

RESEARCH NOTE

What makes policy complex?*

Roman Senninger 

Department of Political Science, Aarhus University, Aarhus, Denmark

Corresponding author. Email: rsenninger@ps.au.dk

(Received 7 February 2022; revised 13 December 2022; accepted 19 April 2023; first published online 4 July 2023)

Abstract

Previous research finds that policy complexity affects important political processes including legislative delegation and policy diffusion. However, policy complexity is not directly observable and the search for a reasonable proxy constitutes a major challenge for scholars. This research note presents a concise and measurable definition of complex policy based on two aspects: a policy's textual sophistication and its ties to other rules and regulations. Using crowdsourcing and a pairwise comparison framework it is shown that the proposed defining features are crucial for humans' understanding of policy text. The proposed definition is then operationalized using a large corpus of European Union rules and is shown to outperform alternative operationalizations of policy complexity in predicting the level of legislative delegation.

Keywords: Public policy; Policy complexity; Legislative delegation; Text and content analysis; Pairwise comparison

1. Introduction

Policy complexity affects many important political processes including delegation and policy diffusion (Kiewiet and McCubbins, 1991; Epstein and O'Halloran, 1999; Braun and Gilardi, 2009; Makse and Volden, 2011). However, policy complexity is not directly observable and is therefore difficult to capture. Researchers rely on different measures, including readability scores or the number of articles, but most prior research only looks at a single aspect of policy complexity at a time. Below I argue and empirically validate that a policy's complexity is best defined by two aspects: its textual sophistication and the number of ties to other policies. The findings are important because they emphasize the role of both internal (i.e., textual) and external (i.e., relational) characteristics in explaining what makes policy difficult to understand. More broadly, they have implications for our understanding of the causes and consequences of complexity in policy-making.

2. Defining complex policies

While some scholars argue that policy complexity depends on the length and detail of a policy (Ehrlich, 2011; Hurka and Haag, 2020), others say that policy complexity is the result of the increasing number of relations between policies (Krehbiel, 1991; Adam *et al.*, 2019). The former count the number of articles and words or rely on readability indexes such as the widely used Flesch Reading Ease formula. The latter investigates the complexity of policies through document network analysis (Katz and others, 2020). Crucially, most prior research focuses on a single aspect of policy complexity at a time.¹

*Earlier versions of this paper have been presented at EPSA 2019, APSA 2019, and Harvard University.

¹It is important to note that this research note is concerned with the definition, operationalization, and measurement of what makes individual policy (i.e., laws, rules, and regulations) complex. The approach is more fine-grained than previous

My definition integrates two approaches. I define complex policies as those that have a high level of textual sophistication and a large number of ties to other policies. For individuals who want or need to engage with a policy there is no way around reading the text of the policy. Reaching a good understanding of the policy can be straightforward if it is written in an accessible manner. By contrast, a policy with a high level of textual sophistication makes it harder for the reader to understand. There exist many potential sources of textual sophistication. The characteristics that contribute to making a text more complex are text length, the use of longer words, the use of uncommon words, and the use of more complex syntactic and grammatical structure (Benoit *et al.*, 2019). Findings from different fields including medicine, communication science, and political science provide evidence that textual sophistication matters a great deal for humans' understanding of text (Leroy *et al.*, 2010; Bischof and Senninger, 2018; Tolochko *et al.*, 2019; Bischof and Senninger, 2022).

The second defining feature of complex policies goes beyond a policy's own characteristics and considers its wider context. This means potential cross references to other laws, rules, and regulations. The reason why ties to other policies are considered to be important is that they provide information that is relevant to fully understanding a policy. The decision to draft a new policy is very often motivated by insufficiencies of already existing policies. However, instead of withdrawing insufficient policies and replacing them with new and better policies, a process of policy layering or policy accumulation is increasingly common (Adam *et al.*, 2019). Whenever two policies build upon each other or regulate very similar policy domains, it is likely that the newer policy makes reference to the older policy to describe the relation between the two (Krehbiel, 1991). A large number of references to other policies can make it more difficult to reach a full understanding of a policy because the consideration of additional related policies is required.

3. Empirical roadmap

The central proposal of this research note is that both textual sophistication and ties with other policies should be used to capture policy complexity. In the following, I first validate the proposed defining features of complex policy by showing that they are crucial for humans' understanding of policy. This section builds on and expands the workflow presented by Benoit *et al.* (2019). Thereafter, the proposed definition is operationalized using a large corpus of policies, and it is shown to outperform alternative operationalizations in predicting a theoretically relevant outcome, namely the level of legislative delegation. Both empirical exercises are conducted in the context of the European Union because it constitutes a large and important jurisdiction that has law-making powers in a broad range of policy areas.² The flowcharts in Figure 1 provide information about the individual steps of the two empirical tests.

4. Validation of proposed defining features

Step 1: First, human judgments of the relative complexity of policy texts using crowdsourcing were collected. The approach involves non-experts who were asked to complete micro-tasks and works particularly well for identifying (latent) document characteristics (Carlson and Montgomery, 2017). The data consist of comparable short passages of text taken from recitals of European Union rules (Thomson *et al.*, 2012). Recitals are listed before the articles of a policy

work that lumps laws together in policy issue areas suggesting that some areas are more complex than others (Epstein and O'Halloran, 1999; Nicholson-Crotty, 2009).

²Section A in the Supplementary information reviews current practices as regards the definition and operationalization of policy complexity in the European Union. Section B provides an introduction to delegation in the context of the European Union.

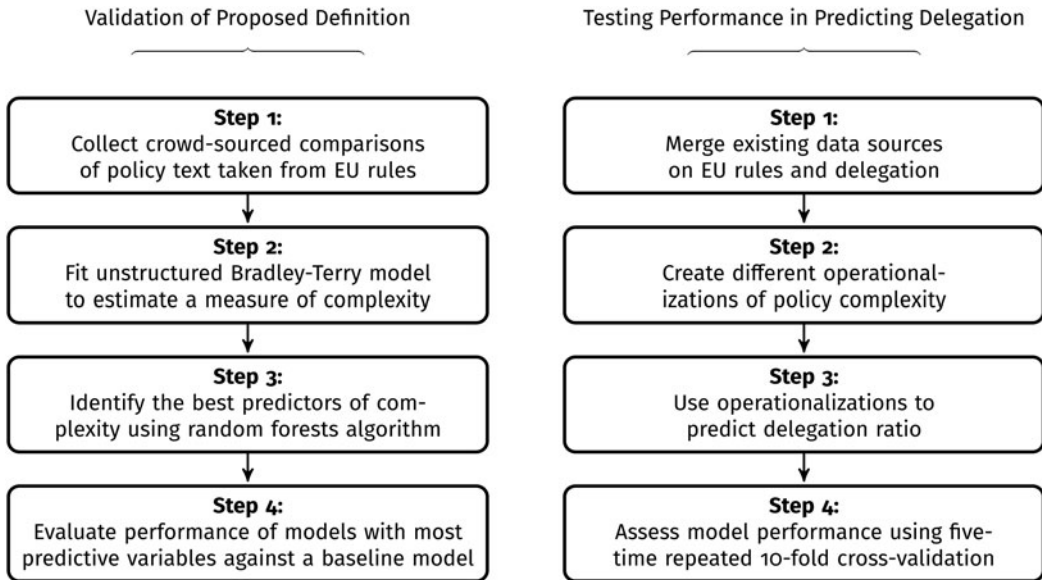


Figure 1. Individual steps of the empirical tests.

act and state the reasons for the provisions, principles, and assumptions on which the act is based.³ From this text corpus, text snippets of varying length were randomly drawn. Following a stratified sampling method, the snippets drawn for comparison were constrained to groups of the same number of sentences and a similar number of characters to avoid comparisons in which coders simply select the one noticeably shorter than the other.⁴

Participants were recruited using the crowdsourcing platform Prolific. The sample is representative of the population of the UK using proportional cross-stratification on sex, age, and ethnicity. In total, 597 individuals participated in the task.⁵ Upon accepting the task, participants were shown a description of the task and a number of examples (see Figure SI 2 and the upper panel in Figure SI 3). Each respondent was asked to compare 15 randomly assigned pairs (two pairs for the purpose of attention checking). To screen respondents' attention, instructive manipulation checks were used (see lower panel in Figure SI 3). For the main analysis, I exclude respondents who failed to pass the attention checks leading to 536 participants and a total of 6962 comparisons. The average number of judgments per snippet is 5.1.

Step 2: The second step is to estimate the underlying complexity using the model for pairwise comparisons developed by Bradley and Terry (1952). The Bradley–Terry model assumes that the odds that *snippet_i* beats *snippet_j* are α_i/α_j , where α_i and α_j are parameters representing the “easiness” of snippets, as respondents were asked which text snippet was easier to understand. The model can be expressed in logit form: $\text{logit} [\text{Pr}(\text{easier than } j)] = \lambda_i - \lambda_j$, where $\lambda_i = \log \alpha_i$ for all i . Fitting the equation to the pairwise data results in estimates of λ_i for each text snippet, representing an unconditional estimate of that text's relative easiness.⁶

³For more information about recitals consult Section A in the Supplementary information. The rationale for using the text from recitals in the validation exercise is described in Section G in the Supplementary information.

⁴See Table SI 1 in the Supplementary information. Text snippets which were outside the 0–121 range of the Flesch Reading Ease measure were dropped, as a simple way to remove very unusual texts. This preprocessing resulted in 1340 snippets which were linked together so that within each group every snippet meets one snippet in a comparison that meets another snippet.

⁵Participants were paid £7.17 per hour, which lies above the minimum reward per hour (£5.5) suggested by Prolific.

⁶In some cases a snippet never wins or never loses a competition raising the issue of complete separation. To address this, the bias-reduction technique embedded in the Bradley-Terry2 R package (Turner and Firth, 2012) was applied.

Step 3: The next step is to select potential predictors of this outcome, considering a model of the form: $\lambda_i = \sum_{i=1}^p \beta_i x_i + U_i$, in which easiness of each snippet i is related to explanatory variables x_1, \dots, x_p through a linear predictor with coefficients β_1, \dots, β_p . U_i represents independent errors (Turner and Firth, 2012). The estimated coefficients $\hat{\beta}$ indicate the marginal effect of each covariate on the perceived relative easiness of the text snippets. To represent textual sophistication, the absolute number of words and characters are considered. In addition, several variables that are part of the best model to explain textual complexity as presented in Benoit *et al.* (2019) are used. These are the mean number of characters per word, the mean number of characters per sentence, and the least frequent word's relative frequency based on the Google books data set. Finally, I add the Flesch Reading Ease, a common readability measure, determined by the number of words and the average number of syllables per word. To represent the second defining feature of complex policy, it was manually coded whether a text snippet refers to any existing legal acts or additional documents including treaties, conventions, communications, and resolutions. The variable comes in two versions. The first version is binary coded and indicates whether a text snippet includes a reference to any rules or documents. The second version indicates the number of such references. The appearance and number of abbreviations are also considered. By convention, recitals can start with the word "Whereas." For each text snippet it was recorded whether this is the case or not. All variables are listed in Table 1.⁷

To assess the predictive power of the listed covariates, random forest models with 1000 trees were used. Random forests are chosen because they are parsimonious, general, and less prone to overfitting (Lantz, 2015). They produce estimates of the relative importance of each variable which is useful information for selecting the best predictors of easiness of text snippets. Figure SI 5 ranks the variables' importance according to the value of the increase in mean squared error (MSE) as a result of a variable being permuted. At each node in each tree, three random variables were tried for the regression. This showed that some of the variables used in previous research also matter for predicting the easiness of short passages of text from recitals (especially the mean characters per sentence). Even more important are the absolute number of words. The mean characters per word also matter. These results provide evidence that textual sophistication is important for humans' understanding of policy text. In addition, it shows that ties with other rules and regulations matter as well. The variable representing the number of references to existing rules and documents contributes to the prediction of the outcome fourth most. When permuting the values of the number of references to other documents over the data, the increase in the MSE is 11 percent. For the binary predictor, the increase is 12 percent.

Step 4: Finally, I use the most predictive variables to fit structured models and assess their performance in predicting the pairwise contests. I compare the models against a baseline model that includes the widely used Flesch Reading Ease score as its only covariate (model 1). Model 2 includes the two most predictive variables of textual sophistication. These are the number of words and the mean characters per sentence. Model 3 keeps the two variables to capture textual sophistication but adds the number of ties with other rules and documents. This third model captures both of my defining features of policy complexity and performs best, with the lowest Akaike information criterion (AIC) (9205.2) and the highest proportion of pairwise comparisons correctly predicted (0.772).⁸ For the first model, we see that the AIC is 9613.2, and the augmented proportion of contests in the data correctly predicted is 0.676. Model 2 outperforms the first model with a lower AIC (9374.4) and a higher proportion of pairwise comparisons correctly predicted (0.752) (Table 2).⁹

⁷For descriptive statistics consult Table SI 3. Figure SI 4 shows correlations between variables.

⁸As discussed in Benoit *et al.* (2019), the upper bound of what any model can achieve as regards performance is dependent on agreement between crowd-coders in their ratings of text snippets and the number of contests. The average upper bound of performance for a model is 0.78 (or 78 percent) correctly predicted.

⁹Table SI 4 presents the performance of an alternative approach combining the FRE and references to other regulation which is outperformed by model 3. Hence, the FRE seems to be a rather poor measure of textual sophistication in the given context.

Table 1. Predictors of policy complexity

Description	Covariate name
Textual sophistication	
Number of words	n_token
Mean characters per word	mean Word Chars
Mean characters per sentence	mean Sentence Chars
Google Books baseline usage	google_min
Flesch Reading Ease	Flesch
Ties between policies	
Reference to other regulation	reg_ref
No. of references to other regulation	reg_ref_count
Additional covariates	
Text snippet with abbreviations	abr
No. of abbreviations	abr_count
Text snippet with “Whereas”	whereas

Table 2. Model performance

Model	AIC	Accuracy
Model 1: Flesch Reading Ease	9613.2	0.676
Model 2: Number of words, mean characters per sentence	9374.4	0.752
Model 3: Number of words, mean characters per sentence, no. of references to other regulation	9205.2	0.772

It is important to note that both defining features matter for our understanding of policy text, as a model that only includes textual sophistication is clearly outperformed by a model that features both textual sophistication and ties to other rules. To demonstrate the face validity of the results, text boxes in Section E in the Supplementary information present text snippets used in the pairwise comparisons which the best performing model identified as having a very low, an average, and a very high level of complexity, respectively.¹⁰

5. Testing performance in predicting delegation

In the following, I test how different operationalizations of policy complexity perform in predicting the level of legislative delegation. *Step 1:* For this purpose, several data sources were merged. For the level of legislative delegation, data come from Anastasopoulos and Bertelli (2020), who use machine learning techniques to measure the amount of delegation to the European Commission and member states’ national administrations in directives and regulations. The predicted values for each provision, effectively articles and sub-articles, are aggregated so that the dependent variable gives the delegation ratio (Δ_i) for each law i . The delegation ratio (Δ_i) represents the number of provisions delegating authority D_i divided by the total number of provisions in the law P_i ($\Delta_i = D_i/P_i$).¹¹ *Step 2:* The data also include the raw text of the provisions of the law. This allows me to operationalize the textual sophistication in a similar manner to that used in the pairwise comparison described above. More specifically, the number of words and the mean number of characters per sentence for each law i were

¹⁰After the pairwise comparisons, respondents were asked to give a description of the features that make a text snippet easier/more difficult to understand. Their answers support the conclusions drawn above and are summarized in Section H in the Supplementary information.

¹¹Delegation comes in two versions, namely delegation to the European Commission and delegation to member states’ national administrations. Tables SI 5–7 present results for delegation to the two agents separately.

Table 3. Model performance - Predicting Legislative Delegation

	RMSE	MAE	R^2	N
Main results: Delegation ratio				
Model 1: Flesch Reading Ease	0.167	0.125	0.070	13366
Model 2: Number of words, mean characters per sentence, and no. of references to other regulation	0.161	0.116	0.152	13366
Additional results: Delegation ratio				
Model 1: Recitals	0.169	0.142	0.103	317
Model 2: Number of words, mean characters per sentence, and no. of references to other regulation	0.163	0.139	0.164	317
Additional results: Delegation ratio				
Model 1: Recitals	0.272	0.198	0.074	302
Model 2: Number of words, mean characters per sentence, and no. of references to other regulation	0.258	0.189	0.176	302

Note: Range of delegation ratio is 0–1.

RMSE, root mean squared error; MAE, mean absolute error.

estimated. Moreover, the ties to existing legislation, treaty articles, and court judgments for each piece of legislation were measured. The number of ties was not directly extracted from the raw text of the provisions but taken from a recently introduced database tracking connections between European Union laws (Fjelstul, 2019).

Merging these data sources provides me with the delegation ratio, the mean number of characters per sentence, the number of words, and the number of ties to other policies for more than 13,000 pieces of legislation enacted by the two co-legislators, the Council of the European Union and the European Parliament, between 1958 and 2015. To compare the performance results against a baseline model, I operationalize complexity using the average Flesch Reading Ease of a law's provisions. In addition, I compare my suggested definition of policy complexity to an operationalization that uses the number of recitals to measure policy complexity. For this purpose, subsets of the data described above were merged with data from two studies including information about the number of recitals (Steunenberg and Rhinard, 2010; Migliorati, 2020). *Step 3:* The models consist of the delegation ratio as the response variable and an operationalization of policy complexity as the predictor variable. The main goal is to get optimal predictions based on a linear combination of the described variables (Cranmer and Desmarais, 2017). *Step 4:* To assess model performance, five-time repeated tenfold cross-validation was applied.

The final model error is the mean error from the various iterations. Table 3 shows the performance results. The proposed operationalization of policy complexity using the number of words, the mean number of characters per sentence, and the number of ties shows a smaller root mean squared error and mean absolute error. The differences are consistent but not very large. In addition, the R^2 , which tells us the proportion of the variance in the response variable that can be explained by the predictor variable(s) in model 2 is clearly larger. Additional results show that my proposed definition and operationalization of policy complexity outperforms an alternative operationalization that is often used in the context of the European Union, namely the number of recitals. All model comparisons show that models with the proposed operationalization have a lower RMSE and higher R^2 which tells us that they are able to fit the data better than the alternative operationalization.

6. Discussion and conclusion

This research note brings forth important implications for scholars interested in the causes and consequences of complexity in public policy. It presents a definition and operationalization of policy complexity that is validated at the individual level and at the same time turns out to be a

relevant predictor of legislative delegation. As such, the approach performs well in a theoretically meaningful test. This stands in stark contrast to existing measures, including readability scores and the number of articles and recitals, that are based on strong implicit assumptions. As a result, future studies are well advised to incorporate operationalizations of textual sophistication and ties between policies to make sure that their measure captures features that actually impact on humans' understanding of policy text. When the proposed operationalization is used to explain a phenomenon like delegation, researchers should ensure that policy complexity and delegation measures are separated, as delegation is sometimes measured by the length of a bill. The proposed definition and operationalization focus upon features that are generally applicable. They do not assume that policies are difficult to understand simply because they belong to a specific policy context but rather focus on general features. This means that the approach can be applied to different topics such as policy diffusion and in different contexts including individual countries within the European Union but also outside of the European context, such as in the USA, and even in sub-national politics.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/psrm.2023.23>. To obtain replication material for this article, <https://doi.org/10.7910/DVN/IPW0M9>

Acknowledgments. I want to thank Jason Anastasopoulos, Kristina Bakkær Simonsen, Jens Blom-Hansen, Steffen Hurka, Heike Klüver, Stefan Müller, Fritz Sager, Bruno Castanho Silva, and Christian Rauh. I acknowledge funding from the Department of Political Science at Aarhus University.

References

- Adam C, Hurka S, Knill C and Steinebach Y (2019) *Policy Accumulation and the Democratic Responsiveness Trap*. Cambridge: Cambridge University Press.
- Anastasopoulos LJ and Bertelli AM (2020) Understanding delegation through machine learning: a method and application to the European Union. *American Political Science Review* **114**, 291–301.
- Benoit K, Munger K and Spirling A (2019) Measuring and explaining political sophistication through textual complexity. *American Journal of Political Science* **63**, 491–508.
- Bischof D and Senninger R (2018) Simple politics for the people? Complexity in campaign messages and political knowledge. *European Journal of Political Research* **57**, 473–495.
- Bischof D and Senninger R (2022) Simplifying politics: The impact of linguistic style on citizens' political knowledge and beliefs about politicians. doi:10.31219/osf.io/cgz4k.
- Bradley RA and Terry ME (1952) Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* **39**, 324–345.
- Braun D and Gilardi F (2009) *Delegation in Contemporary Democracies*. Milton Park: Routledge.
- Carlson D and Montgomery JM (2017) A pairwise comparison framework for fast, flexible, and reliable human coding of political texts. *American Political Science Review* **111**, 835–843.
- Cranmer SJ and Desmarais BA (2017) What can we learn from predictive modeling?. *Political Analysis* **25**, 145–166.
- Ehrlich SD (2011) *Access Points*. Oxford: Oxford University Press.
- Epstein D and O'Halloran S (1999) *Delegating Powers*. Cambridge: Cambridge University Press.
- Fjelstul JC (2019) The evolution of European Union law: a new data set on the *acquis communautaire*. *European Union Politics* **20**, 670–691.
- Hurka S and Haag M (2020) Policy complexity and legislative duration in the European Union. *European Union Politics* **21**, 87–108.
- Katz DM, Coupette C, Beckedorf J and Hartung D (2020) Complex societies and the growth of the law. *Scientific Reports* **10**, 18737–18737.
- Kiewiet RD and McCubbins MD (1991) *The Logic of Delegation*. Chicago: Chicago University Press.
- Krehbiel K (1991) *Information and Legislative Organization*. Ann Arbor: University of Michigan Press.
- Lantz B (2015) *Machine Learning with R*, 2nd Edn. Birmingham: Packt Publishing.
- Leroy G, Helmreich S and Cowie JR (2010) The influence of text characteristics on perceived and actual difficulty of health information. *International Journal of Medical Informatics* **79**, 438–449.
- Make T and Volden C (2011) The role of policy attributes in the diffusion of innovations. *The Journal of Politics* **73**, 108–124.
- Migliorati M (2020) Where does implementation lie? Assessing the determinants of delegation and discretion in post-Maastricht European Union. *Journal of Public Policy* **41**, 1–22.

- Nicholson-Crotty S** (2009) The politics of diffusion: public policy in the American states. *The Journal of Politics* **71**, 192–205.
- Steunenberg B and Rhinard M** (2010) The transposition of European Law in EU member states: between process and politics. *European Political Science Review* **2**, 495–520.
- Thomson R, Arregui J, Leuffen D, Costello R, Cross J, Hertz R and Jensen T** (2012) A new dataset on decision-making in the European Union before and after the 2004 and 2007 enlargements (DEUII). *Journal of European Public Policy* **19**, 604–622.
- Tolochko P, Song H and Boomgaarden H** (2019) That looks hard!: effects of objective and perceived textual complexity on factual and structural political knowledge. *Political Communication* **36**, 609–628.
- Turner H and Firth D** (2012) Bradley-Terry models in R: the Bradley Terry 2 package. *Journal of Statistical Software* **48**, 1–21.