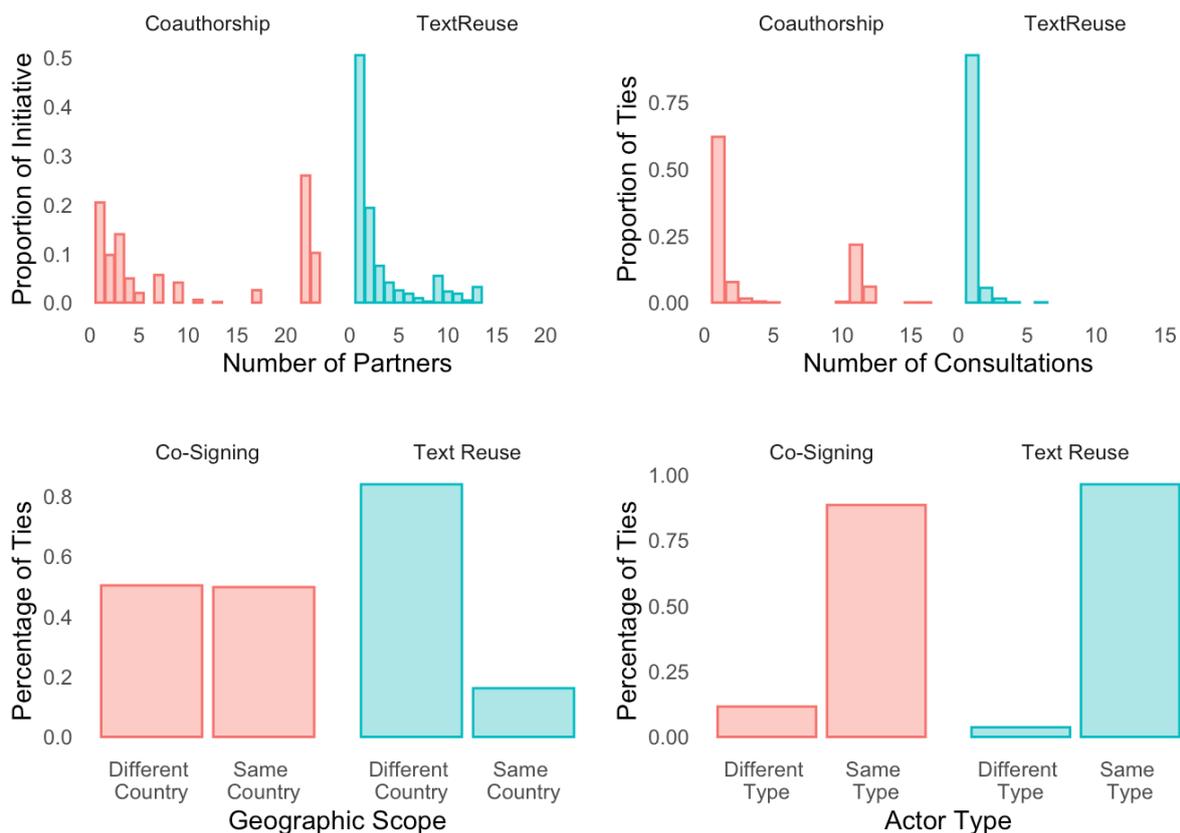
Figure 2: Descriptive statistics of the size (top-left), longevity (top-right), geographic heterogeneity (bottom-left) and sectoral heterogeneity (bottom-right) of informal (measured via text reuse) vs. formal coalitions (measured via co-signing)

First, one pattern emerging concerns the size of different types of coalitions (top-left in Figure 2). Formal coalitions mapped through co-signing of the same letter tend to be larger than information exchange networks mapped through text reuse across different letters. More specifically, when a group co-signs a letter, this is tied - on average - with 10.4 other different groups signing the same letter. The average, however, disguises the fact that the smaller formal lobbying coalitions coexist with much larger ones, bringing together more than 15 co-signatories. Co-signing of the same letter allows interest groups to signal to policymakers the breadth of support that this position enjoys among multiple groups (Nelson and Yackee 2012), thus creating incentives to form large coalitions. Yet within each consultation, a group either sending or receiving text is - on average - tied to 3 other groups. The limited size of informational coalitions based on text reuse is consistent with a different dynamic driving this form of coordination, that is, to acquire relevant information from groups with the relevant expertise rather than maximize the size of the coalition.

Second, information exchange networks differ from formal lobbying coalitions in their longevity across policy battles (top-right in Figure 2). In particular, 37.92% of the combinations of groups that co-sign a letter in response to a policy also extends their collaboration in other policy battles, in some cases co-signing letters in response to more than 10 different policy consultations. While the creation of a formal coalition may entail higher initial transaction costs, once this is in place the same groups tend to lend each other support also in subsequent policy initiatives. Instead, text reuse among the same two interest groups is very unlikely to be found in more than one policy initiative, and only 7.37% of the groups that coordinate informally via text reuse do so across different issues. These results provide support to the insights from the existing literature on how informal coalitions where members retain a high level of autonomy tend to emerge for the duration of a single legislative or regulatory debate (Mahoney 2007; Pijnenburg 1998).

Finally, we investigate how the structure of these different ties varies according to some key characteristics of the groups that engage in coordination. Both in the case of information exchange networks mapped via text reuse and formal lobbying coalitions mapped via co-signing, ties are almost exclusively linking actors of the same type based on the 6 main group categories in the Transparency Register (business, professional services/law, NGOs, research/think thanks, religious entities, public authorities) (bottom-left in Figure 2). 'Strange bedfellows' (Beyers and De Bruycker 2018; Phinney 2017) coalitions such as those linking business interests and non-business remain the exception. This suggests that despite the lower transaction costs associated with the creation of ties based on the exchange of information, stakeholders remain more likely to coordinate with similar entities that are more likely to have similar preferences and goals (Fischer and Sciarini 2016).

One dimension where the two forms of coordination vary significantly, however, is the geographic scope (bottom-right in Figure 2). Organisational coordination through sharing and copying text is characterized by far less evidence of distinct national-level clustering. While 49.72% of the ties among co-signatories to the same letter are based in the same country, only 16.16% of the ties from text reuse are from two groups based in the same country. This finding is consistent with the notion that the formation of informal coordination implies lower transaction costs between organizations, thus facilitating connectivity between more organizations over greater geographical distances.

6 Correlates of Text Reuse and Co-Signing

This section expands the analysis above by investigating what groups engage in information exchange networks, and how does this differ from more formal lobbying coalitions.

Building upon the literature on the determinants of coalitions among interest groups, we identify a range of explanatory variables related to the attributes of the group in question as well as the characteristics of the policy. We specify two separate regression models, whereby each observation represents one of the individual responses to the policy consultations that comprise our dataset. Since the aim of the analysis is to investigate the determinants of different forms of coordination, when multiple groups co-sign a letter each signatory is entered separately in our analysis for that policy consultation. The outcome variable is coded as 1 whether either co-signing or text reuse can be found in the letter, and 0 otherwise. Because of the nested nature of the data, in which each letter is part of a consultation, we used a multi-level logit model, whereby the variables at the consultation are modelled as a distinct level compared to the actor-level variables. Practically, this entails estimating a mixed effects logistic regression with the consultation identifier associated with a random intercept. Because we use a multilevel logit with fixed and random effects, the confidence interval estimates are themselves variable. We thus ran 100 bootstrap estimations of each model in order to ensure greater precision in the standard error estimates. To assess multicollinearity, we checked for variance inflation in a normal logit model. Using a multilevel model in this way helps to address some of the unique structures in these data. However, there are still unmodeled interdependencies in the data that need to be acknowledged, because a tie that any actor forms is

necessarily a tie with another actor in the data.[1] We acknowledge that one limitation is potential autocorrelation, and the estimated confidence intervals should be taken with caution.[1]

We report the logit coefficient estimates in the regression table in the Appendix, and report the odds ratios in Figure 3 for a better substantive understanding of these estimated effects below. We found no high pairwise correlation among our explanatory variables, low variance inflation factors across these models. Summary statistics and correlation table are provided in the Appendix.
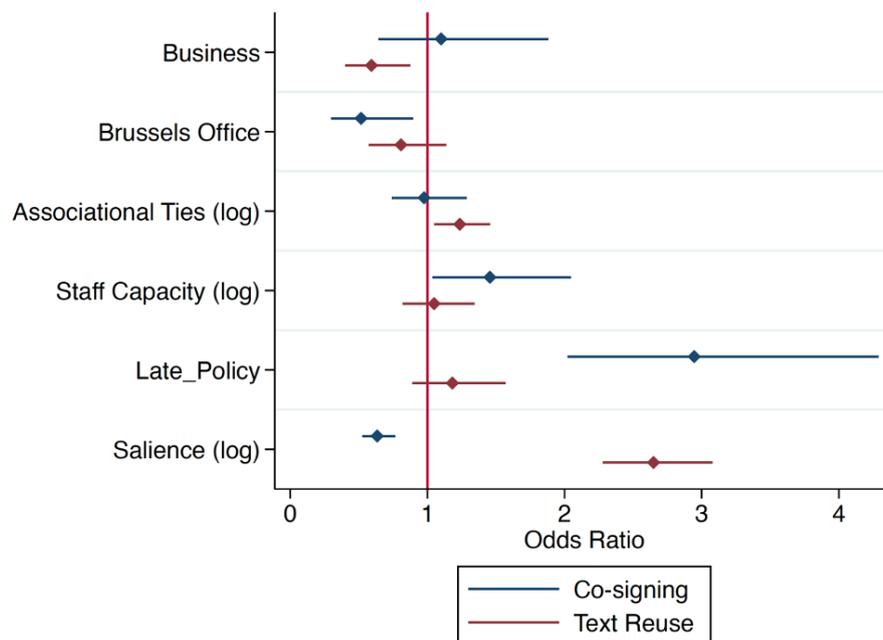
Figure 3: Multilevel Logit Odds Ratio

First, we explore the sectoral provenience of the respondents. Some studies have identified business groups as more likely to form coalitions, while others have found that NGOs and groups representing social or public interests are the ones more likely to join forces into coalitions (Beyers and De Bruycker 2018; Hojnacki 1998; Klüver 2013). We introduce a dummy variable to indicate whether the interest group is a business interest group or not according to the categorization on the Transparency Register. Our results show that the presence of text reuse is more likely to be

---

[1] The way to address this systematically is to analyse tie formation explicitly, using a network analysis method such as quadratic assignment procedure (QAP). We found QAP to be a non-viable option. Given the very large number of actors, and the extremely small proportion of ties, the outcome networks (whether for co-signing or text reuse) were so sparse that they yielded computational problems and highly inconsistent results – a problem made more severe by the fact that QAP operates via repeated iterations, in order to simulate p-values for the interpretation of statistical significance.

found outside of the business community, while this pattern is not found in the case of co-signing. However, this effect is not robust across different measures of text reuse (see below).

Second, existing studies have drawn a link between the resources of a group and the likelihood of joining a coalition. While for some scholars groups with more resources are more likely to be able to build coalitions (Beyers and De Bruycker 2018; Box-Steffensmeier and Christenson 2014; Carpenter, Esterling, and Lazer 2004; Sorurbakhsh 2016), others have claimed that coalition-building remains a primarily 'weapon of the weak' employed by resource-poor groups (Hanegraaff and Pritoni 2019; see also Junk 2019a). We thus include a measure of the number of full-time equivalent staff at each group (which is measured as the log of these values as there are large outliers).

Our results find that groups that engage in co-signing of the letter are statistically associated with greater resources, but this is not the case for text reuse. However, our approach does not allow us to differentiate between the recipients and the source of the information. As a result, it is not possible to discount the possibility that different levels of resources are associated with different roles within an information exchange network.

Third, the participation in common organizations, such as business associations and other interest group organizations represents a critical mechanism through which information can be expected to flow across groups (Kowal 2018) and groups can be exposed to peer pressures (Bunea 2014). We extract from the Transparency Register submission of each group information on what associations they are members of or who are their members in the case of associations. The number of association ties is included as a way of proxying for the level of connectivity that an interest group has in formal associational networks (which is measured as the log of these values as there are large outliers).

We find that the number of associational ties that a group has is correlated with the presence of text reuse in its responses to EU policies, thus providing support to the notion that trade associations and other groups act as conduits for the sharing of information (Kowal 2018). At the same time, the structure of our data does not allow us to pinpoint whether information flows directly between associations and their members or more diffuse mechanisms are in place. Associational ties instead do not affect the likelihood that a group will co-sign a letter, although it is not possible to observe through this analysis cases where groups delegate this form of coordination to an association that they are members of (Chalmers 2018).

Fourth, the rising importance of Brussels as a decision-making site for EU level advocacy could be affecting the capacity of groups to form coalitions, as hypothesized in some existing literature (Bunea 2014; Sorurbakhsh 2016). Sorurbakhsh (2016, 208) argues that 'groups in Brussels will also find it more convenient to coordinate with other groups who also have offices nearby'. We thus generated a variable which indicates if a group has an office in Brussels. We find that having a Brussels office makes no difference for text re-use, but it makes a negative difference for co-signing.

Fifth, we control for two key consultation-level conditions. Policies that are salient and attract significant levels of mobilization among stakeholders may give groups an incentive to join forces to share resources or signal the broad base of support for their position relative to their competitors (Hojnacki 1997; Mahoney 2007). We thus control for the total numbers of letters submitted in response to that specific letter (Junk 2019a). This variable is statistically significant for co-signing yet the coefficient is negative, contrary to the expectation that coalitions are more likely to emerge in contested policies. The same variable instead has a positive relationship with the presence of text reuse. Our analysis however does not allow us to establish the direction of the causal arrow between the level of text reuse and the number of respondents, as it remains possible that the presence of pre-existing communication networks coordination may lower the cost for groups to mobilize and increase the number of comment letters.

We also control for the position of a given consultation in the policymaking process. In particular, it is reasonable to expect that as a policy makes progress through the different stages of the policymaking process, stakeholders will be able to observe the position of other groups and identify potential coalition partners. We generate a dummy variable that is coded as 0 when groups are responding to a policy proposal and consultation document, and as 1 when groups are responding to a final rule-making or implementation measure. We find that the stage of the EU policy that interest groups weigh in on has a large and significant association with co-signing, but not for text reuse.

As an additional check on our results, we re-ran the text reuse regressions with a different threshold for the dependent variable with a differently calibrated text reuse approach. As shown in the Appendix, there is a negligible difference between the results across these two variations, with the exception of the business actor variable.

7 Conclusion

In his review of existing scholarship on networks and interest groups, Heaney calls for pushing research in this area by exploring 'new kinds of data, especially texts that are amenable to content analysis', noting that '[i]n general, the study of social networks could benefit by seeking more synergy with advances in content analysis' (Heaney and Strickland 2016, 446). This paper represents a response to such a call, as we have used text-based analytical tools to establish new measures for the study of how information flows across interest groups in the policymaking process.

Our text reuse approach has enabled us to investigate the way stakeholders interact when responding to EU-level policies and the origin of the information that flows into EU consultative mechanisms. Our analysis provides empirical support for the notion that the way information flows differs from the structure of more formal lobbying coalitions. Our analysis finds that informal communication networks tend to be smaller in size and have less longevity across different policy battles than formal lobbying coalitions, but are also more likely to connect groups from different countries. At the same time, information tends to be exchanged primarily among groups of the same type.

We also find important differences regarding the characteristics of the groups that engage in information exchange networks compared to more formal lobbying coalitions. Informal coordination based on the sharing of text is particularly associated with groups outside of the business community, as well as with the presence of ties with formal associations which remain an important conduit through which information flows across groups. Our analysis also suggests that the prevalence of different forms of coordination also varies across different kinds of policy initiatives.

To conclude, we want to suggest areas of further research to address some of the limitations of this analysis. First, the existing literature has highlighted how groups often play very different roles within coalitions. While some coalition leaders spending significant resources to generate policy-relevant information and to disseminate it among potential allies, others are more passive recipients (Heaney and Leifeld 2018; Junk 2019a). The analysis in this paper does not allow us to differentiate between the specific role that different groups play within an information exchange network. Further work could expand this approach to other forms of interest group communication which are more clearly time-stamped, thus allowing researchers to identify when groups are the source or the recipient of the information being exchanged.

Second, the limited time window covered by our dataset means that we have only been able to provide a static account of the characteristics and determinants of information exchange among

interest groups. However, a key determinant of the decision of a group to work alone or with another group on a given issue is the experience from previous coalitional activities (Hojnacki 1998). Further work is therefore needed to explore how the informal forms of coordination captured through text reuse evolve over time. Moreover, while the existing study has compared the determinants of informal communication networks with formal lobbying coalitions, further work could utilize the approach developed in the paper to better integrate the two sides and analyse - as Heaney argues - 'the mechanisms through which interest groups convert their loose affiliations into collective action' (Heaney and Strickland 2016, 441).

Third, the respondent base from which our analysis is drawn means that most of the organizations being analysed are predominantly business groups. This conforms to established expectations in the literature (Rasmussen and Carroll (2013)) but also raises the prospect that coordination patterns among non-business groups alone may be different than those analysed here.

Finally, our analysis has investigated the determinants of text reuse among responding to EU policy consultations, but it has not examined the nature and type of information that is being exchanged. Further work should extend this approach and leverage our text-as-data approach to explore what type of information and ideas are circulated among groups (Chalmers 2013) rather than the existence of exchange per se. With the use of text analysis tools such as the ones used in this paper and others, pursuing such dynamics remains an inviting possibility open to future researchers.

8 Bibliography

Acree, Brice et al. 2016. 'Comparing and Evaluating Cosine Similarity Scores, Weighted Cosine Similarity Scores, & Substring Matching.'

Arras, Sarah, and Caelesta Braun. 2018. 'Stakeholders wanted! Why and how European Union agencies involve non-state stakeholders.' *Journal of European Public Policy* 25(9): 1257–1275.

Baumgartner, Frank R et al. 2009. *Lobbying and Policy Change. Who Wins, Who Loses, and Why.* Chicago: University of Chicago Press.

Berkhout, Joost et al. 2018. 'Making Inference across Mobilisation and Influence Research: Comparing Top-Down and Bottom-Up Mapping of Interest Systems.' *Political Studies* 66(1): 43–62.

Berkhout, Joost, Marcel Hanegraaff, and Caelesta Braun. 2017. 'Is the EU different? Comparing the diversity of national and EU-level systems of interest organisations.' *West European Politics* 40(5): 1109–1131.

Beyers, Jan, and Sarah Arras. 2019. 'Who feeds information to regulators? Stakeholder diversity in European Union regulatory agency consultations.' *Journal of Public Policy*: 1–26.

Beyers, Jan, and Caelesta Braun. 2014. 'Ties that count: Explaining interest group access to policymakers.' *Journal of Public Policy* 34(1): 93–121.

Beyers, Jan, and Iskander De Bruycker. 2018. 'Lobbying Makes (Strange) Bedfellows: Explaining the Formation and Composition of Lobbying Coalitions in EU Legislative Politics.' *Political Studies* 66(4): 959–984.

Binderkrantz, Anne Skorkjær, Jens Blom-Hansen, and Roman Senninger. 2020. 'Countering bias? The EU Commission's consultation with interest groups.' *Journal of European Public Policy*: 1–20.

Box-Steffensmeier, Janet M., and Dino P. Christenson. 2014. 'The evolution and formation of amicus curiae networks.' *Social Networks* 36: 82–96.

Bunea, Adriana. 2014. 'Explaining Interest Groups' Articulation of Policy Preferences in the European Commission's Open Consultations: An Analysis of the Environmental Policy Area: Explaining interest groups' articulation of policy preferences.' *JCMS: Journal of Common Market Studies* 52(6): 1224–1241.

Bunea, Adriana. 2018. 'Legitimacy through targeted transparency? Regulatory effectiveness and sustainability of lobbying regulation in the European Union.' *European Journal of Political Research* 57(2): 378–403.

Bunea, Adriana, and Robert Thomson. 2015. 'Consultations with Interest Groups and the Empowerment of Executives: Evidence from the European Union: Consultations with Interest Groups.' *Governance* 28(4): 517–531.

Burgess, Matthew et al. 2016. 'The Legislative Influence Detector: Finding Text Reuse in State Legislation.' In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, San Francisco, California, USA: ACM Press, p. 57–66.

Carpenter, Daniel P., Kevin M. Esterling, and David M. J. Lazer. 2004. 'Friends, Brokers, and Transitivity: Who Informs Whom in Washington Politics?' *The Journal of Politics* 66(1): 224–246.

Chalmers, Adam W. 2013. 'With a lot of help from their friends: Explaining the social logic of informational lobbying in the European Union.' *European Union Politics* 14(4): 475–496.

Chalmers, Adam W. 2018. 'Unity and conflict: Explaining financial industry lobbying success in European Union public consultations: Financial industry lobbying.' *Regulation & Governance*.

Chalmers, Adam William, and Kevin L Young. 2020. 'Network effects in the formation of the financial industry's regulatory preferences in the European Union.' *Business and Politics* 22(1): 52–84.

Collins, Paul M., Pamela C. Corley, and Jesse Hamner. 2015. 'The Influence of Amicus Curiae Briefs on U.S. Supreme Court Opinion Content: The Influence of Amicus Curiae.' *Law & Society Review* 49(4): 917–944.

Fischer, Manuel, and Pascal Sciarini. 2016. 'Drivers of Collaboration in Political Decision Making: A Cross-Sector Perspective.' *The Journal of Politics* 78(1): 63–74.

Golberg, Elizabeth. 2018. '"Better Regulation': European Union Style.' *Mossavar-Rahmani Center for Business & Government Working Paper* 98: 109.

Greenwood, Justin, and Joanna Dreger. 2013. 'The Transparency Register: A European vanguard of strong lobby regulation?' *Interest Groups & Advocacy* 2(2): 139–162.

Haeder, Simon F., and Susan Webb Yackee. 2015. 'Influence and the Administrative Process: Lobbying the U.S. President's Office of Management and Budget.' *American Political Science Review* 109(03): 507–522.

Hanegraaff, Marcel, and Andrea Pritoni. 2019. 'United in fear: Interest group coalition formation as a weapon of the weak?' *European Union Politics*: 1465116518824022.

Heaney, Michael T. 2014. 'Multiplex networks and interest group influence reputation: An exponential random graph model.' *Social Networks* 36: 66–81.

Heaney, Michael T., and Philip Leifeld. 2018. 'Contributions by Interest Groups to Lobbying Coalitions.' *The Journal of Politics* 80(2): 494–509.

Heaney, Michael T., and James M. Strickland. 2016. 'A Network Approach to Interest Group Politics.' In *The Oxford Handbook of Political Networks*, eds. Jennifer Nicoll Victor, Alexander H. Montgomery, and Mark Lubell. Oxford University Press.

Hojnacki, Marie. 1997. 'Interest Groups' Decisions to Join Alliances or Work Alone.' *American Journal of Political Science* 41(1): 61.

Hojnacki, Marie. 1998. 'Organized Interests' Advocacy Behavior in Alliances.' *Political Research Quarterly* 51(2): 437.

Holyoke, Thomas T. 2009. 'Interest Group Competition and Coalition Formation.' *American Journal of Political Science* 53(2): 360–375.

Judge, Andrew, and Robert Thomson. 2019. 'The responsiveness of legislative actors to stakeholders' demands in the European Union.' *Journal of European Public Policy* 26(5): 676–695.

Junk, Wiebke Marie. 2019a. 'Co-operation as currency: How active coalitions affect lobbying success.' *Journal of European Public Policy*: 1–20.

Junk, Wiebke Marie. 2019b. 'When Diversity Works: The Effects of Coalition Composition on the Success of Lobbying Coalitions.' *American Journal of Political Science*: ajps.12437.

Klüver, Heike. 2013. *Lobbying in the European Union: Interest Groups, Lobbying Coalitions, and Policy Change*. Oxford University Press.

Kowal, Michael S. 2018. 'Corporate politicking, together: Trade association ties, lobbying, and campaign giving.' *Business and Politics* 20(1): 98–131.

Linder, Fridolin et al. 2018. 'Text as Policy: Measuring Policy Similarity through Bill Text Reuse.' *Policy Studies Journal*.

Mahoney, Christine. 2008. *Brussels Versus the Beltway: Advocacy in the United States and the European Union*. Washington, D.C.: Georgetown University Press.

Mahoney, Christine. 2007. 'Networking vs. Allying: The decision of interest groups to join coalitions in the US and the EU.' *Journal of European Public Policy* 14(3): 366–383.

McKay, Amy Melissa. 2018. 'Fundraising for Favors? Linking Lobbyist-Hosted Fundraisers to Legislative Benefits.' *Political Research Quarterly* 71(4): 869–880.

Mullen, Lincoln. 2016. 'Textreuse: Detect Text Reuse and Document Similarity.' *R package version 0.1.4*.

Nelson, David, and Susan Webb Yackee. 2012. 'Lobbying Coalitions and Government Policy Change: An Analysis of Federal Agency Rulemaking.' *The Journal of Politics* 74(2): 339–353.

Pagliari, Stefano. 2018. 'The Second Half. Interest Group Conflicts and Coalitions in the Implementation of the Dodd-Frank Act Derivatives Rules.' In *Governing the world's biggest market: The politics of derivatives regulation after the 2008 crisis*, eds. Eric Helleiner, Stefano Pagliari, and Irene Spagna. Oxford University Press.

Phinney, Robin. 2017. *Strange Bedfellows: Interest Group Coalitions, Diverse Partners, and Influence in American Social Policy*. Cambridge: Cambridge University Press.

Pijnenburg, Bert. 1998. 'EU lobbying by ad hoc coalitions: An exploratory case study.' *Journal of European Public Policy* 5(2): 303–321.

Quittkat, Christine. 2011. 'The European Commission' s Online Consultations : A Success Story?' *Journal of Common Market Studies* 49(3): 653–674.

Rasmussen, Anne, and Brendan J. Carroll. 2013. 'Determinants of Upper-Class Dominance in the Heavenly Chorus: Lessons from European Union Online Consultations.' *British Journal of Political Science* 44(02): 445–459.

Slapin, Jonathan B., and Sven-Oliver Proksch. 2008. 'A Scaling Model for Estimating Time-Series Party Positions from Texts.' *American Journal of Political Science* 52(3): 705–722.

Smith, Temple F, and Michael S Waterman. 1981. 'Identification of common molecular subsequences.' *Journal of molecular biology* 147(1): 195–197.

Sorurbakhsh, Laila. 2016. 'Interest group coalitions and lobbying environments: Toward a new theoretical perspective.' *Interest Groups & Advocacy* 5(3): 200–223.

Van Ballaert, Bart. 2017. 'The European Commission's use of consultation during policy formulation: The effects of policy characteristics.' *European Union Politics* 18(3): 406–423.

Wilkerson, John, David Smith, and Nicholas Stramp. 2015. 'Tracing the Flow of Policy Ideas in Legislatures: A Text Reuse Approach.' *American Journal of Political Science* 59(4): 943–956.

Yackee, Jason Webb, and Susan Webb Yackee. 2006. 'A Bias Towards Business? Assessing Interest Group Influence on the U.S. Bureaucracy.' *The Journal of Politics* 68(01): 128–139.

Appendix

1 Descriptive Statistics – Consultations
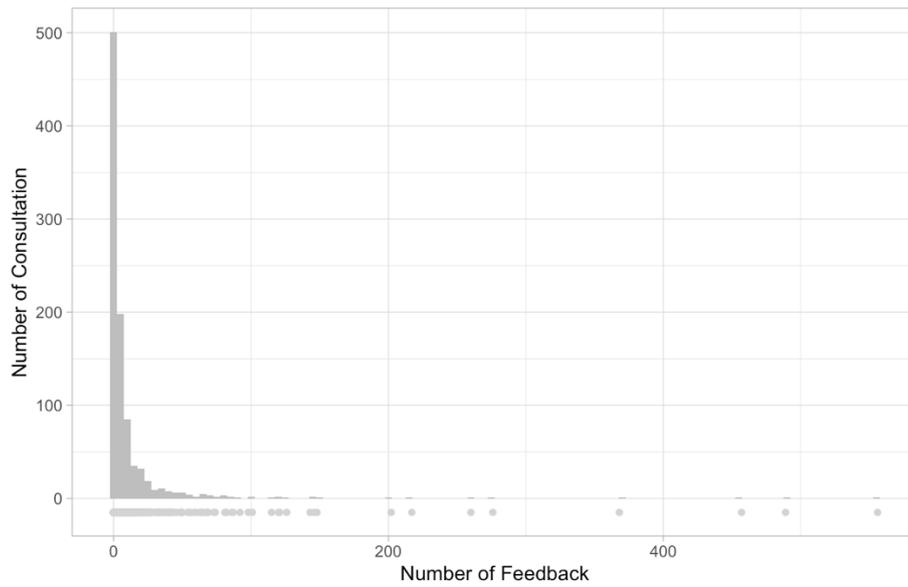
**Figure A.1:** Number of Responses per Consultation
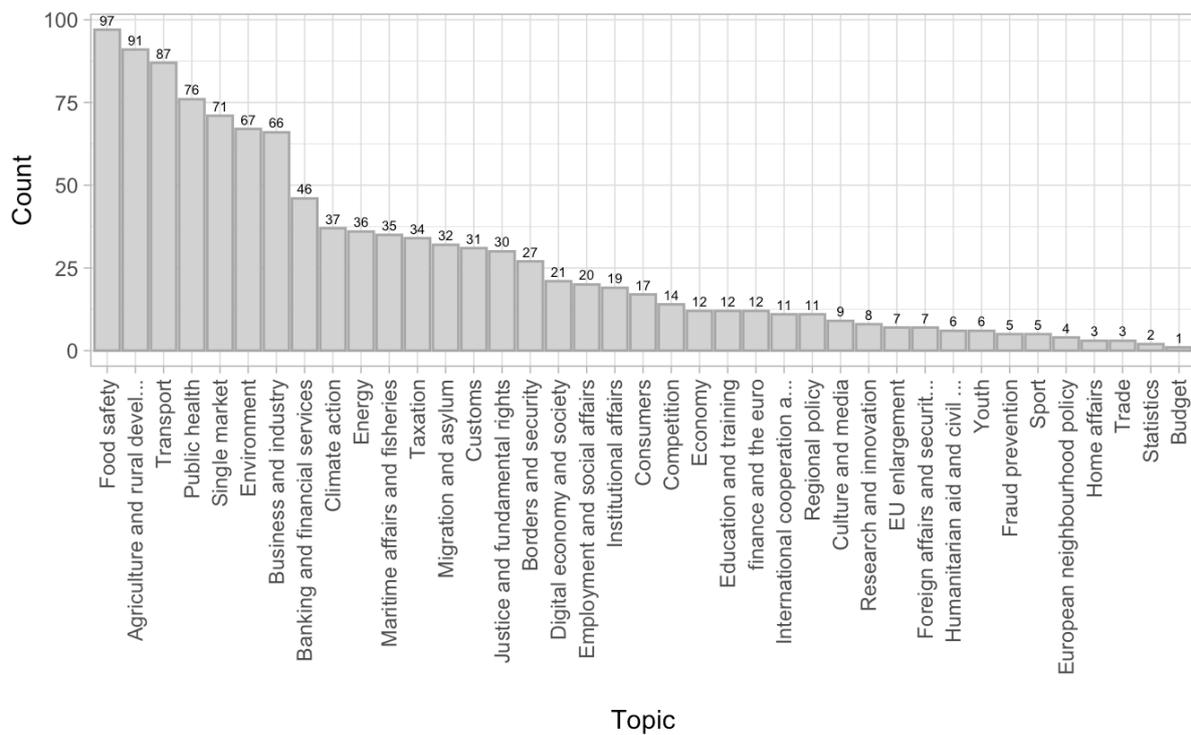
Figure A.2: Most Common Consultation Topics
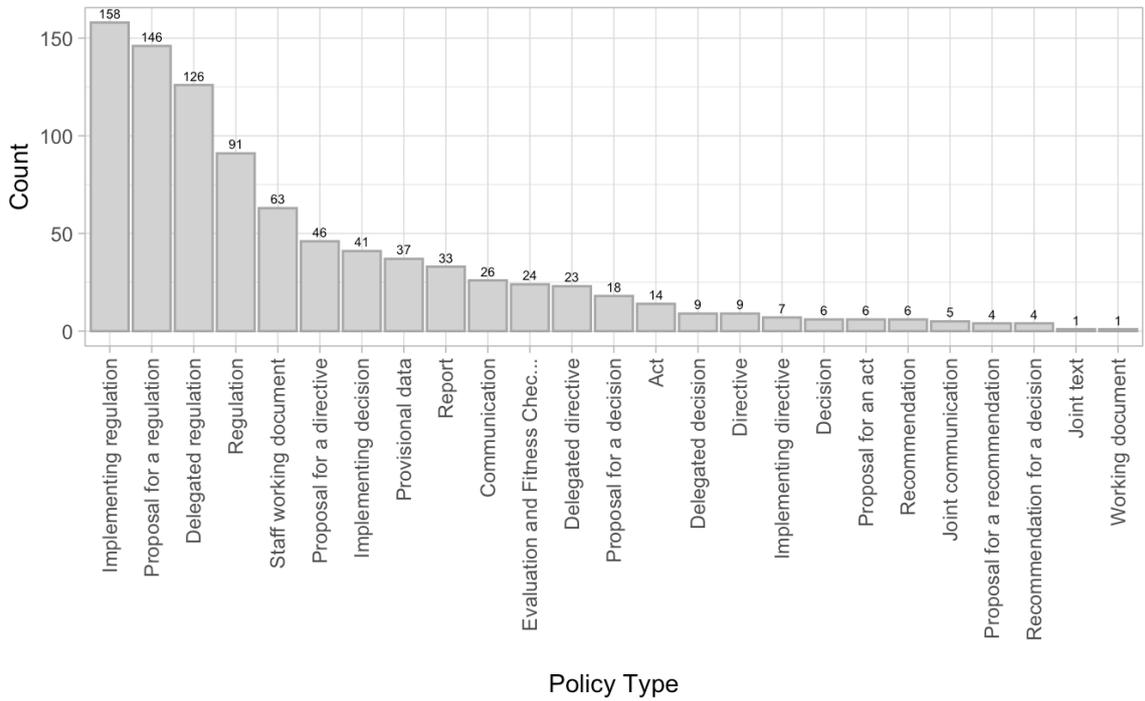
Figure A.3: Most Frequent Policy Types

2 Descriptive Statistics – Stakeholders

Nº Responses with Transparency Register = 5152;
Nº Responses without Transparency Register: 4218

The actor type is derived from the responses on the **Have Your Say** Webpage.
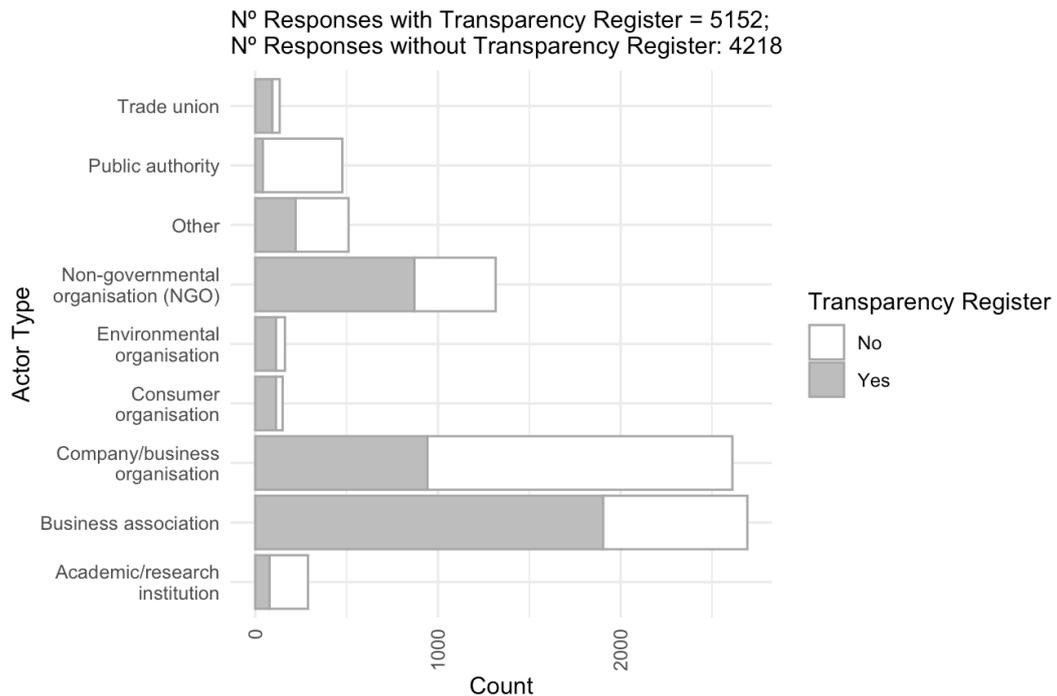
Figure A.4: Transparency Register Coverage by Actor Type
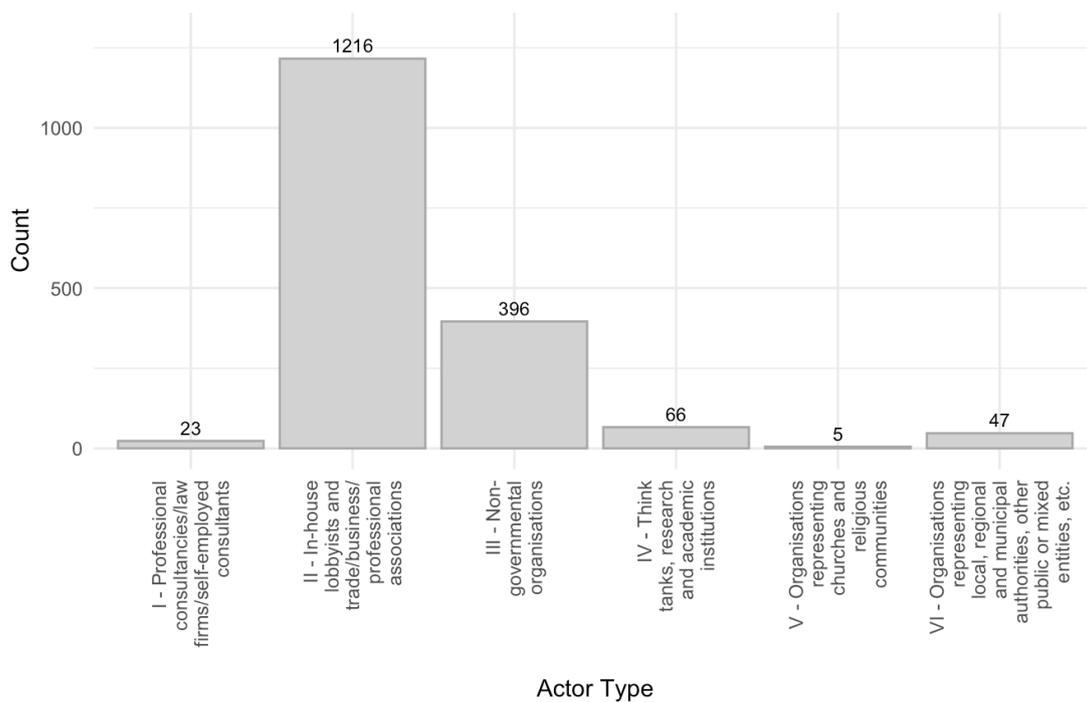
Figure A.5: Actor Types in the Dataset
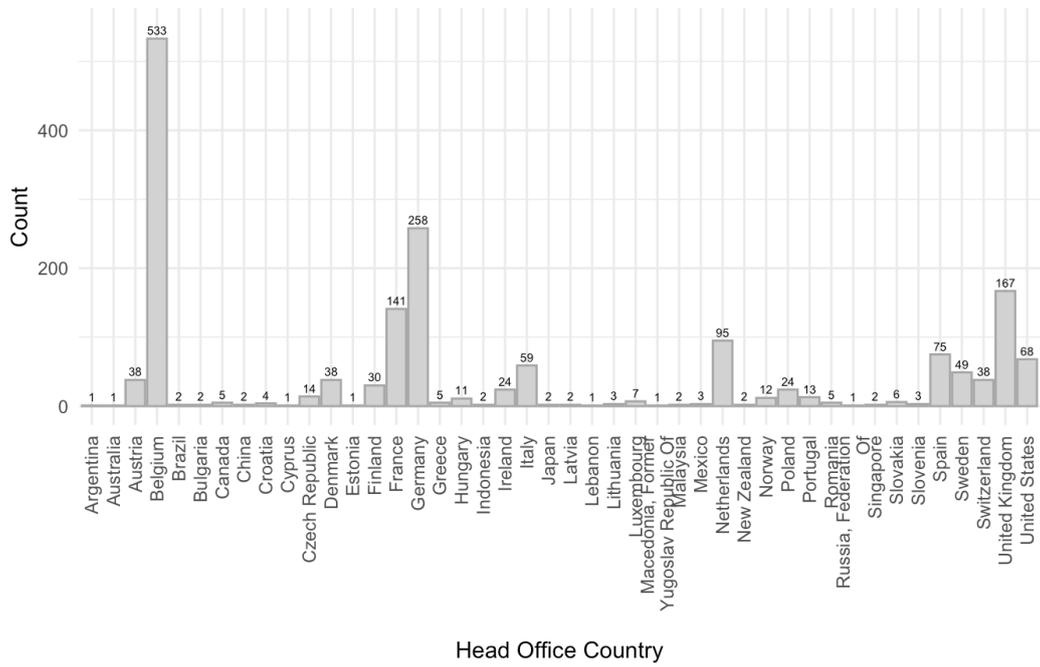
Figure A.6: Country of Origin of Respondents in the Dataset

# 3 Operationalization and Validation of the Method

This appendix illustrates the different steps taken to operationalize the text reuse method described in the main body of the article to detect instances of informational coordination among groups.

Smith-Waterman algorithm

In order to identify instances of text reuse, we run the Smith-Waterman algorithm (Smith and Waterman 1981) using the 'textreuse' package in R (Mullen 2016).

When comparing two sequences of text, the algorithm:

- adds a 'match' score for exactly matching consecutive words,
- subtracts a penalty score when the sequences of text include words that do not match,
- subtracts a penalty score when the sequences of text include placeholder words in matching a shorter sequence to a longer sequence.

These scores are used to identify the substring within the two documents with the highest alignment. In our paper we have followed the approach by Burgess et al. (2016) and we have set these parameters to a match score of 3, mismatch score of -2, and gap score of -3.

In order to identify instances of text re-use in different and non-contiguous parts of the same documents, we divide each document into separate sentences and compare the content of all pairwise combinations of sentences from the responses submitted by different groups to the same policy consultation.

Running a text reuse algorithm on all the pairwise combinations of sentences from the responses to the same consultation would be computationally demanding, a common problem in studies employing text re-use approaches to large corpuses of text. To reduce the computational cost of doing so, we use a 'Minhashing' technique to reduce the number of potential matches by identifying a subset of pair of sentences in the corpus most likely to contain text reuse (for an application of a similar approach see Wilkerson, Smith, and Stramp 2015). This involves converting text strings to shortened numeric references (known as hashes) - a technique first developed to detect duplicated documents on the web - and subsequently filtered out those pairs of sentences below a minimum level of similarity.

Removing False Positives

Next, the data needs to be cleaned to remove false positives, that is, examples of text re-use identified by the algorithm that do not reflect instances of coordination among the authors of the two documents.

First, an important potential source of false positives is represented by cases when different documents cite the same policy document upon which they are commenting or a related one. For instance, in the example in the text both letters cite verbatim a passage from an EU regulatory authority (EIOPA). Moreover, it is common to responses to the same consultation paper to repeat the question that is being answered by the group in their response. In order to filter instances where the patterns of text reuse identified result from citations of the same publicly-available policy documents rather than coordination across the two signatories, we re-run the algorithm between the sentences identified as potential text-reuse and all the policy documents associated with the consultation (e.g. actual policy, consultation documents, background documents). We exclude all the sentences that shared at least 10 consecutive words with the policy documents groups are commenting on (or an equivalent score generated by the Smith-Waterman algorithm).

Second, another important potential source of false positives is represented by the circumstances where different groups may have used the same language independently on any process of direct or indirect coordination, such as in the case of common expressions or the name of an organisation. In order to do this, we identify a minimum number of text reuse that below which overlap between the two text could not be regarded as intentional or meaningful. Table **A.1** provides an illustration of the type of text identified by the algorithm and the scores associated with different amount of text reuse. This reveals how the likelihood that a match is a common expression that could be used independently of a process of intentional coordination is inversely related to the length of the text reuse found within each sentence.

Table **A.1**: Illustration of the Results of the Smith-Waterman Algorithm. '####' indicate a placeholder or missing word in another sentence that was matched to this one

Smith-Waterman Score	Text
30	edcs are a threat to our society s current future public health and prosperity
32	the alternative ###### proposal to harmonise ######### the different general dust limits currently applied in europe
34	an increase in alcohol consumption and #### a substantial increase in deaths from alcohol attributable diseases
36	lead based batteries can withstand internal temperatures which cover all realistic scenarios
38	et al cranberry juice for the prevention of pediatric ######### urinary tract infection ########## a randomized controlled trial
40	comparing eu and us drag testing it ############ appears that a maximum tolerance value of 5 can be technically justified
42	we also have a number of suggestions on how to improve the text further

Table A.1: Illustration of the Results of the Smith-Waterman Algorithm. '####' indicate a placeholder or missing word in another sentence that was matched to this one

Smith-Waterman Score	Text
44	having said this it needs to be ensured that the reviewed annex reflects the ############## progress and lessons ######### learnt ######## as well as the ####### characteristics ###### of ## the sector in the best possible way
46	made the alternative proposal to harmonize ######### the different general dust limits which are currently applied in europe
48	for higher temperatures as mentioned above for lead based batteries additional cooling systems would be required
50	the competent health authorities should participate in the decision process ## when the public has ##### to be informed on exceedances and ### ########## remedial actions
52	europe should take a leading role in regulating edcs ## this will stimulate innovation so that ### industries can ## ### ####### ####### develop and use better and safer alternatives
54	the overall data figure4 suggest that the profile and the amount of the metabolites is not too different
56	annex ii title te ######### since not only measurements but also calculations change the title of annex ii for electric determination methods are allowed for efficiency class determination as 21 stated in article 5 paragraph 4 the title has #### to be
58	however we are concerned that the search for clarity and simplicity risks ######## lowering the overall level of ambition of the directive
60	moreover the study acknowledges that the method 1 is associated to a potential overestimation of waste classification p 13 results
70	lead batteries for vehicle applications in europe are used in an almost closed loop so that today more than 80 of ### lead used ### in the current battery production originates from
70	welcomes the commission s proposal on the 3rd rde #### ######## ######### package as #### ## #### ###### ############ ## a timely step in the right direction to tackle particulate emissions from gasoline direct injection cars and strengthen the rde test procedure
80	corporate action standards have been followed except with ### respect ### to deadlines art 9 and the use of a divergent definition last participant date instead of guaranteed participation date the ## remaining #### significant ##### ## # divergence from the market standards for general meetings
90	the unrivalled cold cranking properties of lead based batteries and the robust re chargeability are key reasons that make this battery type essential and currently irreplaceable for mass market vehicles
100	note that most of our ### comments refer to the interservice consultation drafts and could need to be amended in the coming weeks depending on the importance of changes in the new versions submitted to wto

Second, the likelihood that the presence of text reuse across two documents is indicative of a process of coordination rather than coincidental is higher when this is found across multiple sentences in each document rather than in a single sentence. While it is possible that that two groups may coincidentally use the same sentence, the likelihood this happening multiple times throughout without any coordination is limited.

We consider different combinations of these two filters. Table A.2 summarizes the impact of alternative thresholds in limiting the number of ties captured by the algorithm, as well as the pairs of documents.

**Table A.2**: Descriptive Statistics of Text Reuse

	I	II	III	IV
Threshold	No Filter	Low Filter	Medium Filter	High Filter
Minimum Number of Sentences	0	2	5	10
Minimum Number of Reused Words in a Sentence	5	10	15	20
Number of Pair of Sentences with Text Reuse	693238	680926	593617	575238
Number of Ties	37566	26202	14174	11325
Number of Documents	4573	3749	2150	1604

Figure A.7 illustrates the same results as a social network, whereby two interest groups are connected if they sign two distinct letters in response to the same initiative that share an amount of text-reuse above the different thresholds described above. To facilitate the visual interpretation of these results, we only visualize those groups that are connected via text reuse, while excluding those groups for whom no ties were detected.

No Filter
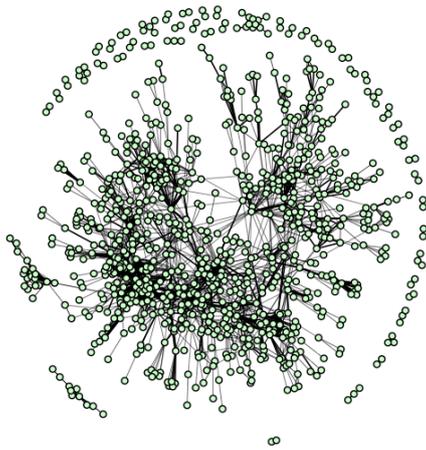
Low Filter
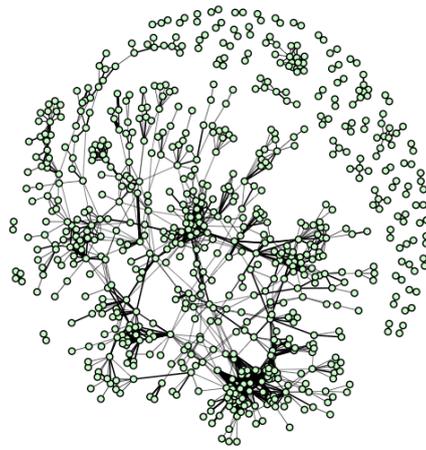
Medium Filter
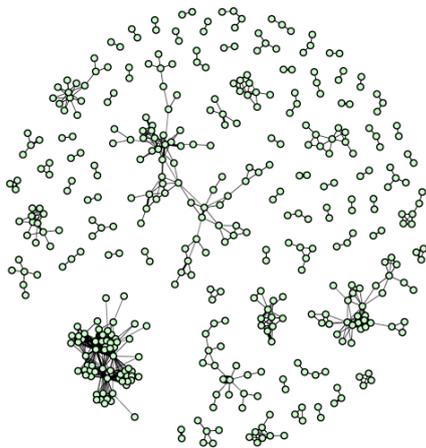
High Filter
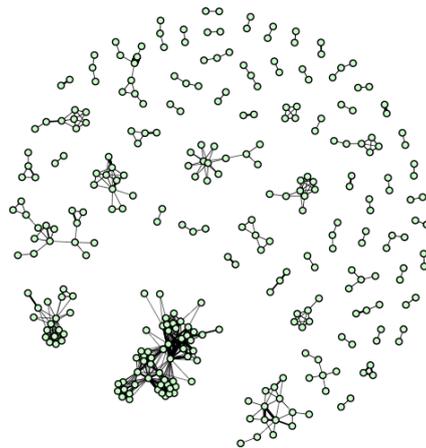
Figure A.7: Text-reuse network. The thickness of the edge connecting two nodes indicates the number of distinct policies where text reuse among those groups occurred

As the minimum threshold of text reuse required for two documents to be regarded as tied increases, the network of documents and groups considered become sparser. While a more stringent filter decreases the likelihood that the analysis will include instances of text re-use that are accidental, it is not possible to exclude the possibility that also legitimate instances of coordination will be excluded because the sentences reused are too short (false negatives). Given the novelty of the approach to study coordination, we decide to take a conservative approach and chose a relatively high threshold for text

reuse to be regarded as evidence of coordination. More specifically, we consider coordination to have occurred among two groups only if their respective submissions share at least 15 consecutive words or an equivalent score (45) when including words that do not match or placeholder words, in at least 10 distinct sentences.

Validation

In order to evaluate more systematically the extent to which our approach could identify instances of text reuse that cannot be attributed to informational coordination across different groups, we run the same algorithm among responses to different consultations whereby coordination should not be expected. Instances of text reuse among responses to different consultations are extremely rare (0.03% of the number of instances of text reuse identified among the responses to the same consultation). A closer look reveals how these few exceptions are the product of the same group submitting some of the same language in response to different policy consultations.

4 Summary Statistics for Variables in Regression Models

Table A.3: Summary Statistics for Variables in Regression Models

Variable	Obs.	Mean	Std. Dev.	Min	Max
Cosigning	5011	0.249	0.4320	0.000	1.000
Text Reuse (high filter)	5011	0.084	0.2770	0.000	1.000
Text Reuse (medium filter)	5011	0.119	0.3240	0.000	1.000
Business Dummy	4950	0.725	0.4460	0.000	1.000
Brussels Office	4950	0.261	0.4390	0.000	1.000
Associational Ties (log)	5011	1.011	0.9980	0.000	3.859
Staff FTE (log)	4950	1.423	0.8160	0.223	3.951
Late Stage Policy	4804	0.682	0.4656	0.000	1.000
Salience (log)	4998	3.411	1.2250	0.693	6.322

Table A.4: Pairwise Correlation Table

Variables		1.	2.	3.	4.	5.	6.	7.	8.
1.	Cosigning	1							
2.	TextReuse (High Filter)	-0.175	1						
3.	TextReuse (Medium Filter)	-0.212	0.823	1					
4.	Business	0.06	-0.046	-0.019	1				
5.	Brussels Office	-0.05	-0.017	-0.016	0.125	1			
6.	Associational Ties	0.09	0.033	0.045	0.013	-0.011	1		
7.	Staff Capacity	0.028	0.031	0.034	-0.109	0.21	0.446	1	
8.	Late Stage Policy	0.177	0.043	0.042	0.079	-0.03	-0.028	-0.036	1
9.	Salience	-0.148	0.112	0.137	0.077	-0.017	-0.042	-0.073	0.022	1

Table A.5 below represents each of the models we report on in the paper, in logit coefficient form, with standard errors in parentheses. Models 1 and 2 are logistic regression models with standard errors clustered at the consultation initiative level. Models 3-5 are multilevel logistic regression models with mixed effects, with the consultation initiative as a distinct level.

Table A.5: Regression Table

Variable	(1) Cosigning	(2) Text Reuse (High Filter)	(3) Cosigning	(4) Text Reuse (High Filter)	(5) Text Reuse (Medium Filter)
Business	0.357 /n(0.324)	-0.563* (0.289)	0.0953 (0.274)	-0.523*** (0.199)	-0.27 (0.186)
Brussels Office	-0.302** (0.141)	-0.101 (0.175)	-0.660** (0.28)	-0.213 (0.174)	-0.213 (0.18)
Associational Ties (log)	0.194 (0.12)	0.125 (0.0841)	-0.0241 (0.141)	0.212** (0.0836)	0.265*** (0.0732)
Staff Capacity (log)	0.0104 (0.112)	0.0681 (0.107)	0.376** (0.173)	0.048 (0.126)	0.0387 (0.134)
Late Stage Policy	1.018** (0.412)	0.367 (0.272)	1.080*** (0.192)	0.167 (0.145)	0.211 (0.132)
Salience (log)	-0.326 (0.288)	0.339*** (0.0892)	-0.455*** (0.0953)	0.974*** (0.0765)	0.942*** (0.0765)
Constant	-1.141 (1.12)	-3.674*** (0.472)	-5.181*** (0.561)	-6.775*** (0.47)	-6.326*** (0.435)
Constant for Initiative Level			1.604*** (0.108)	0.626*** (0.0739)	0.619*** (0.06)
Pseudo R2	0.066	0.034			
Observations	4747	4747	4747	4747	4747

Note:
Notes: Standard errors in parentheses * p < 0.10, ** p < 0.05, *** p < 0.01. For models 3,4 and 5, 100 bootstrap replications were run, based on 1,670 clusters in policy initiative identifier.

One potential concern for our findings is that our results for text reuse - and therefore in the comparison we make between co-signing and text reuse - are sensitive to the threshold we use for what constitutes a text reuse tie. To consider this we re-ran the multilevel logit models reported in the paper with a lower text reuse threshold, corresponding to the 'Medium Filter' in Table A.5.

Figure A.8 compares the odds ratio estimates for both the model reported in the main paper and the alternative dependent variable. The only difference of note is the business actor variable, which is directionally similar but is not statistically significant with the Medium Filter threshold for text reuse.
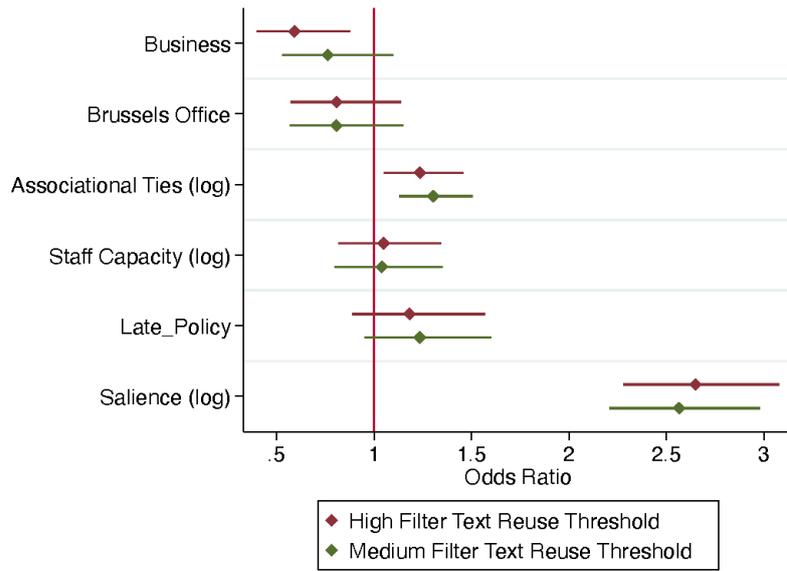
Figure A8: Multilevel Logit Odds Ratio, Comparing Two Different Text Reuse DVs