What's the talk in Brussels? Leveraging daily news coverage to measure issue attention in the European Union

Article *in* European Union Politics · February 2020

DOI: 10.1177/1465116520902530

CITATIONS
3

READS
36

3 authors:

Michal Ovádek
University of Gothenburg
45 PUBLICATIONS **34** CITATIONS
SEE PROFILE

Nicolas Lampach
KU Leuven
37 PUBLICATIONS **76** CITATIONS
SEE PROFILE

Arthur Dyevre
KU Leuven
67 PUBLICATIONS **202** CITATIONS
SEE PROFILE

Article

European Union Politics
2020, Vol. 21(2) 204–232

DOI: 10.1177/1465116520902530
journals.sagepub.com/home/eup
SAGE

What's the talk in Brussels? Leveraging daily news coverage to measure issue attention in the European Union

Michal Ovádek
Centre for Legal Theory and Empirical Jurisprudence, KU Leuven, Leuven, Belgium

Nicolas Lampach
Centre for Legal Theory and Empirical Jurisprudence, KU Leuven, Leuven, Belgium

Arthur Dyevre
Centre for Legal Theory and Empirical Jurisprudence, KU Leuven, Leuven, Belgium

Abstract
Research on issue attention in the European Union has focused on the prominence of EU integration in domestic politics and media and, at EU level, on the salience of individual issues and legislative files, often in relation to lobbying. Existing EU-level measures of issue saliency, though, are limited in scope and periodicity and tend to reflect the policy priorities of a single institutional actor rather than that of the broader EU elite sphere. We present an alternative measure of issue attention leveraging the quasi-institutional nature of the Agence Europe daily bulletin which provides comprehensive but independent news coverage of EU affairs. We use text-mining techniques, including dynamic topic modelling, in combination with manual classification to map issue prevalence between 1979 and 2018. In addition to reporting validation results, we

Corresponding author:
Michal Ovádek, Centre for Legal Theory and Empirical Jurisprudence, KU Leuven, 45 Tiensestraat, 3000 Leuven, Belgium.
Email: michal.ovadek@kuleuven.be

illustrate how our measure relates to other indicators of EU agenda formation and explain how researchers can make use of our new dataset.

Keywords
Dynamic topic model, issue attention, media, validation

Introduction

What issues receive political attention, or not, is a crucial aspect of politics. Political attention is a scarce and precious good, for which aspiring agenda influencers must compete along with institutionalized agenda-setters. So, since agenda formation affects, either directly or indirectly, virtually all facets of politics, it is not surprising that political scientists have expended considerable efforts on mapping and explaining how policy agendas change (Baumgartner et al., 2006; Baumgartner and Jones, 1993; Princen and Rhinard, 2006), issues and their frames evolve (Carmines and Stimson, 1989; Diez-Medrano, 2004) and how democratic institutions respond to public demands and concerns (Alexandrova et al., 2016; Sorace, 2018).

European Union (EU) studies are no exception to the quest for reliable and comprehensive indicators of issue attention. Much of the literature on agenda formation in the context of EU affairs has concerned itself with the salience of EU integration in domestic politics and media (Adam and Eschner, 2008; Diez-Medrano, 2004; Koopmans and Pfetsch, 2006; Netjes and Binnema, 2007; Peter and De Vreese, 2004; Rauh, 2014; Veen, 2011; Wonka, 2016). Owing to the absence of 'a genuinely supranational public sphere on the European level' (Koopmans, 2007; Risse, 2010: 185), measuring the degree of 'Europeanization' of domestic politics has been the principal preoccupation of this strand of research. At EU level, meanwhile, researchers have concentrated on the salience of individual issues and legislative files, often in connection with lobbying efforts (Beyers et al., 2018; De Bruycker and Beyers, 2015; Dür and Mateo, 2014; Klüver, 2011). As with agenda research in other contexts, the development of a general measure of issue attention has had to grapple with the difficulty of measuring salience in a comprehensive manner (Wlezien, 2005), either due to prohibitive cost of measurement techniques (surveys, expert interviews, human coding) or the lack of sufficiently comprehensive data (Warntjen, 2012). Using European Council and European Commission documents, though, researchers have evolved actor-based indicators of agenda formation. Among these is the European Union Policy Agendas Project (Alexandrova et al., 2014), which represents, to date, the most comprehensive attempt to measure issue attention at EU level. The dataset covers a wide range of policies across four decades and constitutes, without any doubt, a major contribution to the study of agenda-setting in the EU context. Still, it

exhibits important limitations. First, it is exclusively based on the items appearing in European Council conclusions. Yet, what matters to the European Council may not matter for other EU actors, and vice-versa. Researchers should therefore be wary of using it as more than a measure of Council agenda priority. Second, European Council conclusions emphasize high politics where national governments wish to set general guidelines and objectives for the bloc. Yet, for many research questions, such as the influence of lobby groups on particular policies or the comparative policy responsiveness of EU institutions on specific issues, researchers need a measure that captures both high and low politics and is independent of EU institutions. More broadly, what they need is a measure that captures issue attention in the Brussels-bound, EU elite sphere that encompasses Commissioners, Members of European Parliament (MEPs), EU civil servants, diplomats, policy experts, EU contractors, consultants, pressure groups and non-governmental organizations (NGOs), as well as national government representatives.

In this paper, we undertake to construct such a measure by leveraging text data from the Agence Europe bulletin (AEB). AEB, we argue, is a 'quasi-institutional' news outlet which caters to EU policy wonks, civil servants and lobbyists. Published daily, it covers all policy areas within the remit of EU competences along with broader foreign policy and business news. The broad scope of AEB ensures that it captures low as well as high politics. Analysing the entire universe of English-language AEB from 1979 to 2018, we apply text-mining techniques, including dynamic topic modelling, to construct a detailed classification of the bulletins' contents. The resulting dataset provides a measure of the proportion of 75 machine-generated topics organized into 19 manually defined meta-categories.[1] We validate our indicator using a random sample of human-coded bulletins. Besides showing how our measure of issue attention relates to other measures of EU agenda formation, we provide illustrations of how particular agenda items track international developments and critical junctures in the European integration process.

Issue attention in the EU

Issue attention, agenda formation or issue saliency – we use these terms interchangeably[2] – can be measured in a variety of ways. A useful distinction is between generic and actor-centred measures (Beyers et al., 2018). Actor-centred indicators approach issue saliency from the viewpoint of aspiring agenda influencers or institutionalized agenda setters. This is the approach adopted by researchers who have constructed measures of issue prevalence from European Council, Council of the European Union and European Commission files (Alexandrova, 2017; Alexandrova et al., 2016, 2014; Carammia et al., 2016; Häge, 2016; Osnabrügge, 2015), European Parliament speeches (Greene and Cross, 2017) or, at the domestic level, from debates in the legislature (Rauh, 2014; Wonka, 2016) or items in party manifestos (Veen, 2011). This approach has the advantage of being both analytically and empirically tractable, in addition to having relevance for various strands

of research on EU politics. Each EU institution produces its own line of policy documents and communications which scholars have been able to dissect and interpret as issue attention (Alexandrova et al., 2014; Greene and Cross, 2017; Häge, 2016; Osnabrügge, 2015). The variance in issue attention among different institutions is itself an object of research (Alexandrova, 2017). The same holds for electoral manifestos and legislative debates (Rauh, 2014; Veen, 2011; Wonka, 2016). In contrast to actor-centred measures, generic measures seek to capture issue attention in the broader mediatized public sphere, assuming that this sphere is imperfectly controlled and transcends the priorities of individual agenda-setters. 'Europeanization' studies that aim to assess the attention paid to the EU in domestic media (Adam and Eschner, 2008; Diez-Medrano, 2004; Koopmans and Pfetsch, 2006; Peter and De Vreese, 2004) typically follow this line of inquiry.

A comparison of the domestic and EU-level literatures on issue attention reveals an interesting disparity. Whereas research on issue attention at the domestic level features both generic and actor-centred approaches, the EU-level literature has considered only the latter. The underlying reason for this asymmetry may seem obvious. The EU lacks an autonomous public sphere (Koopmans, 2007; Risse, 2010: 185). Therefore, actor-based indicators may seem to adequately capture the items that define the agenda in Brussels. The EU Policy Agendas Project is predicated on the notion that the European Council is the most powerful agenda-setter in the EU context (Alexandrova et al., 2014). Thus, what the EU pays attention to should somehow find its way into European Council conclusions, or alternatively into Commission documents or European Parliament speeches.

We believe, however, that actor-centred measures of agenda formation miss important aspects of the EU-level policy debate. While the EU lacks an autonomous public sphere, Commissioners, MEPs, EU civil servants, diplomats, policy experts, EU contractors, consultants, pressure groups, NGOs, and national government representatives operating in the Brussels-bound 'EU bubble'(Busby, 2013) form what we may call an 'elite sphere'. What participants in this sphere debate and discuss goes beyond what a particular EU institution formally puts on its agenda. EU institutions are typically self-focused when articulating their policy priorities, as research comparing the Commission and the European Council has shown (Alexandrova, 2017). Moreover, what they choose to prioritize partly reflects their position in the policy-making process. European Council conclusions set broad goals and articulate general guidelines, which the Commission, the European Parliament and the Council of the European Union then work to implement (Alexandrova et al., 2012). By contrast, while issues debated in the EU elite sphere should be expected to include the items of high politics placed on the European Council's agenda in the ongoing policy cycle, they should also encompass many low-politics topics and sub-topics.

We seek a generic, EU-level measure of issue attention to cover all these items while minimizing the institutional biases associated with actor-centred measures. In the following sections, we construct and validate such a measure.

The Agence Europe bulletin

We propose to construct a generic indicator of issue attention from the AEB. Founded in 1953, originally published in French and from 1979 onward also in English, the AEB specializes in news relating to EU activities and institutions. The AEB sees itself as the leading source of information on European integration: 'For 66 years now, Agence Europe has been widely considered THE source (some might say the "Bible") of information on European economic and political integration.'[3]

Stressing its prominence in the EU elite sphere, some scholars have described the AEB as a 'quasi-institution' (Bastin, 2002; Fougier, 2010; Marthoz, 2008). The bulletin is an elite-sphere publication both *for* and *about* EU decision makers. AEB journalists and correspondents are policy specialists, while EU civil servants, Brussels-bound diplomats and EU policy experts constitute the bulk of AEB's readership (Fougier, 2010). This specialized policy orientation is reflected in the breadth and dry descriptiveness of the bulletins. Aside from institutional and policy developments – which capture change in policy agendas (Peters, 1994; Princen, 2007, 2011; Princen and Rhinard, 2006) – bulletins also carry industry and economic news, in particular those tied to EU affairs. While the AEB along with its leading journalists are associated with a soft-federalist editorial line and EU officials are often their main source of information (Fougier, 2010), AEB is financially and organizationally independent of EU institutions. In early decades, the AEB sometimes re-printed institutional reports which could skew an individual issue towards the frames favoured by its institutional author. Nonetheless, the vast majority of bulletins consist of news items reporting on EU developments without using emotionally charged language, and this largely descriptive tone is consistent over time. We report a minor sentiment analysis exercise in the Online appendix. Opinion pieces (including editorials) represent a very small fraction of the contents of the bulletins.

While political and media attention are, arguably, co-constitutive (Boydstun, 2013), we believe that the AEB provides a comprehensive and relatively objective measure of issue attention. This presumption rests on the small and specialized nature of the AEB's readership. First, the AEB reports on issues which are not sufficiently newsworthy for mass media but of interests to EU civil servants and policy experts. The resulting breadth of coverage makes the AEB suitable for serving as a basis for a general survey of topic attention. Second, its small, specialized audience makes the AEB unable to shape public opinion. Obviously, the AEB is not a publication for the masses. The AEB's journalistic style is descriptive rather than investigative (Fougier, 2010). So, political actors at the EU level are relatively safe to ignore its reporting. Even the increased politicization of EU affairs and the emergence of rival outlets, such as Politico, EURACTIV and EUobserver have made little difference to AEB's descriptive style of reporting (Fougier, 2010). In consequence, it seems reasonable to assume that AEB's comprehensive coverage of EU affairs is comparatively less distorted by the pressures

of electoral and political cycles and the desire to influence voters than transnational mass outlets – such as the Financial Times, the Guardian and the Economist – or, for that matter, national newspapers and broadcasters.

While the EU's elite sphere has seen the emergence of other news outlets – most notably EUobserver, EURACTIV and Politico – and their degree of overlap with the AEB is a question that future research might deem worth investigating, these outlets do not offer the temporal coverage of the AEB. The English version of the AEB employed in the construction of our measure of issue attention goes back to 1979. By comparison, the old rivals of the AEB, EURACTIV and EUobserver, were launched in, respectively, 1999 and 2000.

Data and methodology

This section explains how digitized versions of the bulletins were obtained; how these texts were pre-processed to facilitate computer-aided analysis and how we combine probabilistic topic modelling and human classification to construct our indicator.

Data: AEB, 1979–2018

Our issue attention measure is derived from the entire universe of bulletins (N = 9546) published in English during the period 1979–2018. By relying on a single type of publication from a single media source, we strive to minimize unwanted heterogeneity and maximize temporal consistency in the underlying text data. Between 1979 and 2000, the textual data are available in scanned form courtesy of a digitization project conducted by the European University Institute Library. From 2000 onward, the data were sourced in digital form from the website of Agence Europe (https://agenceurope.eu).

The fact that the data prior to 2000 come from scans of printed publications, as opposed to HTML websites, could introduce unwanted distortion to our measure, as our topic modelling could pick up on technical differences between the two sets of documents. While we acknowledge that the pre-2000 data do not have the perfect quality of the electronically published documents, we strived to apply the best available optical character recognition (OCR) and image conversion techniques to minimize data loss. We point out that 19 bulletins from the early period are missing entirely. Depicting the number of words of each document, Figure 1 suggests that the data retrieval process has not resulted in a systematically biased corpus.

Whereas document length is relatively consistent around the period corresponding to the transition from the scanned to the HTML format (1999/2000), we can see a gradual decline since 2006. We speculate that this reflects editorial decisions driven by readership preferences as a consequence of the changing competitive landscape in the EU-level news market. That said, document length is on the whole remarkably stable, with outliers few and far between. Sentence and

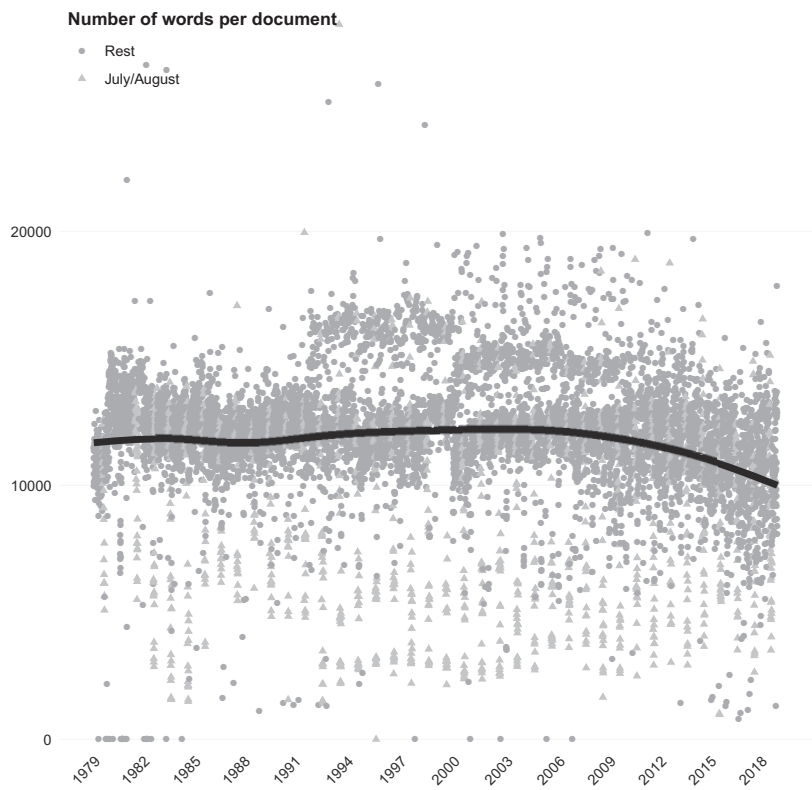

Figure 1. Number of words per document in the raw corpus.

character-level measures produce results very similar to the word-level measure. We report the latter because words constitute the unit of analysis in the topic model. Systematic variation can be observed when comparing summer issues (in particular August) with the rest of the year, which mirrors the lighter workload of EU institutions and the less intense news cycle in this period (see also the Online appendix).

Pre-processing

As is customary for text-mining applications, our pre-processing steps involve removing white-spaces, punctuation, numbers and stop-words together with stemming while converting all characters to lower case. The goal of pre-processing is to strike the right balance between reducing the inherent complexity of textual data and preventing the loss of relevant information (Denny and Spirling, 2018; Grimmer, 2010; Lucas et al., 2015). More specifically, Figure 2 illustrates our pre-processing strategy, which breaks down into three main steps: data collection, cleaning and reduction. As half of our data consist of scanned documents, we

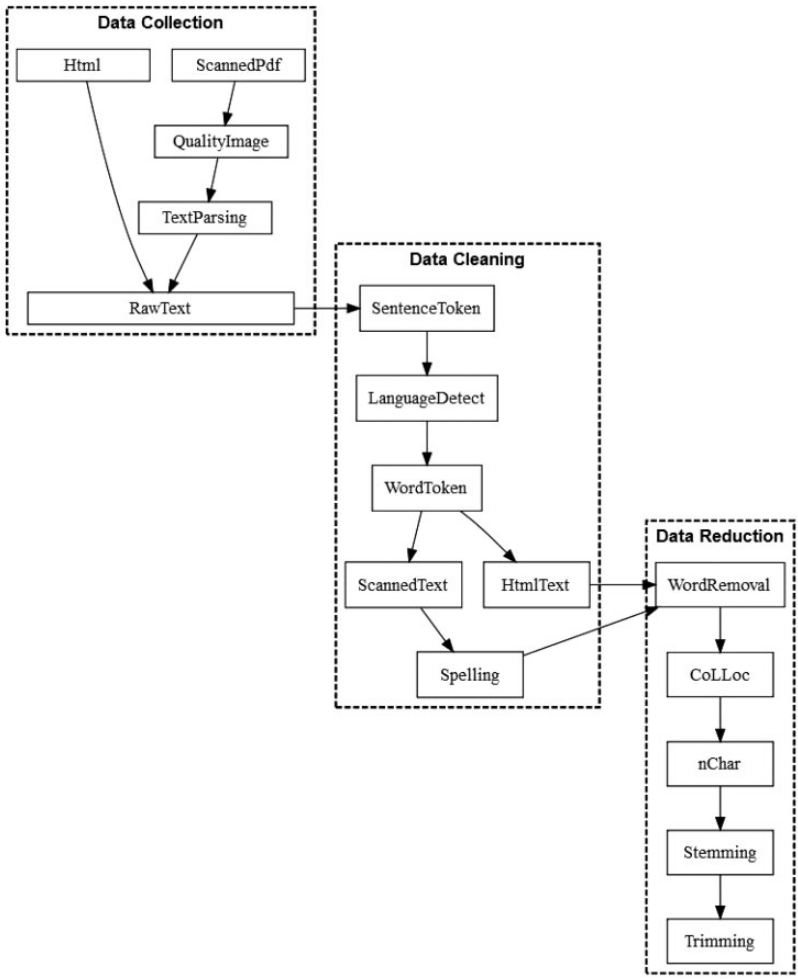

Figure 2. Data pre-processing steps.

attempted to minimize unwanted distortions in the data collection phase by applying a state-of-the-art OCR technique which improves the likelihood of successful text extraction (Smith, 2007). Using recognized words as training data, the OCR algorithm detects the individual component features from which a character is made.

The second stage of pre-processing involves the removal of undesirable document-specific (quasi-)words. A substantial number of bulletins feature sentences or paragraphs written in a language other than English (often French). In order to avoid misclassification problems in the topic model, we tokenize all documents at the sentence level and apply a language detection tool to identify non-English

sentences, which we then remove from the corpus. Owing to frequent misspelling, typos and similar errors resulting from either incorrect character recognition or the original texts, we employ a dictionary-based spell-checker after tokenizing the sentences at the word level (Ooms, 2017). Erroneous terms are replaced by most likely suggested words or dropped if no suggestion is available, as is the case for garbage tokens. In the last step, we apply four common pre-processing techniques aiming at reducing the number of features in the final document-term matrix: removal of uninformative words and terms shorter than three or longer than 15 characters, reducing words to their stem (stemming), and dropping words appearing in more than 95% and fewer than 5% of all documents.[4] During this stage, we also convert frequent – appearing at least 1000 times in the corpus – two-word collocations (bi-grams) into single expressions (for example, 'cohesion' followed by 'policy' becomes 'cohesion_polici').

While the present paper emphasizes our own indicator, which, we are confident, should meet the needs of many researchers in search of generic EU-level measure of issue attention, the textual corpus enables the construction of alternative indicators, a different set of techniques (e.g. supervised classification, vector embeddings) and/or choosing distinct units of time (week, month) or level of generality. All the data we collected and generated – whether the raw and pre-processed corpus or the final EUSSUE indicator – are available on Euthority.eu for researchers to use.

Dynamic topic modelling and human classification

With individual bulletins spanning around 13,000 words on average, coding our corpus manually would require an impossibly large amount of resources. For that reason, we rely on automated content analysis techniques to estimate the latent topical structure of our corpus (Catalinac, 2016; Dybowski and Adämmer, 2018; Grimmer, 2010; Grimmer and Stewart, 2013; Kim, 2018; Quinn et al., 2010; Rothschild et al., 2019). Developments in machine learning and natural language processing have led to the emergence of probabilistic topic modelling as a powerful technique to classify large collections of documents (Blei et al., 2003; Blei and Lafferty, 2006). Originally developed by (Blei et al., 2003), latent Dirichlet allocation (LDA) remains the most popular topic modelling approach. Yet, increasingly sophisticated or specialized topic modelling algorithms have appeared in recent years ranging from non-negative matrix factorization to hierarchical topic models (Blei, 2012; Greene and Cross, 2017; Grimmer, 2010; Rosen-Zvi et al., 2004). All topic models, though, share the same basic assumptions. While reducing texts to bags of words, they model topics as latent probability distributions over words and documents as latent probability distributions over topics.

Although LDA – the plain vanilla approach to topic modelling – is a powerful algorithm for analysing large corpora, it is not well suited for dynamic corpora, where topics may change over time. Dynamic topic modelling addresses this problem by extending the idea of LDA to allow topic representation to evolve over fixed time intervals (Blei and Lafferty, 2006). The multi-purpose structural topic model

developed by Roberts et al. (2014) and implemented in the *stm* package for R allows the estimation of a topic model where each topic is represented not by a single distribution over words but, instead, by a sequence of distributions over the time intervals. Building on the Correlated Topic Model (Blei et al., 2007), the *stm* implementation assumes a logistic normal generalized linear model instead of a latent Dirichlet process for topic proportion (Roberts et al., 2014). Here we choose the year as time interval, as it is a common unit of time to track agenda formation dynamics.

Formally, our time interval enters the document-generating process as a covariate interacting with topic prevalence

$$\theta_{1:D}|t_{1:D}\gamma, \Sigma \sim LogisticNormal(\mu = t_{1:D}\gamma, \Sigma) \tag{1}$$

where t_d is the year in which document d was published; γ is a $p \times (K-1)$ matrix of coefficients for topic proportion and Σ is a $(K-1) \times (K-1)$ covariance matrix. A topic z is drawn from the document-specific distribution, conditional on that topic k, a word is chosen from the multinomial distribution over words denoted by $\beta_{d,k}$

$$\beta_{d,k} \quad \infty \quad exp(m + \kappa_k) \tag{2}$$

where m represents the baseline word frequencies and κ_k the topic deviation per topic k. Conditional on the topic selected and token-level distribution $z_{d,n}$ where $n \in [1, \ldots, N-d]$ for each word in the document, the observed word $w_{d,n}$ is drawn from a multinomial distribution

$$w_{d,n}|z_{d,n}, \beta_{d,k=z_{d,k}} \quad \sim \quad Multinomial\left(\beta_{d,k=z_{d,k}}\right) \tag{3}$$

As implemented in the *stm* package, the posterior distribution for this dynamic topic model is computed via variational expectation maximization (Roberts et al., 2019). The number of topics k in a topic model is set by the researcher and determines the dimension of the topic space. A larger k provides a more detailed picture of the corpus, while a smaller k results in a simpler, more general picture. A variety of metrics, such as harmonic mean, perplexity, log-likelihood, semantic exclusivity and coherence, have been proposed to evaluate the quality of the model and to choose the number of topics (Chang et al., 2009; Mimno et al., 2011; Wallach et al., 2009). However, there is no 'right' k, and the validity of these formal criteria is not well established; perplexity, for one, has been shown to be inversely correlated with human interpretability (Chang et al., 2009). For our main analysis, we ignore these formal metrics. Instead, we set the number of topics on the basis of two criteria. The first is interpretability. There is little point in using topic modelling if the resulting topics, or at least a large number of them, eschew human interpretation. The second criterion is motivated by the aim of our classification exercise: we want a sufficiently large number of

topics to provide a relatively fine-grained picture of agenda variation. After iterating several topic models, we settled on $k = 88$ as providing a sufficiently large number of interpretable topics. We found that a higher k produces few additional meaningful topics, whereas a lower k, though producing interpretable topics, offers a less detailed picture.

Using unsupervised document classification to construct a measure of issue attention has advantages as well as drawbacks. Topic modelling is an inductive, data-driven approach. Whereas supervised classification and manual coding schemes presuppose that the researcher knows the relevant categorizations (Alexandrova et al., 2014), topic modelling lets the data speak first. The topics do not come from the researcher's *a priori* definition but emerge from the data. Thus, a topic model may reveal issues the researcher had not thought of. However, to the extent that one has *a priori* knowledge of the items that a measure of issue salience should include or wishes to relate the results to other, human-coded indicators, the topic model may fail to capture the categories of interest to the researcher. Our approach mitigates this problem by adding a layer of human coding to the topic model. We organize the machine-generated topics into more general themes. For this, we follow the coding scheme of the EU Policy Agendas Project (Alexandrova et al., 2014), from which we only deviate with respect to the European Parliament, the Eurozone and EU Treaties.[5] This makes our measure easier to relate to existing EU-level indicators.

EUSSUE: A generic, EU-level measure of issue attention

Similarly to Quinn et al. (2010), we aim to construct a classification scheme with items that sustain attention over time. We take this to imply that topics must transcend mere events. Nevertheless, as will be seen below, our classification method identifies considerable temporal variation in topic proportion.

The most important output of our dynamic topic model is the topic proportion per document θ. We interpret this value as a measure of issue attention in the k-dimensional topic space. For topic i at time t, a larger θ_{it} indicates greater topic prevalence. Ignoring the temporal dimension for the moment, Table 1 shows an example of the topic proportions in our model (for all 88 topics, see the Online appendix). The table includes also the words most distinctive of each topic, as determined by the word-level parameter β.

In the next step, we proceeded to label all topics and identify those which are not useful for our measurement objectives. The topics were labelled independently by two of the authors with the third one reconciling discrepancies. We use the average cosine similarity to measure inter-coder reliability for topic labelling yielding 0.761. The correlation of assigning the themes of the Policy Agenda project is 0.878. The labels were assigned on the basis of the topic's most distinctive words (based on β value), most distinctive documents (based on θ value) and varying proportion over time. As is common in unsupervised document classification, some topics proved, despite the care we put in selecting the model, impossible to interpret. While

Table 1. An example of topic proportions and top words for 88-topic model.

Topic number	Topic proportion (%)	Words (β)	Topic label
29	2.175	court justic law reason legisl wast infring water fail judg	european_court_of_justice
39	2.157	treati articl provis power common legal legisl charter compet draft	treaty_reform
80	2.140	law power histori citizen research social cultur legal describ integr	garbage: political_principles
77	1.878	summit strategi conclus commit prioriti reform framework integr full invit	summits_organisation
40	1.827	growth deficit rate forecast fall rise economi inflat budgetari figur	economic_outlook_european_economies
5	1.812	brexit vote sourc draft articl law compromis citizen invest mogherini	brexit
17	1.784	price agricultur produc wine farmer milk sugar farm cereal cap	common_agricultural_policy
7	1.782	maastricht treati summit round debat uruguay republ danish yugoslavia packag	maastricht treaty
48	1.757	vote amend resolut plenari rapporteur debat socialist abstent reject green	EP_debates
9	1.695	oil price import unit dollar energi convent neuron jenkin start	oil_price
68	1.689	price export commod import oil deleg start produc energi get	commodity_prices_2
22	1.568	organ summit treati network debat prepar single_curr social spanish access	procedural_summits

(continued)

Table 1. Continued.

Topic number	Topic proportion (%)	Words (β)	Topic label
36	1.564	budget fund budgetari expenditur spend resourc payment commit amount appropri	EU_budget
61	1.553	turnov debat summit american nice aim stake usd enlarg british	garbage: summit
1	1.547	price import export rate demand agricultur united_st dollar american firm	commodity_prices
12	1.526	bank subsidiari share capit firm invest american acquir british turnov	business_activity
21	1.514	technolog enlarg dollar deleg head united_st treati summit lead esprit	garbage: trade_dollar_mid_1980s
24	1.458	convent co-oper enlarg candidate_countri debat summit reform access organis united_st	convention_future of_europe_and_enlargement
27	1.447	resourc special deleg financ reduct athen stuttgart summit agricultur german	community_budget_summit
20	1.422	constitut lisbon_strategi debat research citizen solana financial_perspect get communic prioriti	constitutional_treaty

Note: All 88 topics can be found in the Online appendix.

'garbage' topics may take up a significant proportion of the topic space (Quinn et al., 2010), in our case, this amounts to a tolerable 13.2% of the total. We discarded these topics, thereby bringing the number of topics down to 75.[6]

As can be readily seen in the full table of topics (reported in the Online appendix), several topics relate to a similar overarching theme such as agriculture or the euro. As explained above, we organized the topics into 19 themes following a coding scheme largely inspired by the EU Policy Agendas Project.

This task was completed following the same procedure as topic labelling, with two researchers annotating topics independently and a third one reconciling the discrepancies.[7] We call the resulting generic, EU-level measure of issue attention 'EUSSUE'.

Table 2. An example of EUSSUE topics clustered into themes.

Theme labels	Theme proportion (%)	Topic number	Topic labels	Topic proportion (%)
International affairs and foreign aid	17.109	22	procedural_summits	1.743
		44	humanitarian_crises	1.576
		81	yugoslavian_and_gulf_crisis	1.387
		38	german_reunification	1.303
		25	israel_palestine_conflict	1.268
		55	EU_foreign_policy_banking_crisis	1.195
		10	russia_ukraine_crisis_incl_gas	1.05
		65	south_africa_apartheid	1.000
		16	turkey_accession_to_EU	0.975
		66	EU mediterranean cooperation incl. MENA	0.892
		31	cooperation_with_ morrocco_and_tunisia	0.891
		43	EU_latin_america_cooperation	0.865
		83	balkans	0.843
		33	Cotonou (ACP) agreement	0.776
		79	syrian_crisis	0.723
		63	EU_ASEAN_cooperation	0.622
EU governance	12.328	29	european_court_of_justice	2.351
		77	summits_organization	2.053
		5	brexit	1.987
		36	EU_budget	1.74
		24	convention_future of_ europe_and_enlargement	1.633
		27	community_budget_summit	1.622
		74	eastern_member_states	0.942

Note: The full list can be found in the Online appendix.

Table 2 shows an example of the labelled topics and two themes with the attached (overall) proportions (see table in the Online appendix for a complete overview).

Figure 3 illustrates how issue attention varies over time using the theme level of analysis. Among other things, the plot suggests a decline in the salience of foreign trade but a surge in the salience of immigration around 2015, while agriculture exhibits remarkable stability.[8]

Validation

Because our classification scheme is, for the largest part, machine-generated, it requires validation (Grimmer and Stewart, 2013). Both the nature of the documents and the relative complexity of our classification scheme presented a

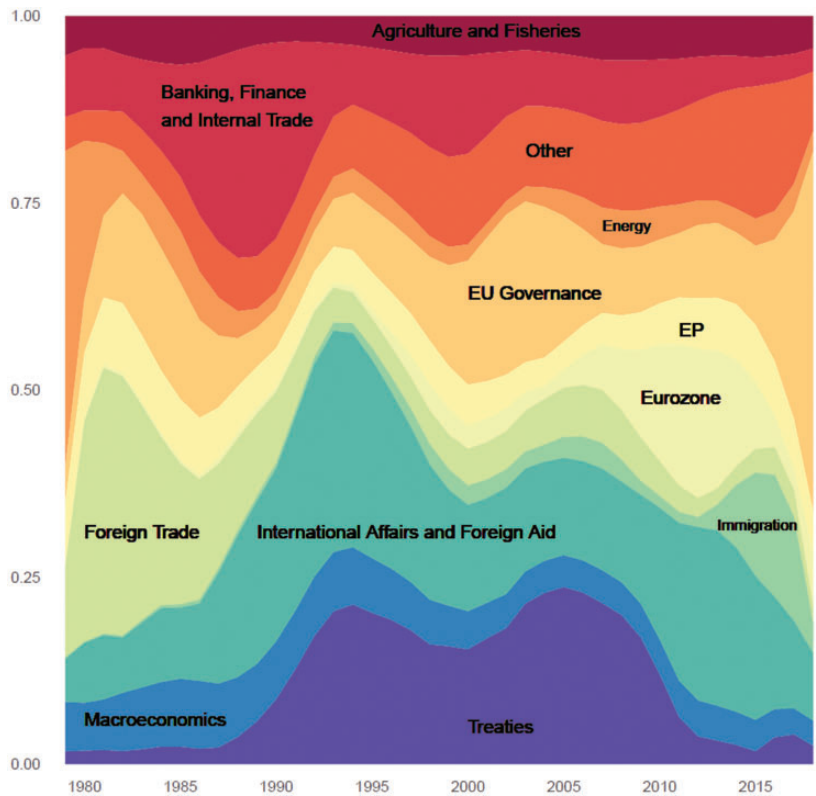

Figure 3. EUSSUE theme-proportion over time.
Note: The theme 'Other' comprises the seven smallest themes in terms of proportion for the sake of visualization. With the exception of themes 'European Parliament', 'Eurozone' and 'EU Treaties', thematic clustering is aligned with the codebook of the Agendas Project (Alexandrova et al., 2014).

validation challenge. The bulletins are lengthy and varied documents, issued in changing economic and geopolitical contexts. Also, because our unsupervised content analysis method models documents as mixtures of topics, a document is assumed to feature all topics, although in proportion varying from large to infinitesimally small. As a consequence, we had reasons to expect that human coders would, at times, lack the historical knowledge or expertise to situate the events described in the bulletins and relate them to the correct topic in our classification scheme. Still, we opted for a simple, conservative validation procedure. We took a random sample of bulletins and then independently asked untrained coders to determine which of the 75 EUSSUE topics they believed to be instantiated in these documents. Each coder had to first read the document and then tick a box in a form listing the 75 topics. The exercise spanned 20 unique documents, for a total of 49 coder-document observations.

Inter-coder reliability was 56%, which is reasonably high for a validation task of such complexity (Hruschka et al., 2004). It bears emphasis, in particular, that, owing to the high number of topics, the probability that coders would agree simply by chance was very small. Cohen's κ, which this probability into account, averages 0.53 across pairwise coder comparisons.

To evaluate how our computer-based measure of topic proportion compares to human coding, we related the documents' θ values from the topic model to the coders' determination regarding the presence of the corresponding topics. Whether we regress human coding on θ (using logistic regression) or θ on the coders' choice (using OLS regression), we find a positive and strongly significant association, as shown in Table 3.

Table 4 further reports the degree of agreement between human coders and our 19 theme-level categories. Correlation coefficients are mostly positive and often significant. Moreover, even if we group the data at the theme level, we observe consistently high correlation across all themes (reported in the Online appendix). These results, we believe, support the validity of our measurement method.

Relation to European Council agenda items

While manual topic validation demonstrates that our method is effective at extracting agenda items from the AEB documents, comparing EUSSUE with an existing measure of EU agenda formation is useful to understand how a generic indicator differs from actor-centred ones. The European Union Policy Agendas Project offers the only quantitative measure comparable to the EUSSUE measure proposed here (Alexandrova et al., 2014). As highlighted above, EUSSUE and the Agendas Project differ in two important respects. First, one is based on European Council conclusions, which are essentially outcomes of political negotiations, whereas the other is built on journalistic reporting. Second, the two measures also diverge in the way they capture the time dimension. AEB is published on average five times a week, while European Council summits have traditionally had biannual recurrence. In sum, the two indicators are designed to measure different things. So, while a degree of overlap

Table 3. Relationship between EUSSUE θ and manual coding of a random sample of Agence Europe bulletins.

	Dependent variable**			
	Coders' choice	θ	Coders' choice	θ
Constant	0.259*** (0.053)	0.010** (0.005)	0.221*** (0.007)	0.011*** (0.0005)
θ	1.856***		1.856***	
	(0.221)		(0.219)	
Coder's choice		0.012*** (0.002)		0.011*** (0.002)
FE author	Yes	Yes	No	No
FE bulletin	Yes	Yes	No	No
Observations	3825	3825	3825	3825
R^2		0.021		0.021
Adjusted R^2		0.014		0.021
McFadden pseudo R^2	0.038		0.018	
Log Likelihood	–2123.067		–2165.995	
Akaike Inf. Crit.	4308.134		**4335.991	
Residual std. error		0.033 (df = 3794)		0.033 (df = 3823)
F statistic		2.758*** (df = 30; 3794)		81.532*** (df = 1; 3823)

Note: *p < 0.1; **p < 0.05; ***p < 0.01. Robust standard errors are given in parentheses.

Table 4. Correlation between EUSSUE theme proportions and human coding at bulletin level.

	Pearson	Spearman	Number of coders
BNo_04921	0.849*** (0.849; 0.849)	0.738*** (0.738; 0.738)	5
BNo_10929	0.837*** (0.837; 0.837)	0.586*** (0.153; 1.019)	5
BNo_11656	0.641*** (0.218; 1.064)	0.338 (−0.13; 0.806)	5
BNo_02687	0.562** (0.123; 1.001)	0.441* (−0.018; 0.900)	2
BNo_02720	0.501** (0.051; 0.951)	0.820*** (0.820; 0.820)	2
BNo_03257	0.546** (0.101; 0.991)	0.818*** (0.818; 0.818)	2
BNo_04900	0.815*** (0.815; 0.815)	0.416* (−0.046; 0.878)	2
BNo_05078	0.592*** (0.155; 1.029)	0.468** (0.014; 0.922)	2
BNo_05151	0.577*** (0.138; 1.016)	0.625*** (0.200; 1.050)	2
BNo_05231	0.641*** (0.218; 1.064)	0.320 (−0.148; 0.788)	2
BNo_05361	0.452* (−0.005; 0.909)	0.273 (−0.198; 0.744)	2
BNo_05541	0.664*** (0.244; 1.084)	0.576*** (0.138; 1.014)	2
BNo_06710	0.625*** (0.200; 1.050)	0.690*** (0.281; 1.099)	2
BNo_07664	0.643*** (0.219; 1.067)	0.362 (−0.104; 0.828)	2
BNo_09499	0.371 (−0.094; 0.836)	0.609*** (0.175; 1.043)	2
BNo_10072	0.297 (−0.174; 0.768)	0.343 (−0.124; 0.810)	2
BNo_10185	0.282 (−0.189; 0.753)	0.403 (−0.059; 0.865)	2
BNo_10331	0.180 (−0.289; 0.649)	0.125 (−0.338; 0.588)	2
BNo_10858	0.330 (−0.137; 0.797)	0.448* (−0.01; 0.906)	2
BNo_11713	0.328 (−0.14; 0.796)	0.187 (−0.283; 0.657)	2

Note: ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$; 95% confidence intervals are given in parentheses. A random sample of Agence Europe bulletins has been selected for the validation.

is to be expected – European Council conclusions are normally reported in the AEB – they should also exhibit a fair amount of divergence.

Generally speaking, our measure attributes greater saliency to items relating to EU institutions (notably European Parliament debates), the Eurozone and EU Treaties than the European Council conclusions coded in the Agendas Project. This is the reason why we created specific themes for these items, thereby deviating from the meta-classification employed in the Agendas Project. For the purpose of comparing the EUSSUE with the Agendas Project, we collapse the Eurozone topic under 'macroeconomics' and European Parliament and EU Treaties under 'EU governance'. For each theme, we computed the annual theme proportion for the period 1979–2014 shared by both indicators.[9] Figure 4 shows the correlations between the two measures with data aggregated by theme and year. Because the data are a mixture of normal and skewed distributions, we report three correlation methods. Coefficients are $\bar{r} = 0.28; \bar{\rho} = 0.24; \bar{\tau} = 0.17$ for, respectively, Pearson, Spearman and Kendall correlation.

Given the underlying differences in measurement design and approach, it should not come as a surprise that the correlations oscillate, for the most part, between low and medium. It is also understandable that rank correlation is overall

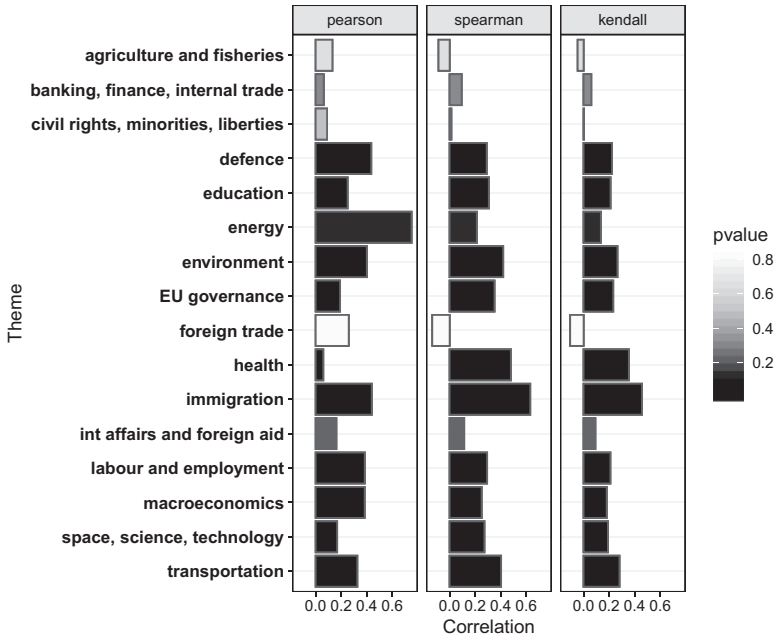

Figure 4. Pearson's, Spearman's and Kendall's correlation between paired theme-year observations in the EUSSUE and Agenda Project datasets.

lower than linear correlation, since issue attention frequently surges considerably in response to crises with the scale of such a phenomenon co-occurring in both measures captured to a larger extent by a linear function. The greatest difference between linear and rank correlation is found for the theme 'energy' ($r = 0.75; \rho = 0.18; \tau = 0.10$). A visual inspection of the data (see Figure 5) demonstrates that both measures share important similarities, in particular for the period of the 1979 oil crisis and in the mid-2000s. Themes which consist of merely one topic – such as health, civil rights, space and science – in the EUSSUE measure tend to be weakly correlated with the corresponding items in the Agendas Projects. Also, whereas agriculture seems to enjoy sustained saliency in AEB, it is much less present in European Council conclusions, which may reflect the fact that this policy has been, in large measure, delegated to the European Commission (Alexandrova, 2017). Overall, these results are consistent with our expectation that while partly overlapping, EUSSUE and the Agendas Project represent distinct measures of issue attention.

Events and temporal patterns

Finally, we relate the EUSSUE to a selected set of exogenous events. While not strictly equivalent to a validation exercise, this helps shore up our assumption that

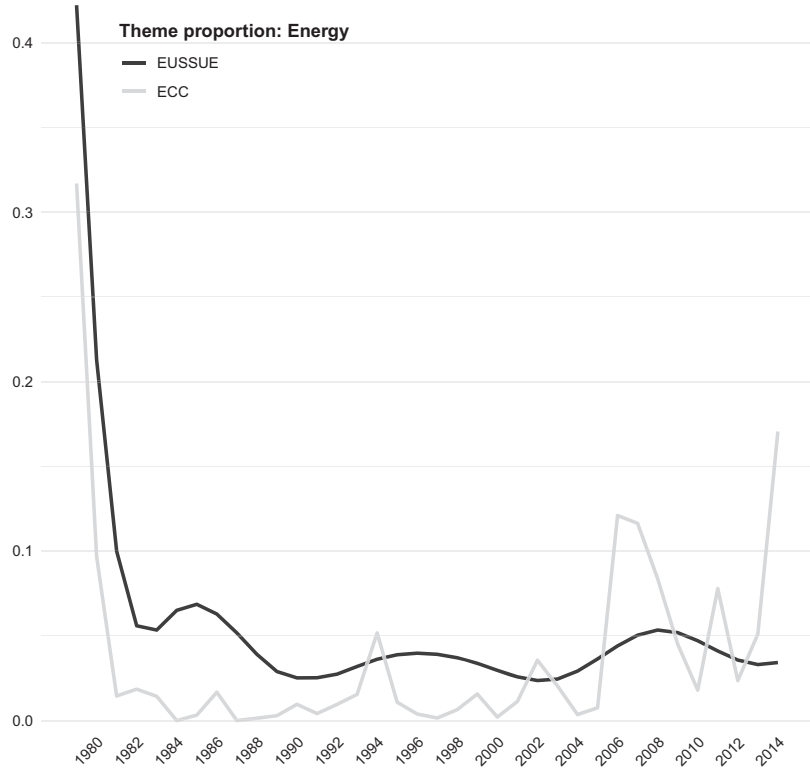

Figure 5. Comparison of EUSSUE and Agendas Project data on attention paid to theme 'energy' between 1979 and 2014.

AEB provides a generic and relatively neutral measure of issue attention. If this assumption is correct, then we should expect topic and theme proportion to correlate with major events, such as armed conflicts and economic crises, where these topics are central. Simultaneously, this exercise helps illustrate how our indicator can be used in research on EU affairs. As a first illustration, we consider the topic 'CO2_emissions'. At the most basic level, we should expect this topic to reflect the rising political and social salience of climate questions since the 1980s and, more recently, the intense diplomatic efforts to bring countries together to cap and reduce greenhouse gases globally. Because the EU has been a key player in this international political effort, we expect the EUSSUE to reflect this. The sizeable literature on the EU's role in the international climate change regime highlights the 2009 Copenhagen summit as a salient juncture, exposing the failure of European leadership (Bäckstrand and Elgström, 2013; Dimitrov, 2010; Groen and Niemann, 2013; Haug and Berkhout, 2010; Oberthür, 2011).

In Figure 6, we can see that the 2009 Copenhagen summit coincides with one of the peaks of the 'CO2_emissions' EUSSUE topic. The θ value for this topic,

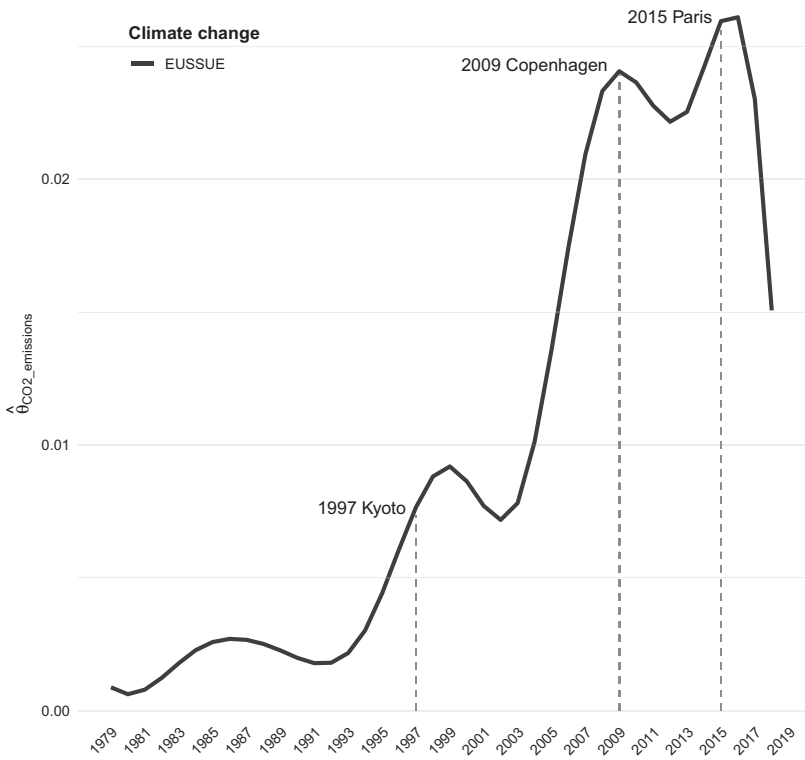

Figure 6. EUSSUE topic 'CO2_emissions' and notable UN climate change summits.

though, reaches its peak in 2016, which corresponds to the signature of the Paris Agreement. Our measure therefore behaves consistently with basic expectations about climate change issue attention in the EU: we observe decade-on-decade increases while peaks correspond to salient junctures.

The impact of climate change summits on issue attention can also be analyzed at a more granular level using days rather than years as unit of time. Summits usually take place at the tail-end of calendar years and might otherwise go undetected when issue attention is aggregated at the year or semester level. Figure 7 reports the document-level θ of the 'CO2_emissions' topic around the time of the 2009 Copenhagen summit.[10]

Figure 7 illustrates how the occurrence of significant events may affect issue coverage outside the strict window when they take place. This makes intuitive sense: in the case of a planned event such as the Copenhagen climate summit, actors prepare and can signal positions in the build-up to the event. The outcome of the event – failure of climate diplomacy and EU leadership – is subsequently discussed in its aftermath. As a result, the period around the Copenhagen summit has a higher $\bar{\theta}$ than the average for 2009, 2010 or overall.

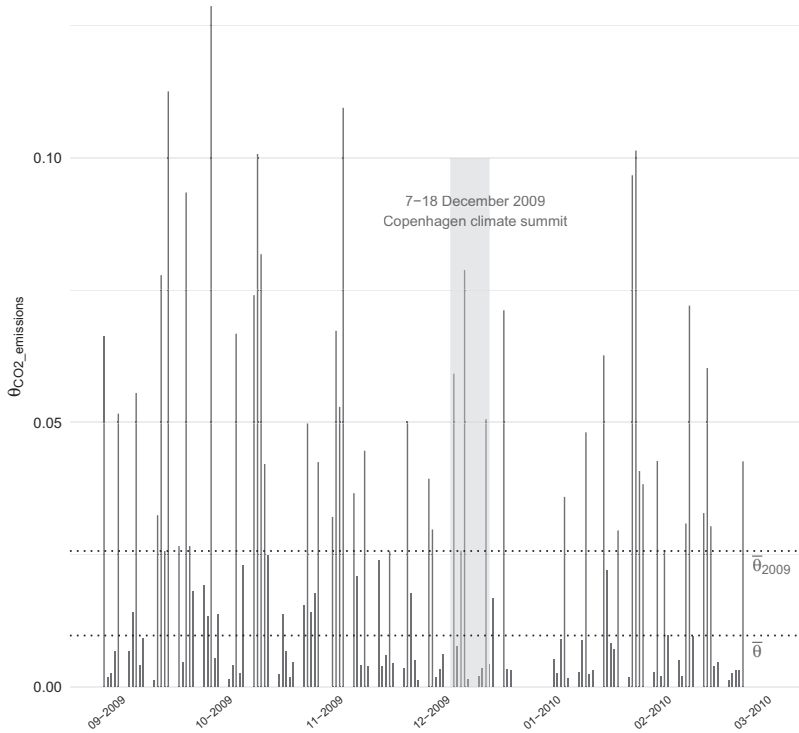

Figure 7. Document-level θ of EUSSUE topic 'CO2_emissions' around the time of the 2009 Copenhagen climate change conference.
Note: The time-series window spans 1 September 2009–1 March 2010.

Events and crises whose intensity can be credibly tracked using continuous indicators instead of individual time-moments provide another illustration of how AEB items respond to these events. Assuming again (imperfect) parallelism between issue attention and real-world developments intuitively recognized as economically or politically salient, we first use Greek bond yields as a proxy for the intensity of the Eurozone crisis. Bond yields capture a combination of risk factors which are sensitive to financial and sovereign debt conditions in a country (Arghyrou and Kontonikas, 2012; Manganelli and Wolswijk, 2009). As the situation in Greece was the most critical to the survival of the Eurozone during the European sovereign debt crisis, the country's bond yields summarize well the evolution of the crisis. Next, we consider the number of asylum applications as a proxy for the 'European migrant crisis'. While asylum-seekers are only one group of migrants, they were at the centre of the European migrant crisis when Germany opened its borders to Syrian refugees in 2015. Last, the Syria conflict is itself the subject of one of our EUSSUE topics. Here, for the comparison with issue attention in the EU, we take battle-related deaths as a relatively objective measure of the

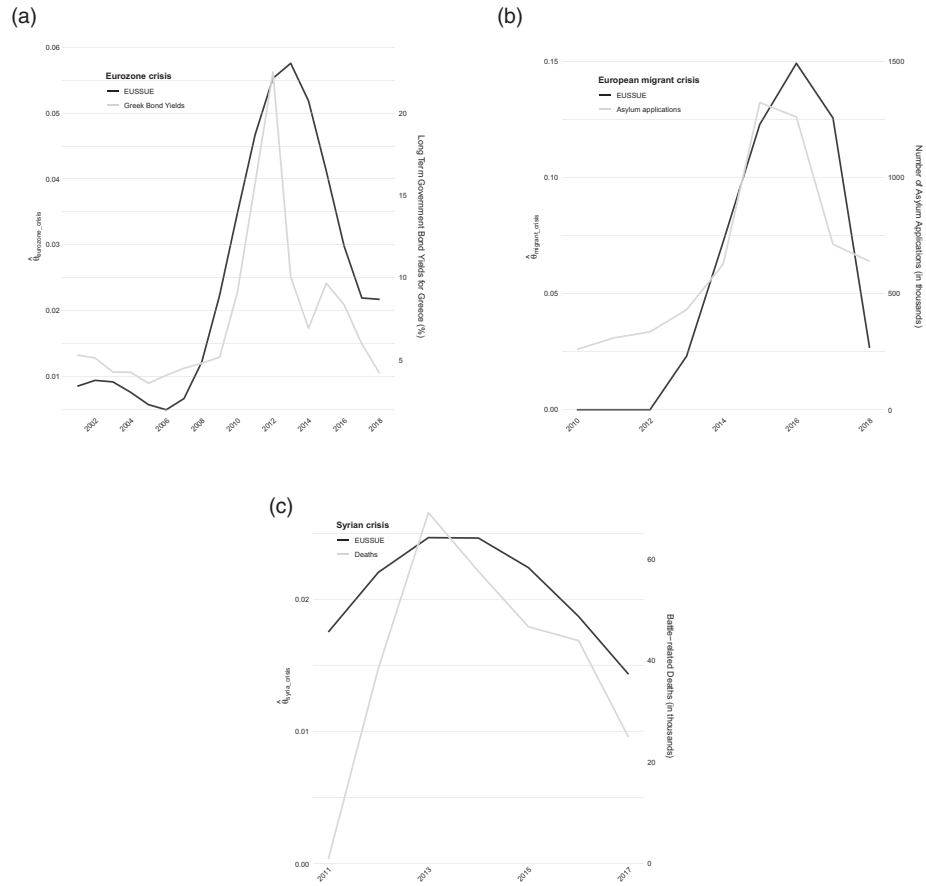

Figure 8. Temporal patterns of four EUSSUE topics and related external measures.
(a) Eurozone crisis, (b) migrant crisis and (c) Syria crisis.
Note: Greek bond yields and the number of asylum application are sourced from Eurostat and battle-related deaths in Syria from the World Bank. On the left, vertical axis is the estimated proportion of a given topic $\hat{\theta}_s$ where $s \in [1, 75]$, while the external indicator is measured on the right vertical axis.

conflict's intensity. The association between these proxies and topics is depicted in Figure 8.

The temporal patterns of external measures largely overlap with EUSSUE topics. The trends support the plausible assumption that issue attention varies for the most part in step with crisis intensity. Crises tend to generate and subsequently lose EU-level attention rapidly.

To assess the statistical strength of these relationships, Table 5 presents the results of several linear regressions, taking the EUSSUE topic proportions as the dependent variable and the external indicators as predictors. The table reports

Table 5. Relationship between three EUSSUE topics and thematically connected real-world quantities.

	Estimated topic proportion ($\hat{\theta}_s$)		
	Eurozone crisis	Migrant crisis	Syria crisis
Constant	0.003 (–0.008, 0.013)	–0.030 (–0.058, –0.001)	0.015** (0.011, 0.019)
Greek long-term bond yields	0.297** (0.178, 0.417)		
Asylum applications		0.135** (0.094, 0.176)	
Battle-related deaths			0.137* (0.042, 0.232)
Observations	18	11	7
R^2	0.598	0.821	0.613
Adjusted R^2	0.573	0.801	0.536
Residual Std. error	0.012 (df = 16)	0.026 (df = 9)	0.003 (df = 5)
F Statistic	23.849*** (df = 1; 16)	41.201*** (df = 1; 9)	7.930** (df = 1; 5)

Note: *p < 0.05; **p < 0.01; Number of bootstrap iterations is 10,000.

effect sizes and 95% confidence intervals with bootstrap standard errors. The normality assumption in the regression analysis is likely to be violated and bootstrapping permits to relax this assumption by producing robust standard errors. The estimates of the external measures are positive and significant at $\alpha = 5\%$.

Conclusion

In this paper, we presented a new measure of issue attention at the EU level, EUSSUE. Constructed from Agence Europe's daily bulletins using a combination of human and computerized text-classification methods, EUSSUE provides a generic EU-level measure of issue prevalence across 75 topics organized in 19 policy themes. Our validation exercise shows that it constitutes a reliable measure of the issues discussed in the entire universe of AEB published in English in the period 1979–2018. EUSSUE should be seen as a complement rather than a substitute to existing, actor-based indicators of agenda priority like the Agendas Project.

We believe that our measure of issue attention should be of broad interest to the EU studies research community. How the Brussels bubble reacts to world events and major crises is just one illustration. The generic nature of the EUSSUE dataset opens the possibility for researchers to deploy it – in entirety or in part – in different ways and contexts to investigate items of both low and high politics. Our indicator can be used as a dependent variable to explain variation in issue

attention; as an independent variable to explain other phenomena; or as a control variable to remove the effect of varying issue saliency on a causal relationship of interest. Alternatively, researchers may also decide to use the underlying textual corpus or the pre-processed data in order to, for example, remodel the topic space in a way tailored to specific research objectives or to apply supervised classification techniques.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors gratefully acknowledge financial support from European Research Council Grant No. 638154 (EUTHORITY).

Notes

1. The textual corpora (.RData) and time-series data (.csv and.RData) of the EUSSUE measure are available in the Online appendix
2. For a conceptual elaboration of these notions, see Beyers et al. (2018) and Netjes and Binnema (2007).
3. See website https://agenceurope.eu/en/about.html
4. See the Online appendix for a comparison with an alternative pre-processing strategy. Custom words removed are also listed in the Online appendix.
5. The deviation rests in the fact that we code these three issues at the theme level rather than as part of the 'EU Governance' theme. The issues receive sufficient attention in the AEB to warrant this more granular approach compared to the EU Policy Agendas project. If complete comparability is desired, these additional themes can be collapsed by researchers into the 'EU Governance' theme.
6. This step also involved merging two topics dealing with commodity prices into one after adjudging that they captured the same issue.
7. Results based on more formal clustering metrics are reported in the Online appendix.
8. The data can be explored online at https://eussue.shinyapps.io/eussue_shiny/
9. We add zeros to the Agendas Project data for years where no sentence is coded as belonging to a given topic and convert the counts into proportions, taking into account the length of European Council conclusions in each year.
10. In order to use document-level values, we must remove the effect of the year-covariate present in the EUSSUE measure otherwise. This does not lead, however, to a large discrepancy. For the topic 'CO2_emissions', the average effect is 0.00004. Documents were annotated with their date of publication to construct the time-series.

ORCID iD

Nicolas Lampach https://orcid.org/0000-0003-2507-501X

Supplemental Material

Supplemental material for this article is available online.

References

Adam BPS and Eschner B (2008) The contribution of the press to Europeanization of public debates. *Journalism: Theory, Practice & Criticism* 9(4): 465–492.

Alexandrova P (2017) Institutional issue proclivity in the EU: The European council vs the commission. *Journal of European Public Policy* 24: 755–774.

Alexandrova P, Carammia M, Princen S, et al. (2014) Measuring the European Council agenda: Introducing a new approach and dataset. *European Union Politics* 15: 152–167.

Alexandrova P, Carammia M and Timmermans A (2012) Policy punctuations and issue diversity on the European council agenda. *Policy Studies Journal* 40(1): 69–88.

Alexandrova P, Rasmussen A and Toshkov D (2016) Agenda responsiveness in the European Council: Public priorities, policy problems and political attention. *West European Politics* 39: 605–627.

Arghyrou M and Kontonikas A (2012) The EMU sovereign debt crisis: Fundamentals, expectations and contagion. *Journal of International Financial Markets, Institutions and Money* 22: 658–677.

Bäckstrand K and Elgström O (2013) The EU's role in climate change negotiations: From leader to 'leadiator'. *Journal of European Public Policy* 20: 1369–1386.

Bastin G (2002) L'Europe saisie par l'information (1952–2001): Des professionnels du journalisme engage aux content coordinators [Europe seized by information (1952–2001): journalism professionals as content coordinators]. *Cahiers Politiques*, pp.19–41. Brussels: L'Harmattan.

Baumgartner F, Green-Pedersen C and Jones B (2006) Comparative studies of policy agendas. *Journal of European Public Policy* 13: 959–974.

Baumgartner F and Jones B (1993) *Agendas and Instability in American Politics*. Chicago: University of Chicago Press.

Beyers J, Dür A and Wonka A (2018) The political salience of EU policies. *Journal of European Public Policy* 25: 1726–1737.

Blei DM (2012) Probabilistic topic models. *Communications of the ACM* 55(4): 77–84.

Blei DM and Lafferty JD (2006) Dynamic topic models. In: *Proceedings of the 23rd international conference on machine learning*. New York: ACM, pp.113–120.

Blei DM, Lafferty JD, et al. (2007) A correlated topic model of science. *The Annals of Applied Statistics* 1(1): 17–35.

Blei DM, Ng AY and Jordan MI (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research* 3(Jan): 993–1022.

Boydstun AE (2013) *Politics, the Media, and Agenda Setting*. Chicago: University of Chicago Press.

Busby A (2013) 'bursting the Brussels bubble': Using ethnography to explore the European parliament as a transnational political field. *Perspectives on European Politics and Society* 14(2): 203–222.

Carammia M, Princen S and Timmermans A (2016) From summitry to EU government: An agenda formation perspective on the European Council. *Journal of Common Market Studies* 54: 809–825.

Carmines EG and Stimson JA (1989) *Issue Evolution: Race and the Transformation of American Politics*. Princeton: Princeton University Press.

Catalinac A (2016) From pork to policy: The rise of programmatic campaigning in Japanese elections. *Journal of Politics* 78:1–18.

Chang J, Gerrish S, Wang C, et al. (2009) Reading tea leaves: How humans interpret topic models. In: Bengio Y, Schuurmans D, Lafferty JD, Williams CKI and Culotta A (eds) *Advances in neural information processing systems*. Vancouver, BC, pp.288–296.

De Bruycker I and Beyers J (2015) Balanced or biased? Interest groups and legislative lobbying in the European news media. *Political Communication* 32: 453–474.

Denny MJ and Spirling A (2018) Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis* 26(2): 168–189.

Diez-Medrano J (2004) *Framing Europe. Attitudes to European Integration in Germany, Spain, and the United Kingdom*. Princeton: Princeton University Press.

Dimitrov RS (2010) Inside Copenhagen: The state of climate governance. *Global Environmental Politics* 10: 18–24.

Dür A and Mateo G (2014) Public opinion and interest group influence: How citizen groups derailed the anti-counterfeiting trade agreement. *Journal of European Public Policy* 21: 1199–1217.

Dybowski TP and Adämmer P (2018) The economic effects of U.S. presidential tax communication: Evidence from a correlated topic model. *European Journal of Political Economy* 55: 511–525.

Fougier E (2010) Le journalisme Européen, un bien nécessaire. *L'Europe en Formation* 3: 149–173.

Greene D and Cross JP (2017) Exploring the political agenda of the European parliament using a dynamic topic modelling approach. *Political Analysis* 25: 77–94.

Grimmer J (2010) A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis* 18: 1–35.

Grimmer J and Stewart BM (2013) Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21(03): 267–297.

Groen L and Niemann A (2013) The European Union at the Copenhagen climate negotiations: A case of contested EU actorness and effectiveness. *International Relations* 27: 308–324.

Häge FM (2016) Political attention in the council of the European Union: A new dataset of working party meetings, 1995–2014. *European Union Politics* 17(4): 683–703.

Haug C and Berkhout F (2010) Learning the hard way? European climate policy after Copenhagen. *Environment: Science and Policy for Sustainable Development* 52: 20–27.

Hruschka DJ, Schwartz D, St. John DC, et al. (2004) Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods* 16(3): 307–331.

Kim SE (2018) Media bias against foreign firms as a veiled trade barrier: Evidence from Chinese newspapers. *American Political Science Review* 112: 954–970.

Klüver H (2011) The contextual nature of lobbying: Explaining lobbying success in the European union. *European Union Politics* 12:483–506.

Koopmans R (2007) Who inhabits the European public sphere? Winners and losers, supporters and opponents in Europeanised political debates. *European Journal of Political Research* 46: 183–210.

Koopmans R and Pfetsch B (2006) Obstacles or motors of Europeanization? German media and the transnationalization of public debate. *Communications* 31: 115–138.

Lucas C, Nielsen RA, Roberts ME, et al. (2015) Computer-assisted text analysis for comparative politics. *Political Analysis* 23(2): 254–277.

Manganelli S and Wolswijk G (2009) What drives spreads in the Euro area government bond market? *Economic Policy* 24: 191–240.

Marthoz J-P (2008) *Journalisme international*. De Boeck Supérieur. Brussels.
Mimno D, Wallach HM, Talley E, et al. (2011) Optimizing semantic coherence in topic models. In: *Proceedings of the conference on empirical methods in natural language processing*. Stroudsburg: Association for Computational Linguistics, pp.262–272.
Netjes CE and Binnema HA (2007) The salience of the European integration issue: Three data sources compared. *Electoral Studies* 26(1): 39–49.
Oberthür S (2011) The European Union's performance in the international climate change regime. *Journal of European Integration* 33: 667–682.
Ooms J (2017) Hunspell: High-performance stemmer, tokenizer, and spell checker for r. R package version 2.3. Available at: https://CRAN.R-project.org/package = hunspell (accessed 25 January 2020).
Osnabrügge M (2015) The European commission and the implementation of its legislative programme. *European Union Politics* 16(2): 241–261.
Peter J and De Vreese C (2004) In search of Europe. A cross-national comparative study of the European Union in national television news. *The International Journal of Press/Politics* 9: 3–24.
Peters G (1994) Agenda-setting in the European community. *Journal of European Public Policy* 1: 9–26.
Princen S (2007) Agenda-setting in the European Union: A theoretical exploration and agenda for research. *Journal of European Public Policy* 14: 21–38.
Princen S (2011) Agenda-setting strategies in EU policy processes. *Journal of European Public Policy* 18: 927–943.
Princen S and Rhinard M (2006) Crashing and creeping: Agenda-setting dynamics in the European Union. *Journal of European Public Policy* 13: 1119–1132.
Quinn KM, Monroe BL, Colaresi M, et al. (2010) How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science* 54: 209–228.
Rauh C (2014) Communicating supranational governance? The salience of EU affairs in the German bundestag, 1991–2013. *European Union Politics* 16(1): 116–138.
Risse T (2010) *A Community of Europeans? Transnational Identities and Public Spheres*. New York: Cornell University Press.
Roberts ME, Stewart BM and Tingley D (2019) stm: R package for structural topic models. *Journal of Statistical Software* 91(2): 1–40.
Roberts ME, Stewart BM, Tingley D, et al. (2014) Structural topic models for open-ended survey responses. *American Journal of Political Science* 58: 1064–1082.
Rosen-Zvi M, Griffiths T, Steyvers M, et al. (2004) The author-topic model for authors and documents. In: *Proceedings of the 20th conference on uncertainty in artificial intelligence*. Virginia, US: AUAI Press, pp.487–494.
Rothschild JE, Howat AJ, Shafranek RM, et al. (2019) Pigeonholing partisans: Stereotypes of party supporters and partisan polarization. *Political Behaviour* 41: 423–443.
Smith R (2007) An overview of the tesseract OCR engine. In: *Ninth international conference on document analysis and recognition (ICDAR 2007)*, Volume 2. New York: IEEE, pp.629–633.
Sorace M (2018) The European union democratic deficit: Substantive representation in the European parliament at the input stage. *European Union Politics* 19: 3–24.
Veen T (2011) Positions and salience in European Union politics: Estimation and validation of a new dataset. *European Union Politics* 12(2): 267–288.

Wallach HM, Murray I, Salakhutdinov R, et al. (2009) Evaluation methods for topic models. In: *Proceedings of the 26th annual international conference on machine learning*. New York: ACM, pp.1105–1112.

Warntjen A (2012) Measuring salience in EU legislative politics. *European Union Politics* 13: 168–182.

Wlezien C (2005) On the salience of political issues: The problem with 'most important problem'. *Electoral Studies* 24: 555–579.

Wonka A (2016) The party politics of the euro crisis in the German bundestag: Frames, positions and salience. *West European Politics* 39: 125–144.