

ARTICLE

From One to Many: Identifying Issues in CJEU Jurisprudence

Philipp Schroeder^{1,*} and Johan Lindholm²

¹Department of Political Science, Ludwig-Maximilians-University Munich, Munich, Bavaria, Germany and

²Department of Law, Umeå University, Umeå, Vasterbotten, Sweden

*Corresponding author. Email: P.Schroeder@gsi.uni-muenchen.de

(Received 04 February 2021; Revised 28 June 2021; Accepted 24 September 2021)

Abstract

Research of judges and courts traditionally centers on judgments, treating each judgment as a unit of observation. However, judgments often address multiple distinct and more or less unrelated issues. Studying judicial behavior on a judgment level therefore loses potentially important details and risks drawing false conclusions from the data. We present a method to assist researchers with splitting judgments by issues using a supervised machine learning classifier. Applying our approach to splitting judgments by the Court of Justice of the European Union into issues, we show that this approach is practically feasible and provides benefits for text-based analysis of judicial behavior.

Keywords: text analysis; machine learning; preliminary rulings; jurisprudence

Introduction

Scholars interested in explaining judicial behavior often use court judgments as a primary source of information to test their theoretical expectations. In doing so, it may seem natural to approach the judgment as the main or only unit of observation (see for example Vanberg 2005; Owens et al. 2013; Corley and Wedeking 2014). After all, the main output of litigation is frequently a single court-produced document. However, this approach overlooks that judgments are neither disconnected from each other nor internally homogeneous. For one, judgments are interlinked by addressing similar and related questions. This is obvious when a court in one judgment explicitly cites existing case law for a rule and then develops that rule to be applied in subsequent cases. Further, a court settling a dispute must frequently address multiple legal questions within a single judgment, including procedural, constitutional, and substantive questions.

Consider, for example, the judgment of the Court of Justice of the European Union (CJEU) in *Laval*.¹ The case concerned a labor conflict between a Latvian corporation, Laval un Partneri Ltd., who had won a contract to construct school buildings in Sweden and Swedish labor unions who blocked Laval's access to the construction site in order to force them to enter into a collective bargaining agreement (CBA) with terms in line with other Swedish CBAs. Although the unions' actions were supported by Swedish law, Laval challenged their legality on the grounds of EU law.² Laval convinced the Swedish court to request a preliminary reference from the CJEU on two questions regarding the interpretation of EU law. In addition to answering those two questions, the Court also had to decide whether the request for a preliminary reference was admissible. Thus, the CJEU's judgment in *Laval* addresses three legal questions, two substantive and one procedural, that are connected to one dispute but distinct from each other.

These characteristics of judgments have implications for scholars of judicial politics who rely on quantitative text data in their work. As scholars' focus shifted toward the evolution of law at the hand of judges, Clark and Lauderdale (2012, 329) diagnosed that the empirical literature "has struggled to keep pace" with theories of judicial decision making that center on the contents of judge-made law. Working closely with the text of judgments appears to be the most promising avenue to close this gap (Panagis and Sadl 2015), and recent studies illustrate the promise of using computer-assisted text analysis in the field of judicial politics (see Lauderdale and Clark 2014; Aletras et al. 2016; Dyeve 2020; Solan 2017; Vogel et al. 2018; Medvedeva et al. 2020).

In order to make full use of these powerful methods to explain how judge-made law develops, how individual judges' preferences feed into their writings, and how external influences shape judges' answers to the legal questions before them (see Clark and Carrubba 2012; Owens and Wedeking 2011; Corley and Wedeking 2014; Staton and Vanberg 2008), we argue that judgment texts ought to be split into blocks addressing individual, internally coherent *issues*, a concept explained in more detail in Section 2. Viewing judgments as combinations of text blocks that address distinct issues unlocks a unit of observation that captures the aspects of judicial decisions we are often most interested in—the written reasoning of courts on the legal questions they need to address. In this article, we show that splitting judgments into issues is practically feasible and identifies patterns in case law and judicial decision making that studies relying on judgments as units of analysis struggle to uncover.

The article proceeds as follows. In Section 2, we conceptualize what we understand as issues in judgments. Throughout the remainder of the article, we then draw on our experiences of working with judgments of the CJEU to illustrate the implementation and benefits of our issue-splitting approach. In Section 3, we show how machine learning classifiers can mitigate the effort needed to split judgments into issues through manual coding. Empirical illustrations in Section 4 demonstrate how working with issue-split judgments improves our ability to identify the topical content and coherent clusters of judge-made law compared to relying on entire judgments. Finally, in Section 5, we discuss the implications of using issues rather

¹Case C-341/05, Laval un Partneri Ltd v. Svenska Byggnadsarbetareförbundet et al., ECLI:EU:C:2007:809, available at <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:62005CJ0341>.

²More specifically, Laval, who already had a valid Latvian CBA, argued that Swedish law's inequitable treatment of CBAs from other Member States violated the fundamental right of free movement of services.

than entire judgments as units of observation for studies applying standard econometric tools to study judicial behavior. We replicate a study conducted by Larsson et al. (2017) on the CJEU's strategic references to its own case law and compare results from an issue-level to a judgment-level analysis. Section 6 offers concluding remarks.

Splitting the judgment

The cases that judges hear engage with the law on various levels of generalization. From a narrow and result-focused perspective, cases are about *rulings*. A ruling refers to the outcome in a specific case and how the court decided the lawsuit and, more specifically, ruled on the claim(s) brought before it by the parties. For example, one could say that a case is about whether a defendant should be required to pay damages to a plaintiff.

However, most modern studies of judicial behavior center on the questions that the court had to answer in order to decide the case (see Clark and Lauderdale 2010; Lax 2011). These questions are typically divided into two types, *questions of fact* and *questions of law*, that is, questions regarding what the law is on a particular point. All courts answer questions of fact and law and frequently multiple of both types. Questions of law are particularly significant in cases heard by apex courts as their answers serve as models for deciding subsequent cases.

Previous research acknowledges that courts often need to address multiple legal questions to decide a case. For example, the Supreme Court Database addresses judgments both as a whole, split by what they refer to as issues and legal provisions, and split by actions (see Baum 2017; Epstein and Knight 1998; Segal and Spaeth 2002). We draw on the terminology of issues in judgments to characterize blocks of text within judgments that address distinct legal questions.³ We conceptualize an issue as a connecting middle layer between the judgment level and the paragraph level, clustering paragraphs addressing the same legal question.

To illustrate, we return to the example of the CJEU's judgment in *Laval*. Figure 1 provides a simplified illustration of the CJEU's judgment in the case, clustering paragraphs into three blocks of text in the judgment: paragraphs 42 to 50 addressing the admissibility of the preliminary reference, paragraphs 53 to 111 addressing the Swedish court's first substantive question,⁴ and paragraphs 112 to 120 addressing its second substantive question.⁵

The nature of issues considered by different courts is dependent on the institutional and procedural context. In preliminary references answered by the CJEU, issues are closely tied to the questions referred by national courts and capture the CJEU's interpretations of different aspects of EU law. The issues considered by the European Court of Human Rights (ECtHR), for instance, concern questions whether participating States failed to respect rights included in the European Convention on Human Rights, while courts like the German Federal Constitutional Court (GFCC)

³In CJEU jurisprudence, paragraphs are appropriately sized units of text as they are relatively short, internally homogeneous in terms of subject and voice, easily identifiable through a unique court-assigned identifier, and used by the Court when referring to its own jurisprudence. A different division may be more appropriate when studying another court.

⁴Whether the Swedish labor union's actions were compatible with EU rules on the provision of services.

⁵On the conformity of a Swedish antidumping law with EU rules.

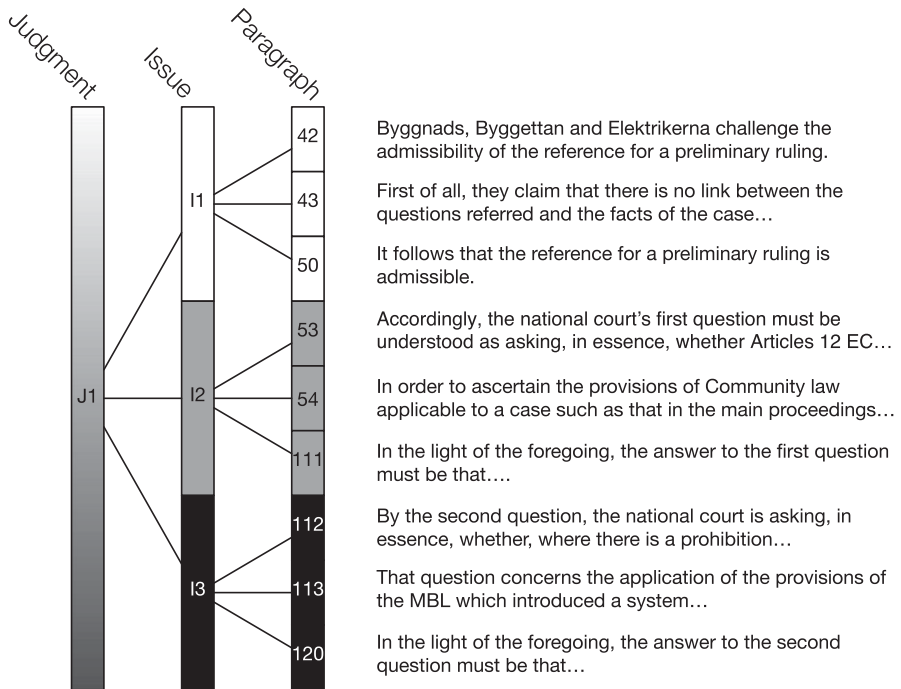


Figure 1. A simple example of a judgment (J1), Case C-341/05, *Laval un Partneri Ltd. v. Svenska Byggnadsarbetareförbundet et al.*, which consists of 121 numbered paragraphs of text, most of which are omitted to enhance readability. Whereas we can determine that the judgment addresses three legal questions, the issue layer (I1–I3) allows us to identify the paragraphs that are associated with each question.

typically interpret different constitutional provisions in their judgments to determine whether a statutory act or ordinance infringed claimants' constitutional rights.

Courts differ not only with regard to the kinds of issues they need to address but also, due to legal and stylistic differences, in how they signal a transition from one issue to another in their judgment texts. Nonetheless, we often find that courts write in a clear, consistent, and structured manner that gives rise to linguistic patterns that mark such transitions. Illustrations in [Tables 1](#) and [2](#) from jurisprudence of the ECtHR, the GFCC, and infringement proceedings at the CJEU show that such patterns tend to be unique to a particular court or even to one type of procedure within a court.⁶

For example, in the ECtHR's judgments, a text block addressing an issue can be distinguished from other issues as the court consistently begins its answer by restating the applicant's claim, using the word "applicant" together with a verb associated with making a claim (e.g., "submit," "complain," or "contend"). Paragraphs concluding whether the State in fact violated a convention article then commonly include words

⁶These examples illustrate how courts use distinctly different words to signal the transition between legal issues; they are internally quite consistent and predictable in their word choices. The GFCC is an interesting example in that it is linguistically consistent and predictable when it starts reasoning on a new issue – in fact, by stating its conclusion – but not when it concludes. Even so, it should be feasible to split its judgments into issues by taking blocks of texts between paragraphs that start an issue.

Table 1. Illustrations of Linguistic Patterns Beginning an Issue in Court Judgments

	European Court of Human Rights	Court of Justice of the European Union: Infringement Cases	German Federal Constitutional Court
Examples	<p>1. The applicants contended that they had been subjected to a collective expulsion without an individual assessment of their circumstances and in the absence of any procedure or legal assistance.</p> <p>2. The applicant submitted that his dismissal by his employer had been based on a breach of his right to respect for his private life and correspondence and that, by not revoking that measure, the domestic courts had failed to comply with their obligation to protect the right in question.</p> <p>3. The applicant complained that he had been denied access to a tribunal to defend his rights in relation to his premature dismissal as President of the Supreme Court.</p>	<p>By its first complaint, the Commission alleges that the French Republic failed to fulfil its obligations under Article 7(1) of Directive 75/442 and Article 6(1) of Directive 91/689 in that it failed to draw up waste management plans for the whole of its territory.</p> <p>By its first head of claim, the Commission asks the Court to declare that, by reserving access to the profession of notary exclusively to its own nationals, the Republic of Austria has failed to fulfil its obligations under Article 43 EC and the first paragraph of Article 45 EC.</p> <p>By its second complaint, the Commission maintains that the Kingdom of the Netherlands has not adopted any measure for the purpose of transposing Article 22(1) of the Directive.</p>	<p>[107] The constitutional complaints are admissible. [111] The constitutional complaints are unfounded.</p> <p>[85] The constitutional complaints of the complainants in proceedings I to III are admissible to the extent that they [...]. For the rest, the constitutional complaints are inadmissible [...].</p> <p>[97] The constitutional complaints of the complainants in proceedings I to III are well-founded to the extent that they [...]. For the rest, the constitutional complaints are [...] unfounded.</p> <p>[56] The constitutional complaint is admissible. [86] The constitutional complaint is well-founded.</p>
Common patterns	applicant; submit/complain/contend that	by its [order] complaint/claim; the Commission (allege/ask/submit/claim); that the [Member State]; infringe/failed to comply/fulfill	constitutional complaint(s); admissible/inadmissible; unfounded/well founded

The table provides typical examples of paragraphs starting an issue within judgments from three courts and the associated, recurring linguistic patterns we can find in these paragraphs.

Table 2. Illustrations of Linguistic Patterns Concluding an Issue in Court Judgments

	European Court of Human Rights	Court of Justice of the European Union: Infringement Cases	German Federal Constitutional Court
Examples	<ol style="list-style-type: none"> 1. Accordingly, there has been no violation of Article 4 of Protocol No. 4. 2. There has therefore been a violation of Article 1 of Protocol No. 1 to the Convention. 3. There has accordingly been a violation of the applicant's right of access to a court, as guaranteed by Article 6 § 1 of the Convention. 	<p>It follows that the Commission's first complaint, alleging breach of the second subparagraph of Article 19(1) TEU, must be upheld.</p> <p>It follows that the second complaint is well founded, to the extent just stated in the preceding paragraph.</p> <p>In the light of all the above considerations, the first head of claim is well founded.</p>	<p>...</p> <p>...</p> <p>...</p>
Common patterns	there has accordingly/therefore been a/no violation of	it follows that/in the light of; complaint/claim; well found/upheld	...

The table provides typical examples paragraphs concluding an issue within judgments from three courts and the associated, recurring linguistic patterns we can find in these paragraphs.

such as “accordingly” or “therefore” together with the patterns “been a violation” or “been no violation.” We should highlight that not all judgments texts follow patterns that allow for the identification of both paragraphs beginning and concluding an issue (as we typically find it in the CJEU's judgments on direct actions and preliminary references; see below). The GFCC is an interesting example here in that it is linguistically consistent and predictable when it starts reasoning on a new issue – in fact, by stating its conclusion – but not when it concludes. Even so, given the GFCC consistently uses particular linguistic patterns it should be feasible to split its judgments into issues by taking blocks of texts between paragraphs that start an issue. Overall, although the legal questions considered by courts such as the CJEU, the GFCC, or the ECtHR are distinctly different, each court makes use of recurring linguistic patterns, and we show below that once we identify the relevant legal context we can exploit these patterns to split judgments into issues.

Going through the trouble of splitting judgments into issues has several uses. First, we can classify what topic the issue concerns by analyzing the associated text or references, for example, using unsupervised approaches such as topic modeling or network-based clustering on the basis of references. Thus, we distinguish between the clustering problem of what text deals with the same legal question (*issues*) and the classification problem of determining the nature of that question (*topics*). While there is no objectively correct number of topics, in order for topics to be a useful analytical tool there should obviously be significantly fewer topics than issues. Second, it allows case-law references to be analyzed on an issue level and for the construction

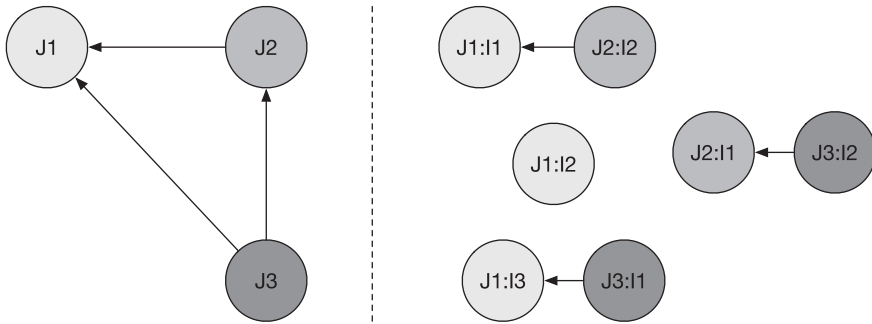


Figure 2. On the left, an example of a judgment-to-judgment network containing three judgments where the middle one (J2) contains a reference to the oldest (J1) and the newest (J3) contains references to both. On the right, an example of a issue-to-issue network based on the same judgments and references but split by issue.

of issue-to-issue citation networks. **Figure 2** illustrates how such a network would more accurately capture relevant references between judgments that allows for more accurate representation and analysis of network structure and centrality. It would also enhance the ability to assess how central a judgment is in the context of a particular topic as judgments must not entirely belong to a single cluster.

Splitting judgments using supervised classification

A researcher familiar with the jurisprudence of a court is capable of reading a judgment and assigning paragraphs within the text to distinct issues. Such manual content analyses are resource intensive, especially when researchers are dealing with large numbers of judgments (see Dyevre 2020). We argue that investing these efforts pays off, particularly where manual coding can be facilitated by computer-assisted methods. Existing research has shown that, once an adequate volume of manually coded data is available, machine learning classifiers can be trained to replicate even complex coding tasks on new data (see Lowe et al. 2011; Anastasopoulos and Bertelli 2020).

In the following section, we draw on our experience of working with the CJEU's judgments in preliminary reference proceedings to highlight the characteristics of judgments that allow researchers to rely on supervised machine learning classification and mitigate the costs of manually splitting large volumes of judgments into issues.

Issues in the CJEU's preliminary rulings

The CJEU's decisions in so-called preliminary reference proceedings have left a deep imprint on the legal systems of EU Member States (Craig and de Búrea 2020). As national courts may (and, in some instances, must) submit a preliminary reference to the CJEU concerning the interpretation of EU law, the CJEU was able to enroll national courts as decentralized "enforcers of EU law" (Craig de Búrca 2020, 497) and expand the reach of EU law even against the interests of Member States (see Alter 2001; Weiler 1994; Stone Sweet and Brunell 1998).

National courts often submit multiple questions concerning different aspects of EU law to the CJEU within a single reference, and the Court's judgments typically comprise a discussion of the relevant national legal context and details from the original case before the national court. A researcher studying CJEU judgment texts to learn from its answers to national courts therefore needs to process the texts prior to analysis to avoid mixing in text elements that have little connection to their research questions.

We hired four research assistants with a background in European law to read a sample of CJEU's judgments in preliminary reference proceedings and separate text segments comprising the CJEU's answers to national court questions from other elements of the judgments (e.g., national legal contexts, case facts, etc.). Our research assistants were instructed to identify paragraphs belonging to one of four classes within each judgment: (1) paragraphs introducing the CJEU's response to a national court's referred question (*question_start*), (2) paragraphs stating the CJEU's concluding response to the national court's question (*question_stop*), (3) and paragraphs stating that a national court question does not require an answer from the CJEU (*question_noanswer*).⁷ All remaining paragraphs in the judgment were assigned to a residual category (*residual*). Table 3 provides typical examples of the semantic structure of these paragraph classes. Section A in the online appendix provides further details on the process of our research assistants' manual coding of

Table 3. Coded Paragraph Classes in CJEU Preliminary Rulings

Paragraph class	Typical examples
<i>question_start</i>	By the second question, the national court is asking, in essence, whether, where there is a prohibition in a Member State against trade unions undertaking collective action with the aim of having a collective agreement between other parties set aside or amended, Articles 49 EC and 50 EC preclude that prohibition from being subject to the condition that such action must relate to terms and conditions of employment to which the national law applies directly [...].
<i>question_stop</i>	In the light of the foregoing, the answer to the second question must be that, where there is a prohibition in a Member State against trade unions undertaking collective action with the aim of having a collective agreement between other parties set aside or amended, Articles 49 EC and 50 EC preclude that prohibition from being subject to the condition that such action must relate to terms and conditions of employment to which the national law applies directly.
<i>question_noanswer</i>	In the light of the answer to the second question, it is not necessary to answer the third question.
<i>residual</i>	In that regard, it must be observed that, in principle, blockading action by a trade union of the host Member State which is aimed at ensuring that workers posted in the framework of a transnational provision of services have their terms and conditions of employment fixed at a certain level, falls within the objective of protecting workers.

Note: Trained research assistants were asked to identify paragraphs marking the beginning of the CJEU's answer to a national court's referred question (*question_start*), and the Court's concluding answer to that question (*question_stop* and *question_noanswer*). All remaining paragraphs were assigned to a residual category (*residual*).

⁷Typically because the CJEU had already provided its answer in response to a separate national court question.

Table 4. Common Features for Paragraph Classes

Paragraph class	Most frequent features (three-grams)
<code>question_start</code>	<code>the_refer_court</code> , <code>be_interpret_as</code> , <code>must_be_interpret</code> , <code>refer_court_ask</code> , <code>in_essenc_whether</code>
<code>question_stop</code>	<code>the_answer_to</code> , <code>answer_to_the</code> , <code>must_be_interpret</code> , <code>be_interpret_as</code> , <code>to_the_question</code>
<code>question_noanswer</code>	<code>there_is_no</code> , <code>is_no_need</code> , <code>no_need_to</code> , <code>to_the_first</code> , <code>question_there_is</code>
<code>residual</code>	<code>eu_paragraph_and</code> , <code>the_main_proceed</code> , <code>in_the_main</code> , <code>to_that_effect</code> , <code>see_to_that</code>

Note: The column “Most frequent features (three-grams)” shows the five most frequent three-grams per paragraph class (N13,797).

the CJEU’s judgments, illustrates with examples how these paragraph classes are embedded in the judgment texts, and discusses the intercoder reliability of the coding.

Once our research assistants had coded all paragraphs within a judgment, we were able to identify text segments that concern a particular *issue*. A text segment capturing an issue begins at the paragraph marking the start of the Court’s answer (`question_start`), ends at the next paragraph marking the conclusion of the Court’s answer (`question_stop`), and comprises all paragraphs between these two. By the time of writing, our hand coders had completed their manual coding task for a sample of 1,080 preliminary references lodged with the CJEU between 1998 and 2011. In total, hand coders had identified 1,804 paragraphs of the class `question_start`, 1,804 corresponding paragraphs of the class `question_stop`, and 189 paragraphs of the class `question_noanswer`. We processed the hand-coded paragraphs’ texts, removed numbers and punctuation, stemmed each term, and tokenized terms into three-grams. We then identified the most frequently occurring three-grams per paragraph class, displayed in Table 4. Note that the most frequent terms for the three classes `question_start`, `question_stop`, and `question_noanswer` displayed in Table 4 appear to have plausible connections to their respective paragraph classes (yet, note also that some terms frequently appear in more than one class, a possible complication we address in Section 3.2).

We show in the following section that these linguistic similarities across paragraphs within the same class allow us to put the collected data to work and train a supervised machine learning classifier that can replicate our research assistants’ task with high accuracy.

Classifying paragraphs

Our data comprise a total of 13,797 hand-coded paragraphs from our sample of 1,080 judgments the CJEU issued in preliminary reference proceedings.⁸ After processing the text as described above and dropping any features that occur in only 10 or fewer paragraphs, we randomly split our data for each paragraph class in half to create

⁸In addition to all hand-classified paragraphs described above, our text data include 10,000 randomly sampled paragraphs classified as `residual`. Paragraphs of the `residual` class are by far the most frequent in the CJEU’s preliminary rulings. By randomly sampling 10,000 paragraphs of this class, we take account of this fact, while keeping the overall data volume low enough to keep the time it takes to compute the classifiers and memory requirements at a reasonable level.

training and test sets to evaluate the performance for three supervised machine learning classifiers: a naive Bayes classifier, a random forest model, and a feedforward neural network.

Naive Bayes classifiers are relatively simple machine learning classifiers that can be fit easily, while providing often reasonable performance (see Kim et al. 2006). Random forests and neural networks are computationally more demanding classifiers, yet well suited to learn more complex patterns for classification tasks. All three classifiers learn from patterns in sparse document-feature matrices, representing paragraphs as bags of words, ignoring the sequence of features. In addition, we programmed a convolutional neural network (CNN) and a long short-term memory (LSTM) network to solve our classification problem, which incorporate the sequence of features when training, but found that the bag-of-words approaches outperform these classifiers (see [appendix Section B](#)). We programmed all classifiers in R, relying on the **quanteda**, **randomForest**, and **keras** packages. All replication material, including text data, R code and files of the trained models, are made available in the [supplementary material](#). Details on the tuning process to define optimal hyperparameters for both the random forest model and the feedforward neural network are provided in [Section B](#) of the appendix.

Performance metrics for the naive Bayes classifier, the random forest model, and the feedforward neural network are reported for each paragraph class in [Table 5](#).⁹ [Table 5](#) shows that the naive Bayes classifier performs reasonably well for the paragraph classes `question_start`, `question_stop`, and `residual` yet poorly for `question_noanswer`. This does not surprise as there are only 94 paragraphs classed as `question_noanswer` in our training data, well below the numbers of the remaining paragraph classes – and arguably below adequate numbers to train a machine learning classifier. However, turning to the metrics for the random forest model and the neural network, we can see that both classifiers perform remarkably well across all four paragraph classes, including `question_noanswer`. We find that the more sophisticated classifiers can effectively replicate the coding decisions of our research assistants (i.e., F_1 metrics for both classifiers are well above or close to 0.90 across the four paragraph classes).

The key to the successful classification is the CJEU's consistent use of linguistic patterns. Patterns such as “by its question the referring court asks in essence” or the “the answer to the question must be” are characteristic of paragraphs marking the beginning and concluding paragraphs of the Court's reasoning on legal issues, while linguistic patterns such as “there is no need to answer” occur virtually exclusively in paragraphs our hand coders had identified as belonging to the class `question_noanswer`. These patterns allow machine learning classifiers to distinguish between paragraph classes even when the available amount of training data is limited. To illustrate, we plot the 30 most important features for the classification of paragraphs for our random forest model in [Figure 3](#). We can see that all three-grams identified as the most important features are connected to linguistic patterns characteristic of the respective paragraph classes. Recall also that some of the features listed in [Figure 3](#)

⁹The classifiers' precision refers to the proportion of predictions (for each class) that were predicted correctly by the model, that is, true positives/(true positives + false positives). Recall refers to the proportion of actual instances of a class the model predicted correctly, that is, true positives/(true positives + false negatives). The F_1 metric combines the results for precision and recall into a single value that gives us a sense of the overall quality of the classifier for each class, ranging from 0–1.

Table 5. Classification Performance for Paragraph Classes in Test Set (N6,898)

Paragraph class	Naive Bayes	Random forest	Neural network
question_start			
Precision	0.77	0.96	0.95
Recall	0.87	0.84	0.86
F ₁	0.82	0.90	0.90
question_stop			
Precision	0.71	0.95	0.96
Recall	0.96	0.93	0.93
F ₁	0.82	0.94	0.95
question_noanswer			
Precision	0.14	0.86	0.96
Recall	0.96	0.87	0.89
F ₁	0.25	0.86	0.92
residual			
Precision	0.97	0.96	0.96
Recall	0.78	0.99	0.99
F ₁	0.87	0.97	0.97

Note: All classifiers were trained on identical 6,899 × 10,175 document-feature matrices (paragraphs × three-grams).

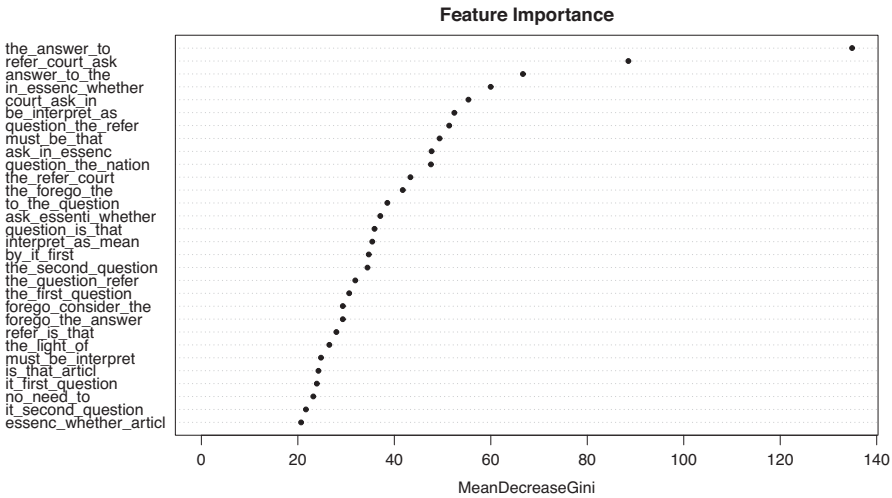


Figure 3. Random forest model’s feature importance ordered by mean decrease in Gini coefficient.

frequently appear in more than one paragraph class (e.g., “must_be_interpret” frequently appears in paragraph classes question_start and question_stop). Although initially a cause of concern, we find that all three classifiers rarely struggled to distinguish between the classes question_start and question_stop but instead that most misclassifications occur between the class residual and the remaining classes, respectively.^{10,11}

¹⁰The F₁ metric is calculated via the formula $F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$

¹¹For our fitted random forest model, we find that only 6 of the 902 paragraphs of the class question_start are misclassified as question_stop, while only 1 of the 902 paragraphs of the class question_stop is misclassified as question_start. Similarly, for our fitted neural network, we find

The results we present here suggest that, instead of instructing our research assistants to manually classify paragraphs in the entire sample of judgments of interest, the task can be performed for a subset of this sample and the manually coded data can be used to train a classifier to complete the job for the remaining judgment texts. In our case, rather than having our research assistants classify paragraphs from all 2,460 preliminary rulings the CJEU issued between 1998 and 2011, we were able to limit their task to a sample of roughly 1,000 judgments.

Limitations

While these results are encouraging, there are limitations to our approach. We need to be confident that the same linguistic patterns present in the training data also appear in the documents that ought to be classified by the trained classifier. For our purposes, given we asked our research assistants to code a sample of preliminary references lodged with the CJEU between 1998 and 2011, the use of the classifier is, strictly speaking, limited to preliminary rulings issued within this time frame.¹² Without evidence suggesting that the linguistic patterns driving the performance of the classifiers discussed above are present in judgments outside the time frame of our analysis (for instance, via manual validation of a number of out-of-sample classifications), we would caution against using the same trained classifier to predict paragraph classes in older or more recent judgments or even judgments issued in other procedures (e.g., direct actions) and other courts. Put simply, the use of a trained classifier is limited to the temporal-, institution- and procedure-specific context of the training data.

Our approach does not make manual coding obsolete, and the investigating researcher still has to carefully select both the judgments on which a classifier is trained and the judgments that ought to be classified for a particular research project. However, even within the limitations outlined above, manually coding a subsample of judgment texts, training a machine learning classifier, and then classifying paragraphs in the remaining set of judgments of interest saves resources and is worth the effort. We believe that classifying paragraphs to split judgments into issues is beneficial for research in the domain of law and politics, and we now turn to demonstrate these benefits for the CJEU's preliminary rulings in the following sections.

Putting issues to the test: The value of issue splitting

In this section, we highlight the practical benefits of issue splitting and show that working with issues as data is preferable to working with entire judgments when

that only 5 of the 902 paragraphs of the class `question_start` are misclassified as `question_stop`, while only 2 of the 902 paragraphs of the class `question_stop` are misclassified as `question_start`. Hence, even though some features are ambiguous, our evidence suggests that they help to separate the classes `question_start` and `question_stop` from the most common class residual, while other — less ambiguous — features then allow classifiers to separate the former paragraph classes from each other.

¹²Given our data collection was embedded in a larger research project and research assistants completed additional data collection tasks besides annotating paragraphs, we had little discretion over the time frame from which our training sample was drawn, yet it allowed us to link information we found within the different text segments that deal with particular legal issues to other information on these legal issues (see our application in Section 6).

analyzing court decisions. First, we set out to identify topics in a sample of the CJEU's case law and compare the results from a topic model estimated on complete judgments and a topic model estimated on judgments that had been split into issues. Second, we identify clusters of case law connected through references, again comparing the results of a network analysis performed on complete judgments and a network analysis of issue-split judgments.

Here, we use a sample dataset consisting of all 206 CJEU judgments in preliminary proceedings concerning free movement of goods referred to the CJEU between 1998 and 2011.¹³ Free movement of goods is one of the fundamental freedoms that serve as the pillars of the internal market that, in turn, constitutes the heart of the European Union and EU law. The right to free movement of goods is enshrined in the Treaties, but the language is enigmatic, which has given rise to a significant body of CJEU case law clarifying the proper interpretation of EU law.

Classification using LDA topic modeling

Judgments concern distinct legal question and knowing what those questions are and which judgments address which question is arguably the most important knowledge that lawyers have, but they are also important to scholars seeking to understand how courts behave in different areas of law. Automated text analysis allows us to predict topic labels from text (see, e.g., Aletras et al. 2016; Ashley and Bruninghaus 2009; Salaün et al. 2020), typically using latent Dirichlet allocation (LDA) topic modeling, which summarizes a corpus assuming an unknown structure of topics reflected in the individual documents of the corpus. Previous studies have used LDA topic modeling to identify topics in different bodies of case law (Carter et al. 2016; Lauderdale and Clark 2014; Panagis et al. 2016; Soh et al. 2019; Trappey et al. 2020; Venkatesh and Raghuvver 2013).

However, applying topic models to entire judgment texts may mask significant differences in the topics addressed in its constituent issues. To illustrate, we trained an LDA model and applied it to the CJEU's judgment in *Fazenda Pública*, a judgment addressing two issues. Figure 4 shows that, applied to the entire judgment text, the model indicates that the judgment predominantly addresses Topic 4 and to a lesser extent Topic 8. However, when applied to its two constituent issues, a different and clearer pattern emerges: Issue 1 predominantly concerns Topic 4, and issue 2 is dominated by Topic 8.

In the following, we use LDA topic modeling to compare how well a topic model performs when classifying complete judgments and issue-split judgments. We examine whether and to what extent a topic model is capable of identifying the topic of an issue with greater probability than for the judgment to which the issue belongs. Specifically, we compare the maximum topic probability of the issue to the maximum topic probability of the judgment, arguing that the better approach is the one that achieves the highest probability.¹⁴

¹³These proceedings were identified using procedure and subject matter classifications provided in the EU database Eur-Lex.

¹⁴Under ideal conditions, that is, applying a perfect model to internally consistent documents, a document would with 100% probability address exactly one topic. Under real-world conditions, this is not possible and maximum topic probability in absolute terms will vary and frequently be significantly lower.

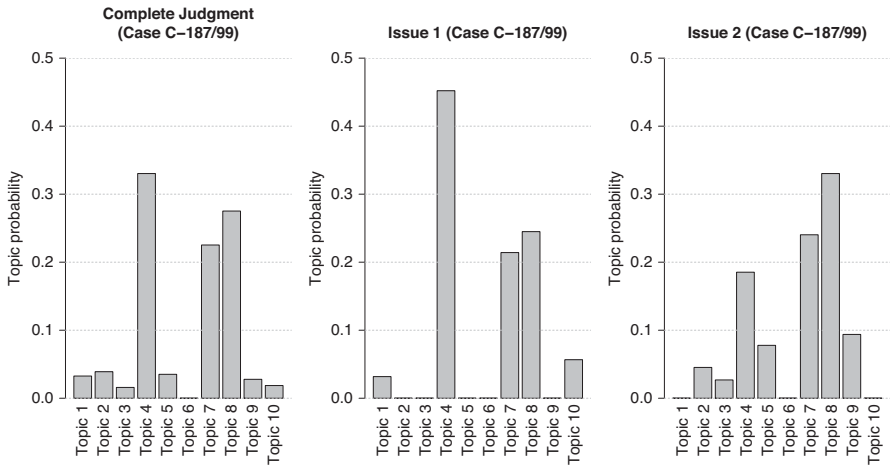


Figure 4. The figure shows the probability distribution for the topics in the model for the CJEU's judgment in Case C-187/99 *Fazenda Pública v. Fábrica de Queijo Eru Portuguesa Lda, interverer Ministério Público, ECLI: EU:C:2001:114*.

We first trained a model based on a random selection of 165 of the 206 judgments for our training set. The process involves identifying words in the corpus that appear more frequently together in the document, and the model is trained over multiple iterations to provide an efficient representation of the entire corpus as well the documents of which it consists (Blei et al. 2003). Following standard text preprocessing,¹⁵ we trained an LDA topic model that, in order to ensure that the model was not biased in favor of complete judgments or issues, included all judgment text twice: with the entire judgment as document and with the issues as documents.¹⁶ The resulting model essentially represents the probability that certain words appear together under 10 topics in free movement of goods case law. In Section C of the online appendix, we describe the topics that the LDA model identified and show that these relate to several distinct themes we would expect to find in CJEU judgments concerning the free movement of goods.

This model was then applied to classify the text of the remaining 41 judgments, our test set, identifying topics in unseen free movement of goods judgments.¹⁷ To test the efficacy of issue splitting, the model was applied to classify (i) the complete text of each judgment, (ii) the text of each issue, and (iii) a filtered version of each judgment consisting only of text belonging to an issue.¹⁸ This returns a classification of each document (here, either a judgment or an issue) expressed as a probability that the document addresses each of the topics in the model. The approach used for

¹⁵Consisting of the removal of stop words, punctuation, and digits, along with converting all letters to lower case, lemmatization, and stemming.

¹⁶This resulted in a document-feature matrix consisting of 564 rows (documents) by 8,790 columns (features). We used the package `topicmodels` for R and set the number of topics $k = 10$.

¹⁷The testing set was processed and structured in the same way as the training set.

¹⁸Compared to the complete judgment, the filtered judgment does not contain, for example, the presentation of the court and parties, the facts of the case, the relevant legislation, and the Court's decision on costs.

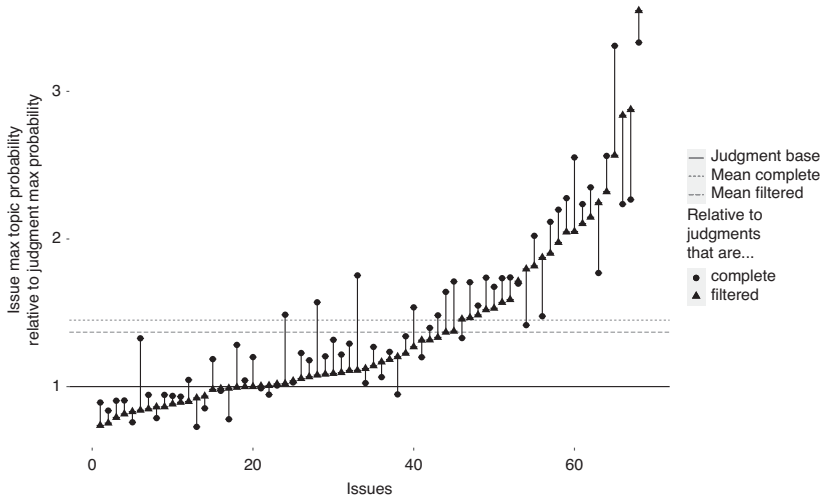


Figure 5. Each observation is an issue in a judgment with at least two issues. Its placement on the y-axis shows its max topic probability relative to the max topic probability of the entire judgment that it belonged to. Issues on average achieve a 45% higher maximum probability than the complete judgment that they belong to and 36% higher than the filtered judgments.

classifying the text is thus constant, the only changing variable being whether the judgment text is analyzed in its entirety or split into issues. We then match and compare the maximum topic probability of the issues to that of the judgments to which they belong, both the complete text and the filtered version.

Results for judgments containing more than one issue are displayed in Figure 5.¹⁹ With few exceptions, the topic model performs significantly better on issues than on judgments in the vast majority of cases. The maximum topic probability for issues is on average 45% higher compared to that of the complete judgments they were taken from and, in some instances, 100–400% higher.

This is in part attributable to a key advantage of issue splitting, “noise-filtering.” Complete judgments contain portions of text unrelated to any legal question, such as presentations of the actors involved and discussion of litigation costs, and removing these improves accurate issue classification. However, even if judgments are allowed to benefit from this, the maximum topic probability for issues are on average 36% higher than the maximum topic probability of the filtered version of the judgments to which they belong.²⁰ Thus, even when disregarding the filtering function, there is a significant advantage to issue splitting *per se* when it comes to text classification. In practical terms, this means that scholars using issue splitting will be able to more accurately classify judgments and, consequently, draw more accurate conclusions. Finally, while we demonstrated the benefits for one specific text classification approach, we expect other approaches to benefit similarly.

¹⁹No comparison is possible for single-issue judgments.

²⁰As discussed in greater detail in the appendix, the level of improvement varies depending on the specific topic model. However, the levels offered by this particular model are common.

Community detection using network analysis

The last decade has seen a significant rise in the use of network analysis for studying courts, for example, on the basis of references between cases, but these studies have generally suffered from a lack of nuanced data (Panagis and Sadl 2015; Winkels et al. 2011). For example, following the approach of studies of other courts (see Fowler et al. 2007; Lupu and Voeten 2012; Winkels et al. 2011), Derlén and Lindholm (2014) studied the CJEU's references to its own previous decisions and concluded that the systemic importance of some decisions, such as *Bosman*,²¹ has been overlooked.

A key use for network analysis is community structure detection, also known as clustering, to identify “densely connected groups of vertices, with only sparser connections between groups” (Newman 2006a, 8577). Community detection has a broad range of uses, including, when applied to case law citation networks, being able to identify communities of judgments addressing similar topics (Mirshahvalad et al. 2012). The leading measurement for assessing the quality of the communities is *modularity*, a value between 0 and 1 calculated by taking the fraction of edges within communities minus the expected fraction if edges were distributed randomly (Newman and Girvan 2004, 7).

We use modularity to evaluate the impact of issue splitting on community detection. We construct two citation networks based on references to CJEU judgments found in the test set described above: one based on references from and to complete judgments (*judgment network*), one based on the same references but from issues to judgment (*issue network*). We then apply six leading community detection algorithms to both networks and compare the modularity. Table 6 shows that the communities in the issue network consistently have greater internal density and lower external density than the communities in the judgment network, in most cases around 10%.

This means that a citation network based on issue-split judgments will more accurately represent the structure of the case law. A more accurate understanding of which judgments belong to the same community is practically important, both as it is a form of topic identification and for the reasons explained immediately above, but also as it enables researchers to more accurately identify the centrality of judgments on a topics.

Application: The CJEU's strategic references to case law

In this final section, we show how moving from judgments to issues as units of analysis affects the specification and estimation of statistical models used to test theories of judicial behaviour. Existing research has argued that judges at the CJEU are aware that EU Member States are instrumental in the implementation of its judgments and may attempt to override unfavorable decisions of the Court (Garrett et al. 1998; Carrubba et al. 2008; Carrubba and Gable 2015; Larsson and Naurin 2016). The CJEU is therefore sensitive – and evidently, responsive – to the interests of

²¹Judgment of the Court of 15 December 1995. *Union royale belge des sociétés de football association ASBL v. Jean-Marc Bosman, Royal club liégeois SA v. Jean-Marc Bosman and others and Union des associations européennes de football (UEFA) v. Jean-Marc Bosman*. Reference for a preliminary ruling: Cour d'appel de Liège - Belgium. Case C-415/93.

Table 6. Modularity of Communities in Judgment Network and Issue Network

Network	Infomap	Walktrap	Louvain	Lead eigenvector	Fast greedy	Edge betweenness
Judgment	0.72	0.73	0.76	0.64	0.76	0.76
Issue	0.76	0.78	0.83	0.72	0.84	0.83

Note: The table displays the modularity of communities in the judgment network and issue network respectively using algorithms introduced by, in order, Rosvall and Bergstrom (2008); Pons and Latapy (2006); Blondel et al. (2008); Newman (2006b); Clauset et al. (2004); Newman and Girvan (2004). Higher values are preferred over lower values.

Member States. Drawing on this literature, Larsson et al. (2017) argue that the CJEU uses legal justifications as a legitimation strategy when its decisions run counter to Member States' interests: "[T]he Court argues more carefully, by means of reference to precedent, when it takes decisions that conflict with the positions of EU governments" (Larsson et al. 2017, 881).

Their empirical analysis draws on two separate datasets. Data provided by Derlén and Lindholm (2014) capture citation patterns between preliminary rulings and is complemented by information on the CJEU and Member States' positions on the questions addressed in preliminary rulings between 1998 and 2011 (see Naurin et al. 2015). While the units of observation for the CJEU's references to precedent are judgments, actors' positions are expressed for the specific questions national courts had referred to the CJEU, with a single CJEU judgment typically dealing with multiple national court questions. Larsson et al. (2017) solve this discrepancy in units of observations by aggregating data on actors' positions to the judgment level.

Avoiding aggregation of data at the judgment level is potentially critical, as we demonstrate with the following example. In Case C-324/99 *DaimlerChrysler AG v. Land Baden-Württemberg*, the CJEU considered four distinct issues within a single judgment. On two of these issues, Member States supported the Court's conclusion; on another issue Member States held no clearly identifiable position, while Member States opposed the Court's answer on the final issue. On the judgment level, aggregating these positions suggests that overall Member States supported the CJEU's conclusions, and the judgment-level data show that the Court made four references to its own case law. Our issue-level data reveal that all four of these references were made in response to the first issue, while the Court made no references in its answer to the final issue despite facing opposition from Member States. Such patterns at odds with Larsson et al.'s expectations are lost in aggregation.

Operationalizations of outcome and explanatory variables

We first identify the CJEU's citations of its previous case law in the texts of the preliminary rulings using regular expressions.²² Given our issue-splitting approach provides us with the text blocks for each issue, we can identify citations both at the judgment level, $N = 206$, and the issue level, $N = 487$. Like Larsson et al. (2017), we then construct a variable *Outdegree*, which counts the number of outward citations

²²Citations of case law in the CJEU's jurisprudence follow a consistent pattern, referring to the cited case name (e.g., *Costa v. ENEL*), often followed by a reference to a particular paragraph in that judgment. Using regular expressions, we are able to capture both which judgment as well as which paragraph is cited.

for a particular unit of observation (i.e., a judgment or an issue).²³ Specific case law can be cited multiple times in a judgment, and we count each of these instances. This means that the count of outward citations at the judgment level equals the sum of outward citations across the issues within the judgment.

Following Larsson et al. (2017), we then construct a variable *MS Conflict*, comprising three categories: (1) *in conflict*, indicating that the CJEU favored an interpretation of EU law that would restrict Member States' autonomy while Member States' net position favored an interpretation of EU law that preserved national autonomy; (2) *in favor*, indicating that Member States' net position aligned with the CJEU's position on a ruling concerning national autonomy; and (3) *ambivalent* indicating that no clear implications regarding the effects of legal integration on national autonomy could be drawn for either the CJEU or Member States' position (or both).²⁴

The same approach was used to measure whether or not the CJEU's positions conflicted with positions of the Advocate General (AG) and the European Commission, captured by the variables *AG Conflict* and *Commission Conflict*, respectively. We reconstruct the variables *MS Conflict*, *AG Conflict*, and *Commission Conflict* for our subset of preliminary rulings, both at the judgment level and issue level.²⁵

Estimation

Not every CJEU judgment actually comprises multiple issues. Out of the 206 judgments, 85 contain only one issue, while we find two or more issues discussed in the remaining judgments. Rather than estimating judgment-level regressions after aggregating values for the issue-level predictors, we incorporate the hierarchical structure of our data in the statistical model and estimate a multilevel regression (Gelman and Hill 2007). Some of the control variables included in the original model by Larsson et al. (2017) are measured at the judgment level, and our multi-level regression model can handle predictors at the both issue and judgment level, while accounting for judgment-level variation by allowing intercepts to vary across judgments.

Given that our explanatory variable is a discrete count variable, we estimate a negative binomial multilevel regression model. In light of the relatively large number of multilevel model parameters that need to be estimated for a relatively small dataset,

²³We consider alternative measures to capture the extent to which the CJEU's rulings are embedded in its existing case law, such as the *Hub Score*, a more sophisticated network centrality measure weighing outward citations based on the precedential authority of the cited units (see Easley and Kleinberg 2010; Lupu and Voeten 2012), as our outcome variable in analyses presented in the online appendix.

²⁴To arrive at Member States' net position on an issue, positions were weighted by the voting power of the Member States in the Council. For full details regarding the construction of the variable, see Larsson et al. (2017, 893).

²⁵Note that as mentioned above, Member State positions on the issues discussed by the CJEU were originally coded for each question a national court had referred to the CJEU. The CJEU occasionally answers several national court questions collectively within a text block, which we would consider an issue in line with our discussion in Section 0. Hence, the original units of observation for the data on actors' positions do not always perfectly match our concept of issues. While we identified a total of 451 issues in our subset of preliminary rulings, data provided by Naurin et al. (2015) indicate that the CJEU dealt with 487 national court questions in this subset of CJEU rulings. We matched the positions coded by Naurin et al. (2015) to the issues we identified in the judgment text without aggregating where the CJEU answers multiple national court questions in a single text block.

we follow advice by Gelman and Hill (2007) and opt for a Bayesian estimation of the model parameters' posterior distributions, specifying uninformative priors and running four chains with 10,000 sampling iterations. All estimations are implemented through the **rstanarm** package for R.

Results

Figure 6 plots the regression coefficients' posterior means along with their 95% highest probability density (HPD) intervals for the judgment-level and the multilevel regression. Reference categories for the three variables *MS Conflict*, *AG Conflict*, and *Commission Conflict* are observations indicating no conflict between the CJEU's position and the positions of Member States, the AG, and the Commission, respectively.

We can spot two patterns in Figure 6. First, coefficient estimates for the categories indicating conflict between the CJEU and respective actors' positions for *MS Conflict*, *AG Conflict*, and *Commission Conflict* are overall similar across the judgment-level and multilevel regressions. The CJEU is more likely to reference its own case law when its position conflicts with the expressed positions of Member States and the AG, relative to instances in which their respective positions align, while no such effects are discernible when the CJEU's position conflicts with the Commission's position. This evidence is consistent with expectations formulated by Larsson et al. (2017) that the CJEU makes an effort to embed its decisions in existing case law when facing an adverse environment to signal a legal legitimacy of an otherwise controversial decision.

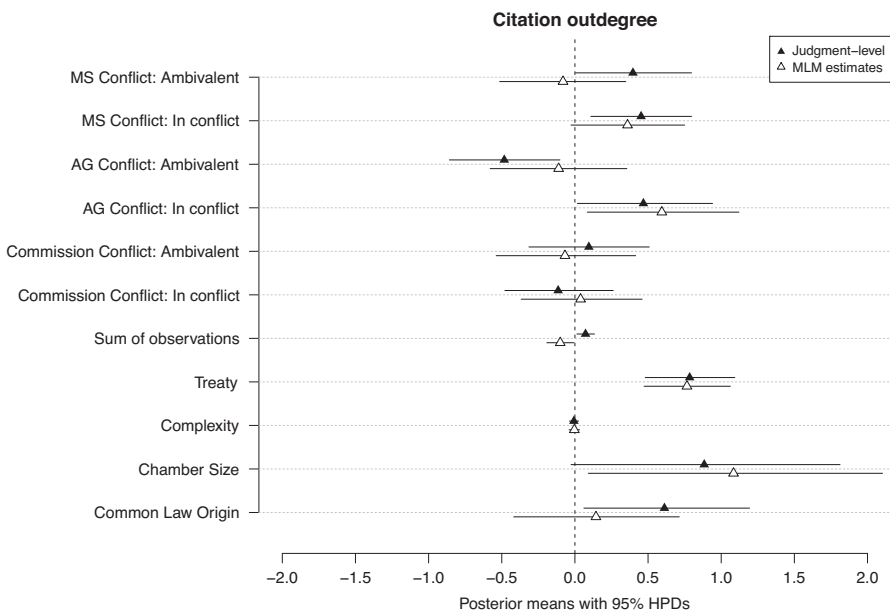


Figure 6. Posterior means with 95% HPD intervals of regression coefficients, displayed for the judgment-level ($N = 206$) and multilevel analyses ($N = 487$). All regression analyses include year fixed effects (not shown here).

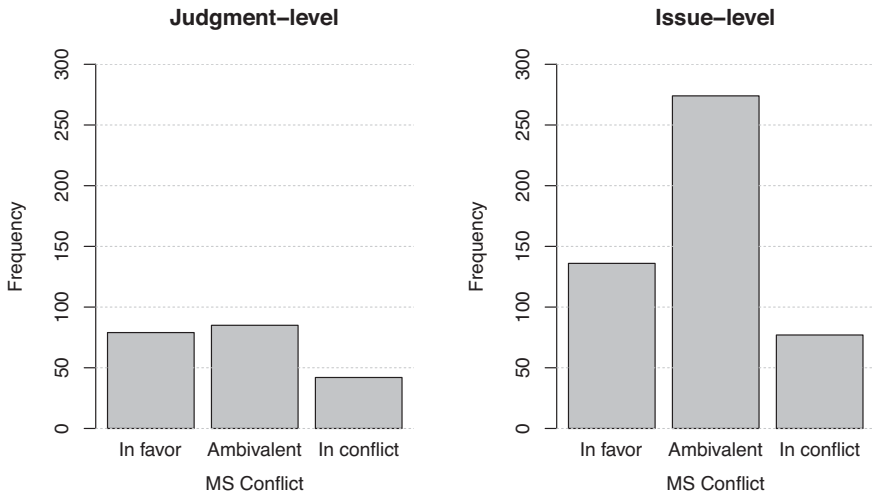


Figure 7. Distributions for the main explanatory variable *MS Conflict* at the judgment level ($N = 206$) and issue level ($N = 487$).

Second, coefficients for *MS Conflict: Ambivalent* and *AG Conflict: Ambivalent* differ markedly between the judgment-level regression and the multilevel regression. Results from the judgment-level regression suggest that the Court makes an additional effort to embed its decisions in existing case law, even when Member States' positions are ambivalent and fewer references to existing case law when the AG's position is ambivalent. The coefficients from the multilevel regression, however, show that neither of these inferences hold once we consider the CJEU's positions on the actual issues discussed in a judgment: The coefficients for *MS Conflict: Ambivalent* and *AG Conflict: Ambivalent* are indistinguishable from zero.

The reason for these differences lies in the nuance in information that is lost when data are aggregated at the judgment level. Figure 7 plots the distribution for the explanatory variable *MS Conflict* at the judgment level and the issue level. Splitting the CJEU's preliminary rulings into issues reveals that for most of the issues considered, the CJEU and EU Member States' positions were ambivalent. Our results suggest that the loss in nuance in information from aggregating data at the judgment level ultimately translates into coefficients that would lead researchers to draw misleading inferences from their analyses.

In the final step, we show that our issue-level data not only uncover substantively important differences to analyses relying on aggregated data but also allows us to make more precise predictions of how many references the CJEU makes to case law in its judgments. We first predict the number of citations for each judgment using the coefficients from our judgment-level regression and compare these predictions to the observed citations. We then predict the number of citations for each issue using our issue-level regression coefficients, sum up the predictions of issues that belong to the same judgment, and again compare these sums to the actually observed citations at the judgment level. Figure 8 shows the distributions of residuals from these two approaches, indicating that residuals from our issue-level analysis are more tightly clustered around zero.

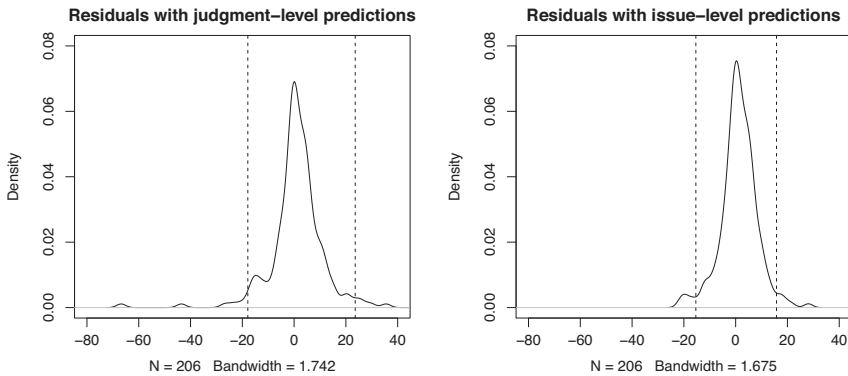


Figure 8. Distributions of residuals for predictions of *Outdegree* at the judgment level. The left panel shows residuals for predictions from the judgment-level model; the right panel shows residuals for predictions from the issue-level model. Vertical dashed lines indicate the 2.5th and 97.5th percentiles of the residuals' distributions.

Conclusion

Scholars of law and judicial politics have urged students of judicial behavior to center their attention on the text of courts' jurisprudence (see, for example, Lax 2011). Tiller and Cross (2006, 523) argue that “[t]he language of the opinion at least purports to establish the rules to govern future cases, but political science researchers have generally disregarded the significance of this language.” Until recently, empirical analyses of judicial behavior have overlooked that “decisions are often most important because of the qualitative changes in law that they effect, rather than because of the decision they provide on the case facing the Court” (Clark and Lauderdale 2010, 871).

However, once scholars shift their attention to the language of court decisions, they are faced with large volumes of text from judgments that commonly address multiple issues. In this contribution, we introduced an approach that structures the text of judgments into clusters of paragraphs that deal with distinct, internally consistent issues. We showed that supervised classification can facilitate the splitting of judgments into issues, exploiting recurring linguistic patterns in judgment texts. Although our approach does not eliminate the need for manual coding, it reduces the time and effort coders would otherwise need to identify distinct issues in a sample of judgments. A key benefit of our approach is that researchers end up with the actual text for each issue that is discussed in a judgment. This opens up a variety of opportunities for empirical research. Rather than having to rely on full judgment texts, which often include more information than we care for, scholars may construct measures connected to relevant aspects of judicial behavior based on word counts, lexical diversity, or sentiment analyses specific to each substantive issue a court considered in its judgment.

Our experience of splitting a subset of the CJEU's preliminary rulings into issues suggests that supervised classification allow us to provide structure to complex judicial decisions without having to read every single word within them, although we are conscious that the context of preliminary references proceedings, with national courts submitting distinct legal questions for the CJEU to resolve, appears particularly well suited to our approach. Nonetheless, we are confident that our

approach can be used or modified to identify similar structures and issues in other courts' jurisprudence as well. Whenever courts use recurring linguistic patterns in their judgments, researchers can employ machine learning classifiers trained to identify such patterns to provide structure to large volumes of unstructured text. Our approach thus helps to reduce the complexity of courts' jurisprudence that would otherwise present obstacles to text-driven research of judicial behavior.

Acknowledgments. We are thankful for thoughtful feedback from Daniel Naurin, Lisa Lechner, Benjamin Engst, Theresa Squatrito, Måns Magnusson, and Andreas Östling on earlier versions of the manuscript. We would also like to thank the two anonymous reviewers and the editor for their useful suggestions on how to improve the manuscript.

Funding Statement. This research was conducted as part of the IUROPA project (www.iuropa.pol.gu.se), financed by the Swedish Research Council Project No. 2018-04215.

Data Availability Statement. All replication material is available at the at the Journal's Dataverse archive.

Supplementary Materials. To view supplementary material for this article, please visit <https://doi.org/10.1086/717421>

References

- Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preotjuc-Pietro, and Vasileios Lampos. 2016. Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective. *PeerJ Computer Science* 2(October): e93.
- Alter, Karen J. 2001. *Establishing the Supremacy of European Law: The Making of an International Rule of Law in Europe*. Oxford: Oxford University Press.
- Anastasopoulos, L. Jason, and Anthony M. Bertelli. 2020. Understanding Delegation through Machine Learning: A Method and Application to the European Union. *American Political Science Review* 114(1): 291–301.
- Ashley, Kevin D., and Stefanies Bruninghaus. 2009. Automatically Classifying Case Texts and Predicting Outcomes. *Artificial Intelligence and Law* 17: 125–165.
- Baum, Lawrence. 2017. *Ideology in the Supreme Court*. Ann Princeton: Princeton University Press.
- Blei, David M., A. Ng, and I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3: 993–1022.
- Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10): P10008.
- Carrubba, Clifford, J., and Matthew Gabel. 2015. *International Courts and the Performance of International Agreements. A General Theory with Evidence from the European Union*. New York: Cambridge University Press.
- Carrubba, Clifford J., Matthew Gabel, and Charles Hankla. 2008. Judicial Behavior Under Political Constraints: Evidence from the European Court of Justice. *American Political Science Review* 102(04): 435–452.
- Carter, David J., James Brown, and Adel Rahmani. 2016. Reading the High Court at a Distance: Topic Modelling the Legal Subject Matter And Judicial Activity of the High Court of Australia, 1903–2015. *University of New South Wales Law Journal* 39: 1300–1354.
- Clark, Tom S., and Clifford J. Carrubba. 2012. A Theory of Opinion Writing in a Political Hierarchy. *The Journal of Politics* 74(02): 584–603.
- Clark, Tom S., and Benjamin E. Lauderdale. 2010. Locating Supreme Court Opinions in Doctrine Space. *American Journal of Political Science* 54(4): 871–890.
- Clark, Tom S., and Benjamin E. Lauderdale. 2012. The Genealogy of Law. *Political Analysis* 20(3): 329–350.
- Clauset, Aaron, M. E. J. Newman, and Cristopher Moore. 2004. Finding Community Structure in Very Large Networks. *Phys. Rev. E* 70: 066111.

- Corley, Pamela C., and Justin Wedeking. 2014. The (Dis)advantage of Certainty: The Importance of Certainty in Language. *Law and Society Review* 48(1): 35–62.
- Craig, Paul, and Gráinne de Búrca. 2020. *EU Law: Text, Cases, and Materials*. Oxford: Oxford University Press.
- Derlén, Mattias, and Johan Lindholm. 2014. Goodbye van Gend en Loos, Hello Bosman? Using Network Analysis to Measure the Importance of Individual CJEU Judgments. *European Law Journal* 20(5): 667–687.
- Dyevre, Arthur. 2020. The Promise and Pitfall of Automated Text-Scaling Techniques for the Analysis of Jurisprudential Change. *Artificial Intelligence and Law*.
- Easley, David, and Jon Kleinberg. 2010. *Networks, Crowds, and Markets*. Cambridge: Cambridge University Press.
- Epstein, Lee, and Jack Knight. 1998. *The Choices Justices Make*. Washington, DC: CQ Press.
- Fowler, James H., Timothy R. Johnson, James F. Spriggs, Sangick Jeon, and Paul J. Wahlbeck. 2007. Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court. *Political Analysis* 15(3): 324–346.
- Garrett, Geoffrey, R. Daniel Kelemen, and Heiner Schulz. 1998. The European Court of Justice, National Governments, and Legal Integration in the European Union. *International Organization* 52(01): 149–176.
- Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Kim, Sang-Bum, Kyoung Soo Han, Hae Chang Rim, and Sung Hyon Myaeng. 2006. Some Effective Techniques for Naive Bayes Text Classification. *IEEE Transactions on Knowledge and Data Engineering* 18(11): 1457–1466.
- Larsson, Olof, and Daniel Naurin. 2016. Judicial Independence and Political Uncertainty: How the Risk of Override Affects the Court of Justice of the EU. *International Organization* 70(2): 377–408.
- Larsson, Olof, Daniel Naurin, Mattias Derlén, and Johan Lindholm. 2017. Speaking Law to Power: The Strategic Use of Precedent of the Court of Justice of the European Union. *Comparative Political Studies* 50(7): 879–907.
- Lauderdale, Benjamin E., and Tom S. Clark. 2014. Scaling Politically Meaningful Dimensions Using Texts and Votes. *American Journal of Political Science* 58(3): 754–771.
- Lax, Jeffrey R. 2011. The New Judicial Politics of Legal Doctrine. *Annual Review of Political Science* 14(1): 131–157.
- Lowe, Will, Kenneth Benoit, Slava Mikhaylov, and Michael Laver. 2011. Scaling Policy Preferences from Coded Political Texts. *Legislative Studies Quarterly* 36(1): 123–155.
- Lupu, Yonatan, and Erik Voeten. 2012. Precedent in International Courts: A Network Analysis of Case Citations by the European Court of Human Rights. *British Journal of Political Science* 42(2): 413–439.
- Medvedeva, Masha, Michel Vols, and Martijn Wieling. 2020. Using Machine Learning to Predict Decisions of the European Court of Human Rights. *Artificial Intelligence and Law* 28: 237–266.
- Mirshahvalad, Atieh, Johan Lindholm, Mattias Derlén, and Martin Rosvall. 2012. Significant Communities in Large Sparse Networks. *PLoS ONE* 7(3): 1–7.
- Naurin, Daniel, Per Cramér, Olof Larsson, Sara Lyons, Andreas Moberg, and Allison Östlund. 2015. *The CJEU Preliminary Reference Procedures Database (1997-2008)*. University of Gothenburg: Centre for European Research (CERGU).
- Newman, Mark E. J. 2006a. Modularity and Community Structure in Networks. *PNAS* 103(23): 8577–8582.
- Newman, M. E. J. 2006b. Finding Community Structure in Networks Using the Eigenvectors of Matrices. *Phys. Rev. E* 74: 036104.
- Newman, M. E. J., and M. Girvan. 2004. Finding and Evaluating Community Structure in Networks. *Phys. Rev. E* 69(Feb): 026113.
- Owens, Ryan J., and Justin Wedeking. 2011. Justices and Legal Clarity: Analyzing the Complexity of U.S. Supreme Court Opinions. *Law and Society Review* 45(4): 1027–1061.
- Owens, Ryan J., Justin Wedeking, and Patrick C. Wohlfarth. 2013. How the Supreme Court Alters Opinion Language to Evade Congressional Review. *Journal of Law and Courts* 1(1): 35–59.
- Panagis, Ioannis, and Urska Sadl. 2015. The Force of EU Case Law: An Empirical Study of Precedential Constraint. In *Legal Knowledge and Information Systems: JURIX 2015: The Twenty-Eighth Annual Conference*, edited by A. Rotolo, pp. 71–81. The Hague: IOS Press.

- Panagis, Yannis, Martin Lolle Christensen, and Sadl Urska. 2016. On Top of Topics: Leveraging Topic Modeling to Study the Dynamic Case-Law of International Courts. In *Legal Knowledge and Information Systems: Frontiers in Artificial Intelligence and Applications*, edited by F. Bex and S. Villata, pp. 161–166.
- Pons, Pascal, and Matthieu Latapy. 2006. Computing Communities in Large Networks Using Random Walks. *Journal of Graph Algorithms and Applications* 10(2): 191–218.
- Rosvall, Martin, and Carl T. Bergstrom. 2008. Maps of Random Walks on Complex Networks Reveal Community Structure. *PNAS* 105(4): 1118–1123.
- Salaün, Olivier, Philippe Langlais, Andrés Lou, Hannes Westermann, and Karim Benyekhlef. 2020. Analysis and Multilabel Classification of Quebec Court Decisions in the Domain of Housing Law. In *Natural Language Processing and Information Systems*, edited by E. Métais, F. Meziane, H. Horacek, and P. Cimiano, pp. 135–143. Cham: Springer International Publishing.
- Segal, Jeffrey A., and Harold J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited*. Cambridge: Cambridge University Press.
- Soh, Jerrold Tsin Howe, Lim How Khang, and Ian Ernst Chai. 2019. Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments. In *Proceedings of the Natural Language Processing Workshop 2019*, pp. 67–77.
- Solan, Lawrence M. 2017. Patterns in Language and Law. *International Journal of Language and Law* 6: 46–66.
- Staton, Jeffrey K., and Georg Vanberg. 2008. The Value of Vagueness: Delegation, Defiance, and Judicial Opinions. *American Journal of Political Science* 52(3): 504–519.
- Stone Sweet, Alec, and Thomas L. Brunell. 1998. The European Court and the National Courts: A Statistical Analysis of Preliminary References, 1961–95. *Journal of European Public Policy* 5(1): 66–97.
- Tiller, Emerson H., and Frank B. Cross. 2006. What Is Legal Doctrine? *Northwestern University Law Review* 100(1): 517–533.
- Trappey, Charles V., Amy J.C. Trappey, and Bo-Hung Liu. 2020. Identify Trademark Legal Case Precedents – Using Machine Learning to Enable Semantic Analysis of Judgments. *World Patent Information* 62.
- Vanberg, Georg. 2005. *The Politics of Constitutional Review in Germany*. Cambridge: Cambridge University Press.
- Venkatesh, Ravi Kumar, and K Raghuvver. 2013. Legal Documents Clustering and Summarization Using Hierarchical Latent Dirichlet Allocation. *IAES International Journal of Artificial Intelligence* 2(1): 27–35.
- Vogel, Friedemann, Hanjo Hamann, and Isabelle Gauer. 2018. Computer-Assisted Legal Linguistics: Corpus Analysis as a New Tool for Legal Studies. *Law and Social Inquiry* 43(4): 1340–1363.
- Weiler, Joseph H. H. 1994. A Quiet Revolution: The European Court of Justice and Its Interlocutors. *Comparative Political Studies* 26(4): 510–534.
- Winkels, Radboud, Jelle Ruyter, and Henryk Kroese. 2011. Determining Authority of Dutch Case Law. *Legal Knowledge and Information Systems* 235: 103–112.