

Policy complexity and implementation performance in the European Union

Maximilian Haag , Steffen Hurka  and Constantin Kaplaner 

Geschwister Scholl Institute of Political Science, Ludwig-Maximilians-University Munich, Munich, Germany

This study examines the relationship between the complexity of EU directives and their successful implementation at the national level. Moving beyond the state-of-the-art, we propose a comprehensive framework considering structural, linguistic, and relational dimensions of policy complexity. We argue that policy complexity entails higher transaction costs, hindering effective implementation. Using a novel dataset covering roughly 1000 directives from 1994 to 2022, we find strong evidence of policy complexity negatively impacting implementation performance. Moreover, we find that states with higher administrative capacity are better able to process high complexity efficiently and that Eurosceptic member states attract fewer infringement proceedings in highly complex policy environments than Europhile member states. This could alternatively point to strategic enforcement behavior of the Commission or to bureaucracies that are less Eurosceptic than their political masters might wish for. Our study thereby contributes to a deeper understanding of the challenges of successful implementation of EU directives.

Keywords: capacity, Euroscepticism, European Union, implementation, performance, policy complexity.

1. Introduction

Effective policy implementation is an essential task for any political system. To solve policy problems and reach policy objectives, political agreements must be put into practice. However, successful policy implementation is a difficult challenge for notoriously understaffed and overstrained public administrations, who continuously strive for optimal resource allocation in increasingly complex policy environments (Knill et al., 2023). Given this difficulty in implementing policies properly, there is a clear danger of a (possibly growing) mismatch between what is being decided at the political level and what is implemented by the bureaucracy. As Pressman and Wildavsky famously put it in their title, policy implementation is all too often a story of “How great expectations in Washington are dashed in Oakland” (Pressman & Wildavsky, 1984). One of the key questions that emerges in this context is whether and to what extent the complexity of the policies can be made responsible for deficits at the implementation stage.

We aim to contribute to answering this question by investigating the relationship between the complexity of a policy and implementation performance in the member states of the European Union (EU). Specifically, we study how the complexity of EU directives affects the timeliness and correctness of policy implementation, both directly and conditional on the administrative capacity and political willingness of the member states. Empirically, our quantitative analysis covers up to 20,509 implementation cases of 998 EU directives in all EU member states and a time period of 28 years (1994–2022). Our theoretical argument focuses on the transaction costs that the complexity of EU directives imposes on national policymakers and bureaucracies. Echoing Pressman and Wildavsky’s (1984) observation on how Washington’s ambitions may come to a halt in Oakland, Brussels’s intentions might well be laid to rest in Paris or Athens.

The EU features interesting properties that make it a promising case for the study of policy complexity and implementation. When the EU adopts laws in the form of so-called directives, it sets out binding objectives that must be achieved by the member states, but leaves discretion to the member states on how to achieve those goals.

Correspondence: Maximilian Haag, Geschwister Scholl Institute of Political Science, Ludwig-Maximilians-University Munich, Geschwister-Scholl-Institute of Political Science, Oettingenstraße 67, 80538 Munich, Germany. Email: maximilian.haag@gsi.uni-muenchen.de

Accepted for publication 10 January 2024.

In turn, the member states are then responsible for transposing EU directives into their national legal systems by adopting and implementing national legislative measures. Because of this setup, the EU provides for a particularly good case to study the relationship between policy complexity and policy implementation. EU decision-making over new policies and their evaluation are at least partially externalized and formally separate from their national implementation in the member states. The system is designed in a way that minimizes mediating or disguising factors between policy decisions and their implementation. This separation lowers the chances of policymakers cherry-picking which policies to pursue or the opportunity to disguise their inability to implement previously decided policies, thereby reducing sampling bias and allowing researchers to study the policy-related factors of implementation performance.

To theorize how the complexity of EU directives affects the quality of implementation in the member states within a common theoretical framework, we adopt a transaction cost perspective and argue that the complexity of policies enhances the costs for those who engage with the policy, including political decision-makers, ministerial officials, and bureaucrats. In particular, those costs manifest themselves in a higher workload, more demanding cognitive challenges, and a more difficult legal environment. Previous research, which has been built on the assumption that all of those costs can be approximated for by the number of recitals that precede a directive's enacting terms (Toshkov, 2010), has resulted in rather inconclusive and contradictory findings on the link between the number of recitals and implementation performance. Therefore, we argue that to properly measure the "burden of implementation" induced by the complexity of EU policy, the actual legal provisions of a given directive should be taken into account. Specifically, we propose to measure policy complexity on the basis of text of the directives and suggest that complexity can feed from different sources, such as structural, linguistic, and relational factors. In particular, we propose a measure of complexity that incorporates various aspects like the level of detail, the readability, the conceptual breadth, the interdependence of legal provisions, and the embeddedness of the directive in the legal landscape. This allows us to fully capture the various types of transaction costs that come with complex policies in one comprehensive measure.

Our study demonstrates that implementation varies with the level of policy complexity across three indicators of implementation performance in the EU that reflect the full cycle of the implementation process: (1) the *notification* of (any) national implementing measures to the Commission, (2) the timeliness of the *transposition* of the directive into national implementing measures, and (3) the Commission's evaluation of the national implementation through the so-called *infringement* procedures. While our results indicate that states are more likely to simply notify transposition measures for more complex directives, complex directives are also associated with less timely transposition and a higher probability of infringement proceedings. In addition, we demonstrate that states with more administrative capacity are somewhat better able to mitigate the negative consequences of policy complexity. Finally, our results indicate that Eurosceptic governments are less likely to attract infringement proceedings when the directives that need to be implemented increase in complexity. This could either point to strategic enforcement behavior of the Commission in more uncertain legal environments or could be read as an indication that euroscepticism at the political level does not easily and quickly translate into euroscepticism at the administrative level. Our findings contribute to a better understanding of the role of policy complexity in the implementation process and provide policymakers with valuable insights into the circumstances under which the complexity of their political decisions makes a difference for whether and how those decisions are implemented. On a broader scale, our research provides new insights into the crucial question of how political and institutional arrangements can mitigate or amplify adverse consequences of policy complexity.

2. Directive complexity and implementation performance in the EU

In the EU, the implementation of EU policies into national law is governed by Article 288 of the Treaty on the Functioning of the European Union (TFEU). While policies of type "regulation" or "decision" are directly binding in their entirety, Article 288 specifies that policies of type "directive" shall be binding only "as to the result to be achieved" leaving "the choice of form and methods" to the member states (Hurka & Steinebach, 2021). Directives are a key component of the EU's legislative output. While most EU laws come in the form of regulations, "the most important regulatory initiatives in the EU take the form of directives" (Toshkov, 2011, p. 175). Given their comparatively open legal formulation, directives need to be translated into national measures by the

member states—a process commonly referred to as *transposition*. Once a member state has adopted a relevant national measure, it is obliged to notify the European Commission, which monitors the timely and correct transposition of EU directives. Since transposition is a political process, however, timely and correct transposition cannot be taken for granted, as is illustrated by the 57,608 infringement decisions recorded between 1991 and 2023 in the Commission's infringement decision database¹ at the time of writing. This is a quite an astonishing number if we keep in mind that according to EUR-Lex, only 1445 directives were adopted in legislative procedures during the same time period (see also Fig. S4 of Appendix S1 for a graphical display of directives and infringements in our sample).²

The literature on policy implementation in the EU in general, and transposition delays in particular, is very extensive (for an excellent review, see Treib, 2014), which is why we focus our state-of-the-art discussion on how previous research treated the main variable we are interested in: the complexity of EU directives. Specifically, we found eight studies that explicitly theorize the link between the complexity of a directive and member states' implementation performance (Haverland & Romeijn, 2007; Kaeding, 2006; König & Luetgert, 2009; Luetgert & Dannwolf, 2009; Mastenbroek, 2003; Steunenberg & Rhinard, 2010; Thomas, 2013; Zhelyazkova, 2013) and seven studies in which complexity is included as a control variable (Finke, 2018; Paasch, 2022; Paasch & Stecker, 2021; Thomson et al., 2020; Zhelyazkova et al., 2017; Zhelyazkova & Thomann, 2022; Zhelyazkova & Yordanova, 2015). Table 1 summarizes those studies' findings.³

Table 1 reveals highly contradictory findings regarding the relationship between directive complexity and implementation performance (see also Treib, 2014, 26 f). While some studies argue that higher complexity leads to worse implementation, others find no correlation between complexity and implementation performance. Surprisingly, some even suggest that higher complexity can improve implementation, as states may anticipate the complexity and allocate additional resources beforehand. In our perspective, there are three main reasons for these inconclusive findings.

Firstly, studies differ significantly in how they conceptualize the dependent variable. Some focus on transposition delay, while others examine the likelihood of correct transposition or the time taken by states to transpose directives. Additionally, some studies go beyond transposition and analyze practical compliance and the initiation of infringement proceedings.

Secondly, the data foundations of these studies vary greatly. Some analyze individual member states, specific policy areas, and only a small number of directives, while others consider hundreds of directives, multiple member states, and several decades of data.

Lastly, the complexity of EU directives is primarily measured by the number of recitals that precede the directive's enacting terms. Although this uniform approach to measuring complexity should lead to highly consistent findings, the opposite is true. The inconsistencies can partly be attributed to the diverse research designs and data foundations mentioned earlier. Moreover, the inconsistent effect of recitals may also be due to the fact that they do not fully capture the complexity of a directive. Most importantly, the number of recitals does not provide insight into their actual content or the content of the directive's enacting terms.

Several issues can be highlighted with regard to the use of recitals as proxies for complexity. First, as pointed out by Toshkov (2010, p. 25), the number of recitals has not only been used to measure the complexity of a directive but also its salience (Beyers et al., 2018; e.g., Häge, 2007) and “political sensitivity” (e.g., Steunenberg & Kaeding, 2009).⁴ Accordingly, it is not entirely clear whether the number of recitals measures a directive's complexity, its importance, or the degree of political conflict it entails. Second, recitals do not hold normative content and as such, they do not have any autonomous legal effects (Baratta, 2014, p. 302). As the EU institutions jointly pointed out in the *Interinstitutional Agreement of 22 December 1998 on common guidelines for the quality of drafting of Community legislation*⁵: “The purpose of the recitals is to set out concise reasons for the chief provisions of the enacting terms, without reproducing or paraphrasing them. They shall not contain normative provisions or political exhortations.” Accordingly, when member states transpose EU directives, they are not bound by the content of the directive's recitals, but by the content of its enacting terms (i.e., its articles and paragraphs). Finally, legal scholars have shown that the way recitals have been used by EU policymakers has changed substantially over time. As den Heijer et al. (2019) demonstrate, policymakers have often misused recitals to send political messages and most importantly, this tendency has increased over time. This can create a major problem: if recitals are used for different purposes over time, their number per directive may be skewed, and the usefulness

Table 1 Overview of EU implementation studies including complexity as an independent or control variable

Study	Dependent variable	Complexity as	Operational	Empirical scope	Finding
Mastenbroek (2003)	Time until transposition	IV	# of measures	229 directives (enacted 1995–1998), the Netherlands	Directives of higher complexity are transposed faster
Kaeding (2006)	Transposition delay	IV	# of recitals	106 transport policy directives (1957–2004), five member states	Complexity strongly increases transposition delay states
Haverland and Romeijn (2007)	Transposition delay	IV	# of recitals	67 social policy directives (1975–1999), five member states	No effect of complexity on transposition delay states
Luetgert and Dannwolf (2009)	Time until transposition (first national measure)	IV	Length of time allocated for transposition, amending nature of directive, legislative procedure	1192 directives (1986–2003), nine member states	Mixed evidence across indicators
König and Luetgert (2009)	Transposition delay	IV	# of measures	1591 directives (1986–2002), 15 member states	Complexity increases transposition delay
Steunenberg and Rhinard (2010)	Time until transposition	IV	# of recitals	2267 cases of transposition (1978–2002), five member states	Directives of lower complexity are transposed faster
Thomas (2013)	Timely transposition	IV	# of recitals	10,655 country-directive observations (1994–2003), 15 member states	Complexity negatively affects timely transposition in election years
Zhelyazkova (2013)	Likelihood of correct transposition	IV	# of recitals/provision, # of sentences/provision	136 provisions in four directives (1999–2001), 15 member states	Weak and unstable negative effect of complexity
Zhelyazkova and Yordanova (2015)	Initiation of infringement proceedings	Control	# of recitals	97 directives (2004–2010), 25 member states	Complexity increases the probability of infringement proceedings
Zhelyazkova et al. (2017)	Initiation of infringement proceedings/practical compliance	Control	# of recitals	65 directives (2000–2010), 27 member states	No effect on transposition, mixed evidence on practical compliance
Finke (2018)	Time until transposition	Control	# of recitals	684 directives (2000–2013), eight member states	No effect of complexity
Thomson et al. (2020)	Transposition delay	Control	# of recitals	Nine directives (2007–2009), 27 member states	No effect of complexity on transposition delay

(Continues)

Table 1 Continued

Study	Dependent variable	Complexity as	Operational	Empirical scope	Finding
Paasch and Stecker (2021)	Transposition delay	Control	# of recitals	850 directives (1990–2018), Germany	Mixed evidence
Paasch (2022)	Time until transposition	Control	# of recitals	762 subnational measures (1990–2018), Länder in Germany	Higher complexity speeds up transposition
Zhelyazkova and Thomann (2022)	Practical compliance	Control	# of recitals	17 directives (2007–2013), 27 member states	No effect of complexity on practical compliance

of recitals as an indicator of the complexity of a directive may be questioned. In fact, if we recall that studies on policy implementation in the EU often cover very different time periods (see Table 1), we should not be surprised that the number of recitals yields highly variant results. Therefore, the number of recitals in a legal text might not only capture a variety of latent features of the decision-making process or the political system, but also creates a measurement problem in efforts to isolate the effect of complexity on transposition in the EU.

Thus, the findings on the role of policy complexity in implementation performance are inconclusive and the reasons for this might lie in the research designs, conceptualizations of the dependent variables, and the questionable tradition of equating policy complexity exclusively with the number of recitals we find in a directive. In this article, we argue that the complexity of a directive can feed from different sources and should be measured on the basis of the directive's enacting terms. Specifically, we propose a more encompassing measure of complexity that is based on the enacting terms of a directive and does not only take structural features into account but also linguistic and relational aspects. Moreover, in the following section, we contribute theoretically to the implementation literature by making the case that the impact of complexity on implementation performance should be moderated by the implementation structures we find in the member states, in particular their administrative capacity and their willingness to comply with EU law. In doing so, we go beyond the existing state-of-the-art, which mostly scrutinized the direct effect of complexity on implementation performance, but did not theorize the circumstances under which we should assume this effect to be dampened or amplified.

3. Theorizing direct and conditional effects of directive complexity on implementation performance

The literature review has shown that existing research on policy implementation in the EU either investigates whether and when EU directives are *transposed* into national law or the extent to which the Commission initiates *infringement proceedings* due to noncompliance in the member states. In this study, we look at both aspects and conceptualize good implementation performance as a combination of timely transposition and correct application. Accordingly, our theoretical approach is geared towards the explanation of implementation performance, while we distinguish between the transposition stage and the application stage empirically (see the research design section). Our main theoretical argument is based on the idea (a) that complex policies generate higher transaction costs than simple policies, (b) that those costs can take different forms, and (c) that they affect policymakers, ministerial officials, and bureaucrats alike.

Policy implementation is a costly endeavor and states need to invest heavily in human and financial resources to put their policies into practice. As those policies become more complex, the transaction costs that accrue for implementers increase. We argue that those transaction costs mainly result from three different aspects, which jointly affect the ease of the knowledge acquisition process for politicians and bureaucrats (Katz & Bommarito II, 2014; Senninger, 2023). First, the size and detail of a policy affect the workload for those who need to

transpose the policy into national law as well as for those who are eventually responsible for practical application. If a directive is very long and contains many intricate provisions, this should prolong the process of translating the directive into national law. Moreover, the greater the number of provisions that can be violated, the higher the probability that such violations will eventually occur. Secondly, the way policies are formulated affects the cognitive burdens faced by policymakers, ministerial officials, and bureaucrats alike (Senninger, 2023). If laws are more difficult to understand and harder to interpret, this should make it harder for legislative actors to efficiently formulate corresponding national laws and it should also increase the risk of noncompliance when policies are put into practice. Finally, when legal provisions interact with each other and laws are integrated into a dense web of existing legislation, the need to take those interdependencies into account makes both transposition and proper application more challenging. Previous research has shown that the web of EU laws has been growing significantly over time (Koniaris et al., 2018) and there has also been a considerable increase in the extent to which legal provisions reference each other in individual EU laws (Hurka et al., 2022, p. 1521). We argue that those patterns contribute to an increasingly complex legal environment, which makes it more demanding for policymakers and ministerial officials to transpose EU directives in time and for implementers to apply the laws appropriately.

Accordingly, not all policies are equally costly to implement, and the size of the burden different policies put on policymakers, ministerial officials, and public administrations varies. At the *transposition* stage, we should expect complexity to impact the time it takes for states to translate a directive's provisions into national law. While short, simple, and legally isolated directives should not entail particularly high transaction costs at the transposition stage, long, convoluted, and legally interdependent directives should make it significantly more difficult for states to comply with transposition deadlines. Specifically, if EU directives increase in terms of their size, their level of detail, their syntactic difficulty, their conceptual variety, their legal interdependence and embeddedness, policymakers and ministerial officials should find it more challenging to formulate national laws that are in line with the directive's spirit in an efficient manner. At the *application* stage, higher complexity entails a higher workload, more demanding cognitive challenges and a more uncertain legal environment for bureaucrats. Theoretically, more complex policies would require public administrations to enhance their cognitive, human, and financial resources to make sure that policies continue to be implemented properly. However, since public administrations are typically not able to quickly increase their resources in lockstep with the requirements set by the policy, the risk of bad application increases. This, in turn, should be reflected in a higher risk of infringement proceedings initiated by the European Commission. We thus formulate our first hypothesis on the direct effect of complexity on implementation performance as follows:

Hypothesis 1. The higher the complexity of the policy to be implemented the worse the implementation performance.

While this direct link between complexity and implementation performance is intuitive, there might be ways how adverse consequences of complex policies can be mitigated or even amplified at the implementation stage. Most importantly, this concerns the role of resources and political will (Sabatier & Mazmanian, 1980). For example, in the implementation literature, the notion that higher administrative capacity positively affects implementation performance is widespread (e.g., Limberg et al., 2021; Toshkov, 2010). Yet, the literature typically evaluates the role of administrative capacity in isolation from the policy that needs to be implemented. König and Mäder (2014), for instance, evaluate the effects of directive complexity and bureaucratic efficiency on member state compliance separately and find that while complexity negatively affects compliance, bureaucratic efficiency does not affect compliance significantly. In our view, the important question that flows from those studies is whether administrative capacity might have the same effect across cases. Specifically, we would expect the payoffs of high administrative capacity to be maximized when the policy that needs to be implemented is of high complexity and that well-equipped bureaucracies can offset the marginal costs of complex directives, at least to some degree.

Hypothesis 2. The higher a state's administrative capacity, the smaller the negative effect of complexity on implementation performance.

Finally, the effect of complexity on implementation performance might also be conditioned by governments' willingness to comply. Most importantly, this willingness has been operationalized by the degree of public and governmental Euroscepticism and the straightforward theoretical expectation of the literature is that more Euroscepticism entails worse compliance. However, the empirical record of this reasonable expectation is inconclusive (Toshkov, 2019). For example, Williams (2018) argues that governments are responsive to public Euroscepticism at the implementation stage and slow down transposition when the public is opposed to European integration. In contrast, Pircher and Loxbo (2020) clearly reject this claim based on their analysis of more than 1000 directives on the single market adopted between 1997 and 2016. To address those mixed findings, Toshkov (2019, emphasis in original) argues that scholars need to "examine more closely the actual *mechanisms* through which Eurosceptic government and public can influence (or not) compliance and enforcement of EU law."

We argue that one potential reason for the inconclusive findings in the literature is that the role of Euroscepticism is hardly ever contextualized and is typically assumed to exert its effect on implementation performance directly. The complexity of a directive might be one such contextual factor that conditions the impact of governmental Euroscepticism on implementation performance. Specifically, while Euroscepticism has various roots and motivations depending on the ideological left-right orientation of political parties (van Elsas & van der Brug, 2015), the reproach that the EU has become excessively complex and out of touch with the needs and preferences of the population is shared across all Eurosceptic parties. Accordingly, Euroscepticism might exert its impact most strongly in situations in which the legislation that comes from Brussels is perceived (or can easily be portrayed as) excessively complex. In general, Eurosceptics should be less compliant with EU legislation than Europhiles, that is, their compliance behavior should already start from a different baseline when policies are not complex. However, as policies grow in terms of complexity (i.e., transaction costs increase), the willingness of Eurosceptics to bear those additional costs should decrease more strongly than the willingness of Europhiles. In other words, Europhiles should be more inclined to bear the complexity-induced transaction costs the EU creates than Eurosceptics and hence, the negative effect of policy complexity on implementation performance should be more pronounced the more a national government opposes the European integration project.

One important caveat to this argument rests with the fact that all stages of implementation in the EU transposition process are impacted by governmental Euroscepticism to at least some extent. While transposition is carried out by legislators and ministry officials, practical implementation is the task of the bureaucracy. Yet, we assume that the lack of willingness to comply at the political level should translate to the administrative level through political influence and the specific adopted national legislative measures.

Hypothesis 3. The more national governments are in favor of European integration, the smaller the negative effect of complexity on implementation performance.

4. Research design

In order to study implementation performance in the EU, we focus on three key dependent variables: notification, transposition delay, and infringement. Whereas *notification* relates to whether or not a member state has reported any transposition measures to the European Commission, *transposition delay* is a measure of the time it takes for a member state to transpose a directive into national law, *infringement* is a measure of whether a member state is found to be in breach of EU law regarding the implementation of a particular EU policy. All variables are measured at the policy-country level. As such, we are able to study the impact of policy characteristics, namely their complexity, on the implementation performance of member states. We consider all legislative processes in the EU between November 1, 1993 and December 31, 2022 that have resulted in a directive and their subsequent implementations in the member states resulting in a sample of 20,509 implementation cases of 998 EU directives in all EU member states and a time period of 28 years (1994–2022).

4.1. Measuring implementation performance in the EU

To create a measure of transposition delay, we rely on the information on member state notification contained in the EU's EUR-Lex⁶ XML document notices for a given directive. These notices include a transposition deadline as well as the national (member state) measures that transpose the given directive into national law along with the notification date for the national measure, that is, when a member state informed the Commission of their transposition via a specific national measure.⁷

It is also important to point out that member states in many cases do not transpose directives with one but many national policy measures. We first create a binary measure of *notification* indicating whether a member state has notified any measures for a particular directive.⁸ We then classify each of the national implementation measures listed for a particular directive as either *existing*, that is, legislation adopted on the national level prior to any directive on the EU level, *on-time*, that is, after the adoption of the directive but before the specified deadline, or *late*, that is, after the specified deadline. We then aggregate the classifications for the individual implementations on the country level per directive such that we obtain one observation per directive-country (see also below). In order to avoid mis-classifying member state efforts where only pre-existing, potentially low-fitting, national measures have been reported to the Commission as a way of signaling compliance only without following up on it (Zhelyazkova & Yordanova, 2015), we only consider cases where new legislation has been passed at the national level as *notified* in our analysis of *notification*. In order to arrive at a measure that captures the delay of transposition, we calculate the ratio of measures that were notified on time versus all notified measures (not including measures for which we do not have information on their timeliness) by a particular member state for a directive and control for the ratio of pre-existing measures. It is important to note that, as König and Luetgert (2009) point out, the amount of measures or presence of any measures at all does not tell us whether the transposition of a particular has been “correct” or complete in a particular member state (see König & Luetgert, 2009, pp. 169–170 for a more extensive discussion).⁹ Table 2 provides an overview of transposition delay by country.

For our measurement of infringement procedures as part of implementation performance, we rely on the information contained in the “Berlin Infringement Database” (BID) dataset (Börzel, 2021). The dataset records infringement procedures of the European Commission against member states by CELEX reference of the legislative document. We match the CELEX id of the legislative document to the CELEX id of the directive in our dataset and then aggregate these cases into a binary variable indicating whether an infringement procedure has been started against the respective member state for a particular directive. In order to capture infringements due to the quality of implementation rather than late transposition, we only consider infringement procedures that have been started due to incorrect transposition or bad application of directives (cf. Börzel, 2021).¹⁰

It should be noted that the use of infringement procedures as a measure of implementation performance or compliance are, by design, imperfect. As has been pointed out in the literature (König & Mäder, 2014; e.g., Toshkov, 2019, 2010), infringement procedures are not only a measure of implementation quality, but also a measure of the Commission's enforcement efforts and subject to strategic considerations on both the Commission's and the member states' side. Yet, employing infringement data can nonetheless help us draw a more complete picture of implementation in the EU as it is a measure of implementation quality that is not based on self-reporting by member states, such as the notification of transposition measures.¹¹

4.2. Measuring policy complexity for the study of implementation performance

The review of the literature on the implementation of EU directives has revealed that the measurement of policy complexity is neither uniform nor exhaustive. In contrast, our operationalization of policy complexity in this study encompasses structural, relational, and linguistic factors of a policy text to capture the transaction cost associated with the implementation of a policy (Hurka et al., 2022; Hurka & Haag, 2020; Katz & Bommarito II, 2014; Schuck, 1992). To measure the complexity of directives, we employ data from the EUPLEX dataset (Hurka et al., 2022), extending it by final texts.

Structural complexity encompasses the size and depth of a legislative text. Size is measured by the number of policy elements (e.g., recitals, paragraphs, points, and indents), while depth is operationalized as the average hierarchical level of the policy elements within the text (e.g., articles, paragraphs, points/indents). Our measurement thereby includes the often-used number of recitals and text length but goes beyond these measures by providing

Table 2 Summary statistics for implementation performance by country

MS	N	Impl. / Dir.	Notif. rat.	Exist rat.	On time rat.	Late rat.	NA rat.	Infring. rat.
AUT	987	5.314	0.806	0.059	0.246	0.695	0.063	0.054
BEL	998	4.420	0.759	0.030	0.229	0.741	0.031	0.090
BGR	447	2.528	0.678	0.047	0.263	0.691	0.011	0.036
CYP	551	2.336	0.730	0.068	0.127	0.805	0.023	0.023
CZE	551	13.506	0.922	0.222	0.189	0.589	0.004	0.028
DEU	998	3.035	0.845	0.124	0.271	0.605	0.043	0.062
DNK	998	3.260	0.721	0.036	0.290	0.674	0.087	0.029
ESP	998	2.152	0.878	0.125	0.262	0.613	0.021	0.077
EST	551	3.230	0.697	0.016	0.330	0.654	0.013	0.022
FIN	987	5.304	0.789	0.078	0.255	0.667	0.023	0.044
FRA	998	3.304	0.749	0.028	0.225	0.747	0.020	0.093
GBR	980	4.362	0.732	0.055	0.189	0.756	0.281	0.057
GRC	998	1.317	0.757	0.054	0.127	0.820	0.179	0.079
HRV	171	5.439	0.673	0.009	0.262	0.730	0.006	0
HUN	551	8.998	0.706	0.022	0.187	0.791	0.024	0.036
IRL	998	2.016	0.747	0.108	0.166	0.727	0.027	0.049
ITA	998	1.530	0.758	0.031	0.144	0.824	0.042	0.102
LTU	551	9.263	0.840	0.129	0.274	0.597	0.028	0.012
LUX	998	1.539	0.748	0.036	0.253	0.711	0.012	0.044
LVA	551	3.964	0.871	0.153	0.261	0.587	0.024	0.020
MLT	551	1.966	0.717	0.056	0.255	0.688	0.041	0.019
NLD	998	2.338	0.761	0.015	0.315	0.670	0.045	0.047
POL	551	4.871	0.746	0.051	0.176	0.773	0.014	0.047
PRT	998	2.353	0.842	0.089	0.150	0.761	0.035	0.071
ROU	447	3.539	0.738	0.109	0.189	0.702	0.028	0.016
SVK	551	5.443	0.840	0.139	0.289	0.572	0.015	0.014
SVN	551	3.721	0.715	0.046	0.183	0.771	0.011	0.035
SWE	987	4.309	0.671	0.033	0.254	0.713	0.224	0.048

Note: Ratios (rat.) relate to average ratios of the respective categories per country. Notification rates include pre-existing measures.

a more fine-grained measurement of the structural complexity of a policy text. Relational complexity increases with the degree of integration of a proposal within the existing legal order and the interconnections between individual legal provisions. This complexity is measured by the number of cross-references (both internal and external) per article in the Commission proposal and the final text. Linguistic complexity is associated with the syntactic and semantic properties of the text. Linguistic complexity is operationalized using the LIX score, which considers average sentence and word lengths, and word entropy, which measures the variety of the words contained in a text. To create an aggregated measure of complexity, we standardize and average all six individual text measures (see Hurka et al., 2023 for more detail on this procedure).

4.3. State capacity, EU position, and controlled factors

Based on our theoretical argument, we expect that member states with higher levels of administrative capacity are better able to deal with the complexity-induced cost of transposition and thus are able to implement EU directives faster than member states with lower levels of administrative capacity. In order to measure administrative capacity, we employ the World Bank's Worldwide Governance Indicators (WGI) (Kaufmann et al., 2010).¹² The WGI covers the timespan from 1996 to 2021 in 2-year intervals. We interpolate missing years with the information from the following year as data are based on retrospective reports and assessments. For 2022, we use WGI data from 2021. We employ the WGI's measure of *government effectiveness* as a proxy for administrative capacity.

The WGI's government effectiveness measure is based on the perception of the quality of public services, the quality of the civil service, and the degree of its independence from political pressures.

In order to measure critical government positions towards the EU at the time of the publication of the respective directive, we rely on positions extracted from manifesto data (Lehmann et al., 2023; Lowe et al., 2011) taking into account positive and negative mentions of the EU (per 108 for positive per 110 for negative references to the EU) in the manifesto dataset. We follow Lowe et al. (2011) and use negative mentions of the EU as the *R* and positive mentions as the *L* category resulting in a measure that takes on lower/negative value for more positive stances towards the EU. While the use of party manifestos for assessing party positions is not without its flaws and has attracted rightful criticism, such as the inability to capture intra-party dissent or issues regarding the differing length manifestos across countries and parties (e.g., Marks et al., 2007), the data allow researchers to obtain position measure over long periods of time for a great variety of parties as is required for our analysis. Our approach also reflects how other recent studies measured government euroscepticism, both in studies on policy implementation (e.g., Pircher & Loxbo, 2020; Williams, 2018) and in other empirical contexts (e.g., Rauh et al., 2020; Wratil et al., 2023).

We also consider a range of control variables: In order to capture possible effects on the political fit and will of member states to transpose a particular measure, we include a dummy variable for EP involvement in the decision-making process. Whereas some procedures, such as co-decision, include the EP as a veto player, others, such as consultation, allow the member state governments in the Council to act more or less independently of the EP. Given the greater likelihood of more dispersed preferences when the EP is involved, we could expect that EP involvement leads to more complex directives that are also more likely to be further away from the collective Council's ideal positions (see also Hurka, 2023). We additionally include a dummy variable indicating whether a directive amends another act in order to control for the effects on the complexity measurement (Hurka et al., 2022) and possibly lightened burden of implementation for cases where measures already exist from a previous transposition effort.

4.4. Model specification

To study the impact of policy complexity on the implementation of EU directives, we proceed in multiple steps. For each dependent variable (DV) (notification [N], transposition delay [D], and infringement [I]) we estimate four models. As a baseline model (1) we first estimate a simple linear model using ordinary least squares containing only our main variables of interest. Then, (2) we employ linear mixed effect models including our main independent variables and our control variables with directive and country-year level random intercepts of the following form¹³:

$$\begin{aligned}
 DV_i &\sim N(\alpha_{j[i],k[i]}, \sigma^2) \\
 \alpha_j &\sim N\left(\gamma_0^\alpha + \gamma_1^\alpha(\text{complexity}) + \gamma_2^\alpha(\text{anti} - \text{EU Position}) + \gamma_3^\alpha(\text{amending}_1) + \gamma_4^\alpha(\text{EP} - \text{involvement}_1), \sigma_{\alpha_j}^2\right), \text{ for directive } j = 1, \dots, J \\
 \alpha_k &\sim N\left(\gamma_0^\alpha + \gamma_1^\alpha(\text{statecapacity}), \sigma_{\alpha_k}^2\right), \text{ for country} - \text{year } k = 1, \dots, K
 \end{aligned}$$

Lastly, we estimate two more linear mixed effect models additionally containing interaction effects between policy complexity and state capacity (3) and between policy complexity and EU position (4).

$$\begin{aligned}
 DV_i &\sim N(\alpha_{j[i],k[i]}, \sigma^2) \\
 \alpha_j &\sim N\left(\gamma_0^\alpha + \gamma_1^\alpha(\text{complexity}) + \gamma_2^\alpha(\text{anti} - \text{EU position}) + \gamma_3^\alpha(\text{amending}_{\text{yes}}) + \gamma_4^\alpha(\text{EP} - \text{involvement}_1), \sigma_{\alpha_j}^2\right), \text{ for directive } j = 1, \dots, J \\
 \alpha_k &\sim N\left(\gamma_0^\alpha + \gamma_1^\alpha(\text{statecapacity}) + \gamma_2^\alpha(\text{complexity} \times \text{statecapacity}), \sigma_{\alpha_k}^2\right), \text{ for country} - \text{year } k = 1, \dots, K \\
 DV_i &\sim N(\alpha_{j[i],k[i]}, \sigma^2) \\
 \alpha_j &\sim N\left(\gamma_0^\alpha + \gamma_1^\alpha(\text{complexity}) + \gamma_2^\alpha(\text{anti} - \text{EU position}) + \gamma_3^\alpha(\text{amending}_1) + \gamma_4^\alpha(\text{EP} - \text{involvement}_1) + \gamma_5^\alpha(\text{complexity} \times \text{anti} - \text{EU position}), \sigma_{\alpha_j}^2\right), \text{ for directive } j = 1, \dots, J \\
 \alpha_k &\sim N\left(\gamma_0^\alpha + \gamma_1^\alpha(\text{statecapacity}), \sigma_{\alpha_k}^2\right), \text{ for country} - \text{year } k = 1, \dots, K
 \end{aligned}$$

Alternative modeling approaches for transposition delay data include generalized linear, count and survival models (see also Toshkov, 2010). We do not consider these kinds of models appropriate to examine the relationship between policy complexity and transposition efforts. First, there are two reasons why we rely on a combination of linear models (LM) and linear mixed effects models (LME) instead of generalized linear (GLM) or generalized linear-mixed effect models (GLME) with a logit link function.¹⁴ On the one hand, we deem LMs more fitting when including grouping fixed-effects, since all-zero or all-one groups are not dropped from the estimations which can be problematic given the amount of fixed-effects (see also Beck, 2020). From a practical standpoint, LMEs additionally allow us to include more stringent random-effects specifications without running into convergence issues as we would do with GLME models with a logit link function. Therefore, we chose to opt for mainly LME models with robust standard errors although those come with their own downsides such as the potential for out-of-bounds predictions.¹⁵

Second, count data models are not appropriate as we are not interested in the amount of national legislation resulting from EU directives, but in how member states are able to deal with the transaction cost induced by policy complexity and thus their (in)ability to implement EU policy given this complexity. Here, we do not deem the amount of legislation a suitable proxy for implementation performance as such a measure (a) does not take into account the length of the individual national measures, which in itself may be a by-product of the national legal landscape or tradition and is not necessarily related to effort, and (b) cannot tell us anything about member states' ability to deliver on time or the "completeness" (see above) of transposition efforts. Second, survival models would require the selection of a specific national measure (e.g., first or last) as the reference point which involves similar problems regarding the extent of transpositions. Alternatively, modeling every single national measure, we would need to classify cases where no transposition was notified at all as single measure, ongoing, "censored," cases even if these cases do not involve any transposition efforts at all. This would lead to a misclassification of cases and thus a biased estimation of the relationship between policy complexity and transposition delay.

Lastly, König and Luetgert (2009) employ a selection model to account for the fact that we can only investigate implementation delays for those observations where states notified their transposition measures. Non-notification (of any measures) does not necessarily imply that a member state was not able to transpose a directive, but could indicate that some member states seem to simply not transpose certain directives for unobserved reasons, other than but not unrelated to complexity, such as the salience or importance of a particular directive to a member state or in general. This can result in a biased sample of selected observed national transpositions for directives. Nevertheless, the high level of uncertainty surrounding this distortion and lack of a suitable exclusion restriction lead us to forego such models (Brandt & Schneider, 2007). We do, however, model the notification of measures in relation to our independent variables separately as part of our three-pronged measure of implementation performance.

5. Empirical analysis

This section presents the results of our empirical analysis along our three dependent variables of implementation performance at the directive-country level: the notification of (new) measures, transposition delay (if measures were notified) and the occurrence of infringement procedures. Turning to the descriptive statistics, presented in Table 2, first, we see that the number of national implementing measures varies greatly by member state with the Czech Republic averaging 13.5 and Greece implementing, on average, only 1.3 measures per directive. These figures do not necessarily reflect implementation performance, but are subject to member states' discretion and national legal requirements. The notification rates across member states are fairly similar with an overall average of roughly 77 percent (Table 3). In terms of pre-existing measures, some countries seem to have less need to adopt new measures—with up to roughly 22 percent of already existing measures (Czech Republic)—than others. Of the newly adopted measures, only 23 percent (Table 3) of measures are notified on-time. In terms of data quality of the transposition recorded in EUR-Lex XML notices, Great Britain, Sweden, and Greece show high levels (>10 percent) of missing notification date data. Finally, the occurrence of infringement procedures is generally low with some outliers—Italy with roughly 10 percent, France and Belgium with both roughly 9 percent—compared to an overall infringement rate of roughly 5 percent (Table 3).

Table 3 Summary statistics of the dataset

	N	NA	Type	Mean	SD	Min	Median	Max
Dependent variables								
Notification	21,494	0	Binary	0.769	0.421	0	1	1
Notification (new acts)	21,494	0	Binary	0.686	0.464	0	1	1
Impl. measure on time ratio	15,873	5,621	Cont.	0.227	0.393	0	0	1
Infringement	15,987	5,507	Binary	0.052	0.222	0	0	1
Independent variables								
Complexity (agg.)	21,494	0	Cont.	-0.007	0.685	-1.563	0.001	3.210
Capacity (WGI)	21,494	0	Cont.	1.291	0.581	-0.372	1.450	2.346
Position EU	20,509	985	Cont.	-2.306	1.617	-5.979	-2.398	4.709
Controls								
EP involvement	21,494	0	Binary	0.701	0.458	0	1	1
Exist. impl. measures ratio	15,873	5,621	Cont.	0.072	0.257	0	0	1
Amending legislation	21,494	0	Binary	0.337	0.473	0	0	1

To examine our first expectation, we analyze the relationship between policy complexity and implementation performance. Building upon the transaction cost perspective, we expect that as the complexity of EU directives increases, the costs associated with their implementation also rise. The regression models employing this variable as the dependent variable, presented in Table 4 (N1, N2), without any interactions, do not support this view. Contrary to our expectation, we observe a significant positive coefficient for directive complexity indicating that new national measures are more likely to be notified for more complex directives. At first glance, this does seem puzzling, but there are some possible explanations for this finding. First, it is possible that more complex proposals might be more salient than less complex proposals. Thus, states might be more pressured or more inclined to notify their actions correctly or to take action at all. Second, proposals with higher complexity might necessitate more national measures than those of lower complexity. With a higher number of measures it might become more likely that states notify (at least) some measures.

If we turn to the models with implementation delay as the dependent variable (D1, D2), we observe a significant negative coefficient for directive complexity. In contrast to the notification model and in line with our expectation, this indicates that increasing directive complexity is associated with a lower ratio of national implementing measures that were notified on time, that is, worse implementation performance. In terms of effect size, based on the full model (D2) if the complexity of a proposal increases by one standard deviation the ratio of on-time measures decreases by roughly 3 percentage points. Moving from the least to the most complex proposal would reduce the ratio of on-time measures by 21 percentage points, again based on the estimated effect size of D2 and as demonstrated in the predicted values in Figure 1.

Finally, the models with infringement procedures as the dependent variable (I1, I2) show a positive significant coefficient for complexity. This is again in line with our expectation that more complex directives are harder to implement correctly and thus more likely to be subject to infringement procedures. Increasing complexity by one standard deviation increases the chances of an infringement procedure by 2 percentage points. Consequently, compared to the lowest complexity proposal the highest complexity proposal in our sample has a 14 percentage point higher likelihood of leading to an infringement procedure (see also Fig. 1).

In addition to policy complexity, administrative capabilities play a crucial role in the successful implementation of policies. We propose that administrative capacity can act as a mitigating factor, potentially offsetting the negative impact of complexity on implementation performance. Governments with stronger bureaucratic capacities should be better equipped to handle the costs and challenges posed by complex policies.

The regression model employing notification as a dependent variable (N3) shows a significant positive coefficient for the interaction between complexity and state capacity (0.01) with a positive significant individual coefficient for directive complexity (0.09) and a significant negative individual coefficient for state capacity (-0.02). The predicted values, depicted in Figure 2, show that states with higher state capacity tend to notify slightly less on low complexity proposals, while notifying more with raising complexity overtaking lower capacity states.

Table 4 Regression results for notification, delay, and infringement models

	Notification				Delay				Infringement			
	N1 (LM)	N2 (LME)	N3 (LME)	N4 (LME)	D1 (LM)	D2 (LME)	D3 (LME)	D4 (LME)	I1 (LM)	I2 (LME)	I3 (LME)	I4 (LME)
(Intercept)	0.66 *** (0.01)	0.70 *** (0.03)	0.70 *** (0.03)	0.70 *** (0.03)	0.23 *** (0.01)	0.21 *** (0.02)	0.21 *** (0.02)	0.21 *** (0.02)	0.02 *** (0.00)	0.09 *** (0.01)	0.09 *** (0.01)	0.09 *** (0.01)
Complexity	0.11 ***	0.11 ***	0.09 ***	0.12 ***	-0.05 ***	-0.04 ***	-0.06 ***	-0.04 ***	0.03 ***	0.03 ***	0.04 ***	0.02 *
Anti-EU position	(0.00)	(0.02)	(0.02)	(0.02)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
	-0.09 ***	-0.04	-0.04	-0.04	0.11 ***	0.08 *	0.08 *	0.08 *	-0.09 ***	-0.04 ***	-0.04 ***	-0.04 **
State capacity	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
	0.01	-0.02 *	-0.02 *	-0.02 *	0.03 ***	0.04 ***	0.04 ***	0.04 ***	0.01 *	-0.01 **	-0.01 **	-0.01 **
Amending proposal	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
	0.02	0.02	0.02	0.02	0.04 **	0.04 **	0.04 **	0.04 **	(0.00)	-0.06 ***	-0.06 ***	-0.06 ***
EP involvement	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.01	-0.01	-0.01	-0.01
Existing measures	(0.02)	(0.02)	(0.02)	(0.02)	-0.25 ***	-0.28 ***	-0.29 ***	-0.28 ***	(0.00)	(0.01)	(0.01)	(0.01)
	0.01 **	0.01 **	0.01 **	0.01 **	(0.00)	(0.02)	(0.02)	(0.02)	(0.00)	-0.01 *	-0.01 *	-0.01 *
Complexity: state capacity	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Complexity: anti-EU position												-0.04 ** (0.01)
N	20,509	20,509	20,509	20,509	15,180	15,180	15,180	15,180	15,325	15,325	15,325	15,325
R ²	0.03	0.69	0.69	0.69	0.04	0.29	0.29	0.29	0.01	0.41	0.41	0.41
N (Directive)		998	998	998		987	987	987		918	918	918
N (Country-year)		631	631	631		603	603	603		479	479	479
AIC	26,015.33	5,589.24	5,583.72	5,590.64	14,124.24	11,761.69	11,760.56	11,762.76	-2,406.37	-7,701.93	-7,704.78	-7,707.59
BIC	26,054.97	5,660.59	5,663.00	5,669.92	14,170.01	11,837.97	11,844.46	11,846.67	-2,368.18	-7,633.19	-7,628.41	-7,631.22
R ² (fixed)		0.03	0.03	0.03		0.04	0.04	0.04		0.03	0.03	0.03

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; LME, linear mixed effects model. Note: Standard errors are heteroskedasticity robust. ordinary least squares models with robust standard errors in parentheses. MLE models include a random intercept on the directive and the country-year level.

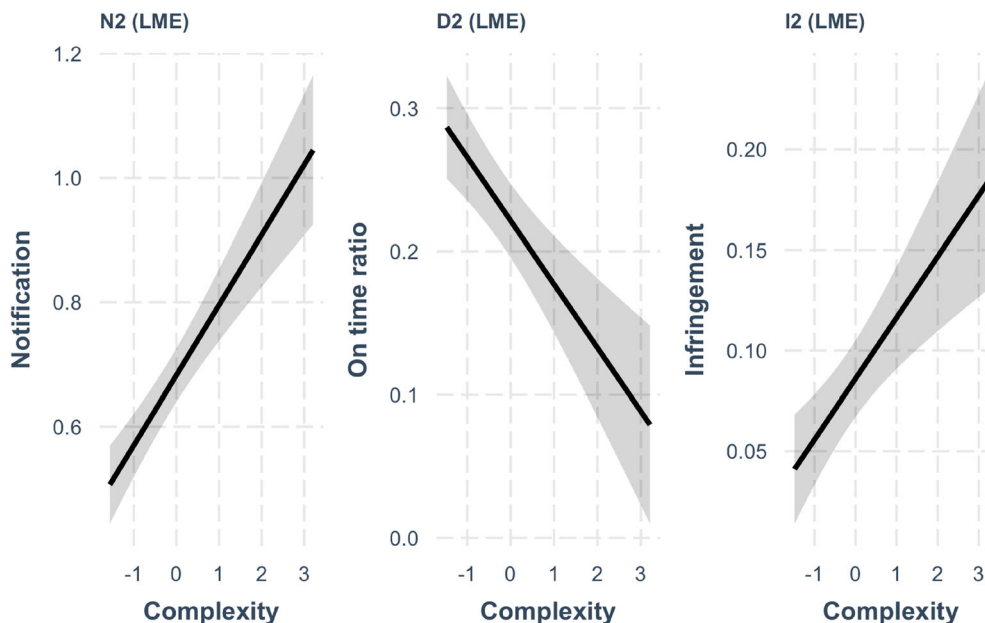


Figure 1 Predicted values based on models N2, D2, and I2 (includes only variation based on fixed effects).

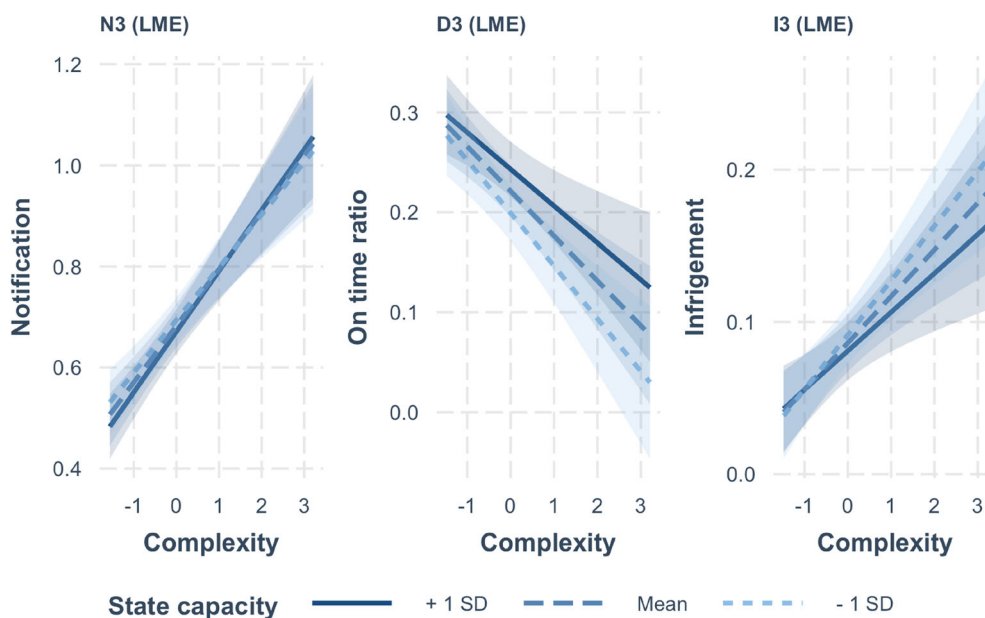


Figure 2 Predicted values for the interaction effects based on models N3, D3, and I3.

While the interaction effect is significant on the 1 percent level, the differences in the predicted values is rather small, leading to an overall very high rate of notification for high complexity proposals.

For delay, in contrast, we observe a negative significant coefficient for complexity and a positive significant coefficient for state capacity, but no significant interaction effect. Nevertheless, the predicted values in Figure 2 hint at a difference between low and high capacity states, with high capacity states being slightly less affected by rising complexity. This is indicated by the less steep slope identified for higher capacity states. Looking at the difference between the predicted values high capacity states (+1 SD) deliver around 15 percent of measures on time, while lower capacity states (-1 SD) fall below 5 percent.

Finally, the model with infringement procedures as the dependent variable (I3) shows a positive significant coefficient for complexity (0.04) and a negative significant coefficient for state capacity (-0.01) and a significant negative interaction effect (-0.01) between the two. This indicates that higher capacity indeed moderates the influence of complexity on implementation performance in terms of the likelihood of infringement procedures. The predicted values in Figure 2 show that higher capacity states (+1 SD) have a 5 percentage point lower likelihood of infringement than low capacity states (-1 SD).

Overall, the results for the interaction effects between directive complexity and state capacity for our three measures of implementation performance cautiously suggest that state capacity can indeed absorb complexity-induced costs and increase implementation performance. This is most evident from the lower predicted values for the initiation of infringement procedures for high capacity states in high complexity cases.

The political context in which policy implementation takes place can significantly affect the outcome. Governments that are supportive of European integration may exhibit a stronger commitment to implementing complex EU policies, potentially mitigating the negative effects of policy complexity. Overall, our results do not corroborate this hypothesis. We do not find significant interaction effects in the notification (N4) nor in the delay model (D4). Looking at the predicted values in Figure 3 we can identify no discernible difference for notifications, but we see that a higher anti-EU stance leads to a higher on-time ratio and lower infringement likelihood with only the latter being significant.

The latter finding goes against our hypothesized mechanism, but it connects with some recent literature (Pircher & Loxbo, 2020; e.g., Toshkov, 2019), which similarly found that the Commission tends to initiate fewer infringement proceedings against Eurosceptic governments than against Europhile governments. Our data indicate that this pattern is slightly stronger for highly complex directives, while the degree of government Euroscepticism does not seem to make a detectable difference for simpler directives. In our view, at least two different interpretations of this findings are plausible. First, the empirical pattern could indicate that the complexity of the directive serves as an informational shortcut for the Commission when deciding over which battles it should pick with Eurosceptic member states. The Commission tends to pick legal battles it is likely to win in court (König & Mäder, 2014). Losing in court against a Eurosceptic member state might be potentially more costly for the Commission because Eurosceptics have a higher political payoff from winning against the Commission. Accordingly, the Commission could prefer to initiate infringement proceedings against Eurosceptic governments when victory in court is more certain. It is possible that complexity might make it harder for the Commission to gauge its chances for success in Court and, thus, influence the pattern we observe. In fact,

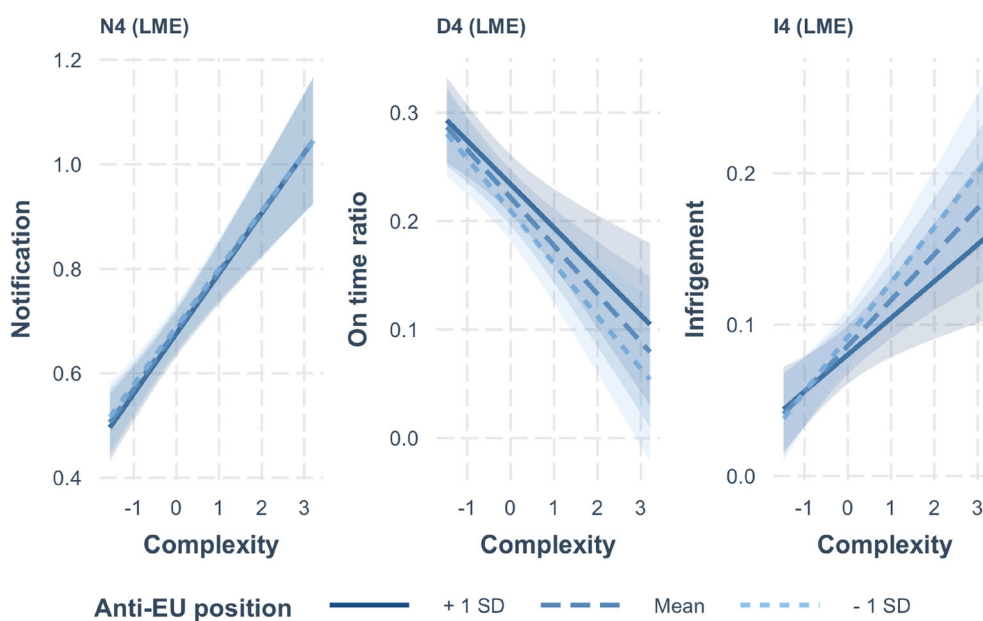


Figure 3 Predicted values for the interaction effects based on models N4, D4, and I4.

the Commission might also strategically delay infringement action against Eurosceptic governments for noncompliance with highly complex EU directives when it can reasonably expect a Europhile government to win the next national election (Cheruvu, 2022) and improve implementation performance.

Alternatively, our counterintuitive finding for H3 could imply that higher degrees of euroscepticism in government do not automatically and quickly translate into higher degrees of euroscepticism in the bureaucracy. Similarly, as Toshkov (2019) argued, highly efficient bureaucracies tend to be located in rather Eurosceptic member states, which might lead to the empirical pattern we identified.¹⁶ Ultimately, our data do not allow us to conclusively answer whether the finding is driven by strategic behavior of the Commission or by highly efficient bureaucracies clustering in highly Eurosceptic member states and we have to leave the task of testing which explanation is better to future research. The conclusions we draw heavily depend on whether infringements can be considered to measure actual noncompliance or strategic Commission behavior.

6. Conclusion

Political systems are constantly under pressure to allocate notoriously scarce resources to address increasingly complex governance challenges. The extent to which states are able to solve this difficult task is crucial for the success and effectiveness of public policies. Against this background, this study investigated the implementation performance of member states in the EU and put a special theoretical focus on the role of policy complexity, that is, the transaction costs contained in individual pieces of EU legislation.

We went beyond existing operationalizations of complexity and built a complexity indicator that is based not only on the structural aspects of a law but also on its linguistic and legal sophistication. Based on an empirical analysis of 998 EU directives adopted between 1994 and 2022, we showed that the impact of policy complexity on implementation performance varies depending on which implementation task we are interested in. When it comes to mere *notification*, EU member states even perform better when legislation becomes more complex, which is presumably either due to a higher salience of complex legislation or due to the fact that more complex legislation requires more implementation measures, which increases the probability that at least some of those measures get notified. However, more complexity is clearly and strongly associated with a greater likelihood of delayed *transposition* and also significantly increases the probability that the Commission initiates *infringement* proceedings for noncompliance. Accordingly, policy complexity critically affects states' implementation performance, which strongly underlines the importance of the European Commission's recent efforts at simplifying EU legislation in the context of its regulatory fitness and performance program (REFIT).

To some degree, states can counter the negative impact of policy complexity if they dispose of high administrative capacity. When the law to be implemented is relatively simple, states with high and low administrative capacity have a similar probability to be found in violation of the law. When complexity increases, however, administrative capacity helps to mitigate the impact of complexity on noncompliance, at least to some degree. This implies that even though the effect is small, the extent to which states decide to invest in their implementation resources makes a difference when implementation tasks increase in difficulty. Thus, our findings provide empirical support for the claim that in the EU, "preventive capacity building and rule clarification [...] reduce the risk of violations due to incapacity or inadvertence" (Tallberg, 2002, p. 632).

Our analysis of the interaction between political preferences and the costs induced by complexity found that more Eurosceptic governments are less likely to be under scrutiny for noncompliance than Europhile governments when the complexity of the implementation task increases. This counterintuitive finding resonates with some of the more recent literature on the relationship between Euroscepticism and implementation performance (Pircher & Loxbo, 2020; Toshkov, 2019). At least two interpretations of this finding are possible, in our view. First, we could suspect that if we follow the argument that the Commission picks its legal battles with the member states strategically (König & Mäder, 2014), it might not only be more hesitant to go to court against Eurosceptic member states, who have a lot to gain politically and electorally from winning against the Commission, but it might also pick court battles in more certain legal environments. As more complex laws might create more legal uncertainty for the Commission's prospects in court, it might be more inclined to forgo an infringement proceeding against a Eurosceptic member state when laws become more complex. Those findings add an important dimension to existing research on the Commission's strategic enforcement of noncompliance

(Cheruvu, 2022; König & Mäder, 2014) and its strategy of “supra-national forbearance” in the context of rising Euroscepticism (Kelemen & Pavone, 2022). Alternatively, however, the finding might also simply indicate that having a Eurosceptic government does not automatically and quickly translate into a more Eurosceptic bureaucracy. Against the background that highly efficient bureaucracies can often be found in rather Eurosceptic member states (Toshkov, 2019), the empirical pattern could indicate that bureaucracies operate rather independently from their political masters when implementing EU legislation. Eventually, the question of which interpretation has more explanatory power depends on whether we accept infringement proceedings as a valid measure of actual implementation performance or as an expression of strategic behavior by the European Commission.

This study has shown that the quality of policy implementation not only depends on states’ ability and willingness to comply but also on the difficulty of the implementation task at hand. By implication, if policy complexity is associated with worse implementation performance, then rising policy complexity will aggravate existing implementation problems. What makes this pattern particularly problematic is that policies do not only become more complex individually, but that they also accumulate over time (Adam et al., 2019). In this context of rising policy complexity and policy accumulation, states need to critically review their implementation resources and make sure that their administrative capacities are in sync with the demandingness of the tasks at hand.

Acknowledgments

This study is a research output of the EUPLEX project, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation - 407514878). We thank our three anonymous reviewers for their constructive feedback on the original version of this article. The authors also thank Annika Kolbe for excellent research assistance and all fellow panelists, discussants, and commenters at the EPSA 2023 and CES 2023 conferences for their valuable feedback and suggestions. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

Endnotes

- ¹ https://ec.europa.eu/atwork/applying-eu-law/infringements-proceedings/infringement_decisions, last accessed 9 June, 2023.
- ² <https://eur-lex.europa.eu/statistics/legislative-acts-statistics.html>, last accessed 23 October, 2023.
- ³ We would like to point out that this list of studies might not be exhaustive.
- ⁴ Warntjen (2012), however, shows that recitals and interview-based measures of salience are at best weakly correlated.
- ⁵ [https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31999Y0317\(01\)](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31999Y0317(01)).
- ⁶ <https://eur-lex.europa.eu/>.
- ⁷ See Appendix S1 for notes on data collection from EUR-Lex and our data collection choices.
- ⁸ Note that for our analysis of *notification* itself, we only consider cases where member states have notified new (previously non-existing) measures as *notified* (see also below).
- ⁹ For some cases, we observe missing notification dates in the notices which we treat as non-notified.
- ¹⁰ The BID dataset only records “proceedings launched between 1978–2017 that have reached the reasoned opinion stage by 7 March 2019” (Börzel, 2021 codebook). The reasoned opinion stage is a formal request to the member state to comply with EU law but precedes possible referral to the Court, that is, it indicates the Commission’s conviction that the member state is continuously in breach of its obligation to transpose a directive. We thus exclude all directives that have been adopted after 2016 from our analysis of infringement procedures.
- ¹¹ For a more extensive discussion of the use of infringement procedures as a measure of implementation quality, see Toshkov (2010), pp. 16–17.
- ¹² <http://info.worldbank.org/governance/wgi/>, 2022 version.
- ¹³ The following equations were produced using the equatiomatic R package (version 0.1.0.9000).
- ¹⁴ We additionally provide GLM regression models with fixed effects in Table S6 of Appendix S1.

- ¹⁵ Note that predicted probabilities in linear probability models are only reliable between the range of 0.2 and 0.8, where they are a close approximation of a logistic regression model (Long, 1997). We therefore focus our interpretation of the results on the coefficient estimates.
- ¹⁶ Denmark and Finland are prominent examples in this context.

References

- Adam, C., Hurka, S., Knill, C., & Steinebach, Y. (2019). *Policy accumulation and the democratic responsiveness trap*. Cambridge University Press.
- Baratta, R. (2014). Complexity of EU law in the domestic implementing process. *The Theory and Practice of Legislation*, 2(3), 293–308. <https://doi.org/10.5235/12050-8840.2.3.293>
- Beck, N. (2020). Estimating grouped data models with a binary-dependent variable and fixed effects via a logit versus a linear probability model: The impact of dropped units. *Political Analysis*, 28(1), 139–145. <https://doi.org/10.1017/pan.2019.20>
- Beyers, J., Dür, A., & Wonka, A. (2018). The political salience of EU policies. *Journal of European Public Policy*, 25(11), 1726–1737. <https://doi.org/10.1080/13501763.2017.1337213>
- Börzel, T. (2021). *Berlin infringement database*. Freie Universität Berlin.
- Brandt, P. T., & Schneider, C. J. (2007). *So the reviewer told you to use a selection model? Selection models and the study of international relations*.
- Cheruvu, S. (2022). When does the European Commission pursue noncompliance? *European Union Politics*, 23(3), 375–397. <https://doi.org/10.1177/14651165221087635>
- den Heijer, M., van den Abeelen, T. O., & Maslyka, A. (2019). *On the use and misuse of recitals in European Union law* (SSRN Scholarly Paper No. 3445372). <https://doi.org/10.2139/ssrn.3445372>
- Finke, D. (2018). Institutional conditions for effective parliamentary oversight of European Union politics. *West European Politics*, 42, 1–24. <https://doi.org/10.1080/01402382.2018.1474418>
- Häge, F. M. (2007). Committee decision-making in the Council of the European Union. *European Union Politics*, 8(3), 299–328. <https://doi.org/10.1177/1465116507079539>
- Haverland, M., & Romeijn, M. (2007). Do member states make European policies work? Analysing the EU transport deficit. *Public Administration*, 85, 757–778. <https://doi.org/10.1111/j.1467-9299.2007.00670.x>
- Hurka, S., & Haag, M. (2020). Policy complexity and legislative duration in the European Union. *European Union Politics*, 21(1), 87–108. <https://doi.org/10.1177/1465116519859431>
- Hurka, S., & Steinebach, Y. (2021). Legal Instrument Choice in the European Union. *JCMS: Journal of Common Market Studies*, 59(2), 278–296. <https://doi.org/10.1111/jcms.13068>
- Hurka, S., Haag, M., & Kaplaner, C. (2022). Policy complexity in the European Union, 1993–today: Introducing the EUPLEX dataset. *Journal of European Public Policy*, 29(9), 1512–1527. <https://doi.org/10.1080/13501763.2021.1938174>
- Hurka, S. (2023). The institutional and political roots of complex policies: Evidence from the European Union. *European Journal of Political Research*, 62(4), 1168–1190. <https://doi.org/10.1111/1475-6765.12555>
- Hurka, S., Haag, M., & Kaplaner, C. (2023). Proposal complexity and report allocation in the European Parliament. *European Union Politics*, 24(2), 307–326. <https://doi.org/10.1177/14651165221144888>
- Kaeding, M. (2006). Determinants of transposition delay in the European Union. *Journal of Public Policy*, 26(3), 229–253. <https://doi.org/10.1017/S0143814X06000547>
- Katz, D. M., & Bommarito, M., II. (2014). Measuring the complexity of the law: The United States code. *Artificial Intelligence Law*, 22(4), 337–374.
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). *The worldwide governance indicators: Methodology and analytical issues* (SSRN Scholarly Paper No. 1682130).
- Kelemen, R. D., & Pavone, T. (2022). *Where have the guardians gone? Law enforcement and the politics of supranational forbearance in the European Union* (SSRN Scholarly Paper No. 4188118). <https://doi.org/10.2139/ssrn.3994918>
- Knill, C., Steinebach, Y., & Zink, D. (2023). How policy growth affects policy implementation: Bureaucratic overload and policy triage. *Journal of European Public Policy*, 31(2), 324–351. <https://doi.org/10.1080/13501763.2022.2158208>
- Koniaris, M., Anagnostopoulos, I., & Vassiliou, Y. (2018). Network analysis in the legal domain: A complex model for European Union legal sources. *Journal of Complex Networks*, 6(2), 243–268.
- König, T., & Luetgert, B. (2009). Troubles with transposition? Explaining trends in member-state notification and the delayed transposition of EU directives. *British Journal of Political Science*, 39(1), 163–194. <https://doi.org/10.1017/S0007123408000380>
- König, T., & Mäder, L. (2014). The strategic nature of compliance: An empirical evaluation of law implementation in the central monitoring system of the European Union. *American Journal of Political Science*, 58, 246–263. <https://doi.org/10.1111/ajps.12038>
- Lehmann, P., Franzmann, S., Burst, T., Regel, S., Riethmüller, F., Volkens, A., Weßels, B., & Zehnter, L. (2023). The manifesto data collection. In *Manifesto project (MRG/CMP/MARPOR)*. Berlin für Sozialforschung/Göttinger Institut für Demokratieforschung. <https://doi.org/10.25522/manifesto.mpd.2023a>
- Limberg, J., Steinebach, Y., Bayerlein, L., & Knill, C. (2021). The more the better? Rule growth and policy impact from a macro perspective. *European Journal of Political Research*, 60(2), 438–454. <https://doi.org/10.1111/1475-6765.12406>
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Sage Publications.
- Lowe, W., Benoit, K., Mikhaylov, S., & Laver, M. (2011). Scaling policy preferences from coded political texts. *Legislative Studies Quarterly*, 36(1), 123–155. <https://doi.org/10.1111/j.1939-9162.2010.00006.x>

- Luetgert, B., & Dannwolf, T. (2009). Mixing methods: A nested analysis of EU member state transposition patterns. *European Union Politics*, 10(3), 307–334. <https://doi.org/10.1177/1465116509337772>
- Marks, G., Hooghe, L., Steenbergen, M. R., & Bakker, R. (2007). Crossvalidating data on party positioning on European integration. *Electoral Studies*, 26(1), 23–38. <https://doi.org/10.1016/j.electstud.2006.03.007>
- Mastenbroek, E. (2003). Surviving the deadline: The transposition of EU directives in the Netherlands. *European Union Politics*, 4(4), 371–395. <https://doi.org/10.1177/146511650344001>
- Paasch, J. (2022). Revisiting policy preferences and capacities in the EU: Multi-level policy implementation in the subnational authorities. *JCMS: Journal of Common Market Studies*, 60(3), 783–800. <https://doi.org/10.1111/jcms.13286>
- Paasch, J., & Stecker, C. (2021). When Europe hits the subnational authorities: The transposition of EU directives in Germany between 1990 and 2018. *Journal of Public Policy*, 41(4), 798–817. <https://doi.org/10.1017/S0143814X20000276>
- Pircher, B., & Loxbo, K. (2020). Compliance with EU law in times of disintegration: Exploring changes in transposition and enforcement in the EU member states between 1997 and 2016. *JCMS: Journal of Common Market Studies*, 58(5), 1270–1287. <https://doi.org/10.1111/jcms.13030>
- Pressman, J. L., & Wildavsky, A. B. (1984). *Implementation: How great expectations in Washington are dashed in Oakland* (3rd ed.). University of California Press.
- Rauh, C., Bes, B. J., & Schoonvelde, M. (2020). Undermining, defusing or defending European integration? Assessing public communication of European executives in times of EU politicisation. *European Journal of Political Research*, 59(2), 397–423.
- Sabatier, P., & Mazmanian, D. (1980). The implementation of public policy: A framework of analysis. *Policy Studies Journal*, 8(4), 538–560. <https://doi.org/10.1111/j.1541-0072.1980.tb01266.x>
- Schuck, P. H. (1992). Legal complexity: Some causes, consequences, and cures. *Duke Law Journal*, 42(1), 1–52.
- Senninger, R. (2023). What makes policy complex? *Political Science Research and Methods*, 11(4), 913–920.
- Steuenberg, B., & Kaeding, M. (2009). “As time goes by”: Explaining the transposition of maritime directives. *European Journal of Political Research*, 48(3), 432–454. <https://doi.org/10.1111/j.1475-6765.2009.01832.x>
- Steuenberg, B., & Rhinard, M. (2010). The transposition of European law in EU member states: Between process and politics. *European Political Science Review*, 2, 495–520. <https://doi.org/10.1017/S1755773910000196>
- Tallberg, J. (2002). Paths to compliance: Enforcement, management, and the European Union. *International Organization*, 56(3), 609–643. <https://doi.org/10.1162/002081802760199908>
- Thomas, M. (2013). The relationship between national elections and the delay in transposition of E.U. Directives. *Politics & Policy*, 41(6), 911–946. <https://doi.org/10.1111/polp.12052>
- Thomson, R., Torenvlied, R., & Judge, A. (2020). Incentives to comply: The impact of national governments’ and stakeholders’ preferences on compliance with EU laws. *Journal of European Public Policy*, 27(12), 1801–1818. <https://doi.org/10.1080/13501763.2019.1699593>
- Toshkov, D. (2010, February). *Taking stock: A review of quantitative studies of transposition and implementation of EU law*. Working Papers of the Vienna Institute for European integration research, Institute for European integration research.
- Toshkov, D. (2011). Public opinion and policy output in the European Union: A lost relationship. *European Union Politics*, 12(2), 169–191. <https://doi.org/10.1177/1465116510395043>
- Toshkov, D. (2019). *Does Euroscepticism influence compliance and enforcement of EU law in the member states?* Hart Publishing. <https://doi.org/10.5040/9781509927685.ch-002>
- Treib, O. (2014). Implementing and complying with EU governance outputs. *Living Reviews in European Governance*, 9(1), 1–33. <https://doi.org/10.12942/lreg-2014-1>
- van Elsas, E., & van der Brug, W. (2015). The changing relationship between leftright ideology and euroscepticism, 1973. *European Union Politics*, 16(2), 194–215. <https://doi.org/10.1177/1465116514562918>
- Warntjen, A. (2012). Measuring salience in EU legislative politics. *European Union Politics*, 13(1), 168–182. <https://doi.org/10.1177/1465116511428495>
- Williams, C. J. (2018). Responding through transposition: Public Euroscepticism and European policy implementation. *European Political Science Review*, 10(1), 51–70. <https://doi.org/10.1017/S1755773916000187>
- Wratil, C., Waeckerle, J., & Proksch, S. O. (2023). Government rhetoric and the representation of public opinion in international negotiations. *American Political Science Review*, 117(3), 1105–1122.
- Zhelyazkova, A. (2013). Complying with EU directives’ requirements: The link between EU decision-making and the correct transposition of EU provisions. *Journal of European Public Policy*, 20(5), 702–721. <https://doi.org/10.1080/13501763.2012.736728>
- Zhelyazkova, A., Kaya, C., & Schrama, R. (2017). Notified and substantive compliance with EU law in enlarged Europe: Evidence from four policy areas. *Journal of European Public Policy*, 24(2), 216–238. <https://doi.org/10.1080/13501763.2016.1264084>
- Zhelyazkova, A., & Thomann, E. (2022). “I did it my way”: Customisation and practical compliance with EU policies. *Journal of European Public Policy*, 29(3), 427–447. <https://doi.org/10.1080/13501763.2020.1859599>
- Zhelyazkova, A., & Yordanova, N. (2015). Signalling “compliance”: The link between notified EU directive implementation and infringement cases. *European Union Politics*, 16(3), 408–428. <https://doi.org/10.1177/1465116515576394>

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site:

Appendix S1 Supporting information.