

This article was downloaded by: [UB Mannheim]

On: 19 September 2012, At: 12:07

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954

Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of European Public Policy

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/rjpp20>

HOW ROBUST ARE COMPLIANCE FINDINGS? A RESEARCH SYNTHESIS

Mariyana Angelova, Tanja Dannwolf & Thomas König

Version of record first published: 24 Jul 2012.

To cite this article: Mariyana Angelova, Tanja Dannwolf & Thomas König (2012): HOW ROBUST ARE COMPLIANCE FINDINGS? A RESEARCH SYNTHESIS, *Journal of European Public Policy*, 19:8, 1269-1291

To link to this article: <http://dx.doi.org/10.1080/13501763.2012.705051>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Research Agenda Section

Edited by Berthold Rittberger

HOW ROBUST ARE COMPLIANCE FINDINGS? A RESEARCH SYNTHESIS

Mariyana Angelova, Tanja Dannwolf and
Thomas König

ABSTRACT This study presents a synthesis of quantitative and qualitative research on compliance with EU directives. We identify and code 12 theoretical arguments tested in 37 published compliance studies and evaluate the robustness and representativeness of their findings. Our synthesis reveals robust findings for the ‘goodness-of-fit’ and ‘institutional decision-making’ arguments, while results on ‘actors’ policy preferences’ and ‘administrative efficiency’ remain ambiguous. A closer examination of the studies’ research design suggests policy and country selection effects. Specifically, most studies focus on environmental and social policies and rarely include complying Scandinavian states and non-complying Southern states. We therefore recommend a cautious interpretation of existing compliance findings and, for future compliance research, a more careful selection of countries and policy fields.

KEY WORDS Compliance; European Union; research synthesis.

Compliance with European Union (EU) directives is a central issue in the analysis of European integration. Directives are binding and enforceable legislative acts of the European Union, often adopted with supermajorities after long deliberations in the Council of Ministers and, more increasingly, with the support of the European Parliament. Member states are obliged to transpose directives into domestic law, but are given discretion on the instruments they can use for the implementation of a directive’s goals. This discretion is raising concerns about possible non-compliant behaviour that has been studied by agencies and scientists in a variety of disciplines (Börzel 2001; Falkner *et al.* 2005; Haas 1998; Hartlapp and Falkner 2009; Haverland 2000; Héritier *et al.* 2001; Knill and Lenschow 1998; Krislov *et al.* 1986; Mastenbroek 2005; Mbaye 2001; Siedentopf and Ziller 1988; Steuernberg 2006; Tallberg 2002; Toshkov 2010; Toshkov *et al.* 2010; Treib 2003, 2008; Weiler 1991). From a methodological point of view, the fact that a growing number of member states with different political and economic backgrounds

must transpose the same directives into domestic law provides us with an excellent laboratory to study (non-)compliance across policy fields and countries.

However, existing reviews of compliance studies provide a mixed picture with sometimes contradictory conclusions on the factors promoting (non-)compliance. For example, Mastenbroek (2005: 1108) points to the inconclusiveness of the results with respect to the main legal, administrative and political variables. She states that the prominent goodness-of-fit hypothesis is 'neither a necessary nor a sufficient condition for smooth compliance, and vice versa' (Mastenbroek 2005: 1109).¹ Treib (2008: 11) criticizes the mixed picture on the impact of veto players and support for European integration and reports that 'the only factor that seems to find support in most quantitative analyses so far is various aspects of administrative capabilities'.² Toshkov (2010) concludes that administrative efficiency, parliamentary scrutiny and co-ordination strength have a positive influence, whereas federalism, corruption, number of veto players, number of ministries and the existence of domestic conflict all have a negative impact on compliance. In a separate review of qualitative studies, Toshkov *et al.* (2010) find little support for the goodness-of-fit hypothesis, but highlight the importance of administrative capacity, co-ordination and domestic level factors – such as domestic conflict between veto players – for the level of compliance. Although these reviews summarize a variety of compliance studies, it remains an open question whether their insights are robust – especially if we take into consideration the large variety of methodologies and research designs used in the extant literature.

To answer this question we evaluate compliance findings of qualitative and quantitative studies by a research synthesis, which assesses robustness of findings and controls for research design. In the first step of our synthesis we identify the most prominent theoretical arguments to which we assign the corresponding explanatory factors reported in the selected studies.³ This allows us to compare compliance findings in a systematic manner across all types of studies on transposition, infringement and application with directives, and to control for different methodological approaches and research designs. We secondly control for research design regarding the selection of countries and policy fields where we find considerable variation. This variation might not only explain why existing reviews of compliance studies provide a mixed picture with sometimes contradictory conclusions, but it also suggests that findings of compliance research risk a bias by non-random case selection. This risk exists when the samples include only those policy fields and/or countries with characteristics prompting non-compliance. To examine this risk we relate the selection of policy fields and countries to the relative level of compliance by the corresponding infringement records.

Methodologically, we believe that our research synthesis moves beyond earlier reviews of compliance studies in several respects. First, we are able to compare the findings regarding a large number of explanatory variables across both qualitative and quantitative studies with different research designs. Second, to reduce coding errors we use well-defined guidelines that help us to organize the vast

amount of empirical work on compliance in the EU. Third, to ensure the cross-validity of our classification of the reported findings on compliance we code them independently twice and compare the results. Finally, we examine the robustness of the results across compliance studies statistically, while controlling for representativeness by research design.

More substantially, we find robust evidence supporting the goodness-of-fit and the institutional decision-making capacity argument. This suggests that the similarity between goals of directives and existing national regulations enhances compliance, and that institutionally restricted governments have a lower likelihood of compliance. These findings are confirmed by qualitative and quantitative studies. The effect of the prominent arguments on administrative efficiency and governments' preferences is less consistent. The closer inspection of the research design of these studies reveals that most case studies explore the fields of environmental and social policies. This policy field bias, we find, only includes a minor share of directives, which are characterized by a high number of infringements. Compliance scholars also tend to exclude well-complying Scandinavian and poorly complying Southern member states from their samples. This kind of country selection is also problematic for a representative overview and can be responsible for the lack of robust findings of frequently studied theoretical arguments on compliance.

SAMPLE SELECTION, CODING AND INTER-CODER RELIABILITY

For our analysis of the robustness of compliance findings we decided to consider both qualitative and quantitative studies (Hunter and Schmidt 2004).⁴ Reviews which separate quantitative from qualitative compliance studies have been released in two recent working papers of Toshkov (2010) and Toshkov *et al.* (2010) – neither of which accounts for the statistical robustness of findings nor controls for the selection of policy fields and countries. Furthermore, Toshkov and colleagues counted the effects of similar *measurements*, which makes it difficult to draw conclusions about the reasons for non-compliance and their robustness.⁵ Therefore, to compare the results across different designs and methods of analysis we assess the robustness of reported findings at their *theoretical level* for both qualitative and quantitative studies and control for the selection aspects of the research design. However, the methodological diversity of compliance studies and often missing statistical information impedes the conduction of a meta-analysis with overall effect estimations. As a consequence, our synthesis uses the less sophisticated vote counting method to compare the reported findings.⁶

A critical point of any research synthesis concerns the sampling of studies. Because several findings of the same study are reported multiple times in working papers, conference papers, journal articles and book chapters, we limited our sampling to international scholarly journal articles. As recommended by Wilson (2009), we developed a coding scheme before storing our data in a relational database. This database contains all results, which we

extracted from compliance research following strict coding instructions and controlling for inter-coder reliability. This part of the procedure makes it possible to compare the explanatory consistency of the theories on compliance, while controlling for each study's research design, in particular for countries and policy fields selected for investigation.

Our final sample comprises 37 studies published in scholarly journals (see references). We started with the studies listed in the literature review of Mastenbroek (2005), which we extended by searching in the Thomson Reuters Social Sciences Citation Index (SSCI) database⁷ for articles on compliance of national law with EU directives published in the period between 1996 and 2009. To ensure reliability and to avoid over-representation we include only articles published in international peer-reviewed journals because such studies underwent external quality checks.⁸ Furthermore, we offer our coding scheme and data for replication and future expansion purposes. Please note that we do not claim that our sample comprises the universe of studies on the EU or compliance. For example, we eliminated studies that did not evaluate the explanatory power of theoretical arguments on compliance.⁹ Our sample is also restricted by our focus on transposition and implementation of directives, which excludes studies with a broader focus on Europeanization or on general law enforcement. Owing to the difficulty of coding the impact of interaction effects,¹⁰ we further excluded studies where all explanatory variables constituted an interaction.¹¹

For the remaining studies we collected information on the reported direction of effects as well as the key characteristics of their research design, including methodology, sample size and coverage with respect to time period, policy fields and countries. Our evaluation relates to 12 broader categories of major theoretical arguments, which have been frequently referred to in the compliance literature (e.g., Börzel 2003; Falkner *et al.* 2005; Tallberg 2002).¹² Following the authors' theoretical reasoning, each explanatory factor from the selected studies was assigned to one of the 12 general theoretical arguments. To give an example for our coding procedure, a common argument found in many studies is that high institutionally driven decision-making capacity of governments has a positive effect on compliance. We summarize the effect of several indicators of institutionally induced decision-making capacity, such as federalism, number of veto players and the degree of centralization. Finally, we adjust the direction of effects so that a positive effect always means a confirmation of the theoretical argument regarding a higher likelihood of compliance. For instance, we code the effect of a higher level of federalism – which introduces additional veto-points – as decreasing institutional decision-making capacity of governments. Hence, we count a positive effect of federalism as a negative effect of institutional decision-making capacity on compliance. Missing causal interpretations in qualitative studies, complex interaction effects in regression analysis and categorical variables are coded as an unclear effect.

A major challenge of our synthesis is the reliable classification of the variables, findings and research designs. To evaluate inter-coder reliability, we coded all

studies and findings twice and independently as recommended by Krippendorff (2004). Differently coded descriptive information and findings were resolved by deliberation between the two coders or otherwise excluded. The Kappa inter-coder reliability measure for the dependent variable is Kappa 0.78. Kappa is only a little lower for the 12 theoretical accounts and their operationalization with values of 0.64 and 0.69 respectively (Orwin 1994: 152).¹³ Given the variation of studies in compliance research, we believe that these numbers represent high inter-coder reliability measures.

To evaluate robustness of the reported findings with respect to each theoretical argument we apply a binomial test of the proportion of positive results with Wilson confidence intervals, which are recommended for evaluating heterogeneous studies of our sample size (Bushman and Wang 2009; Deeks *et al.* 2008).¹⁴ The null hypothesis is that the probability of finding a positive effect is 0.5, which corresponds to an equal proportion of positive and negative results. A probability of 0.5 suggests that finding a positive or negative effect of a given theoretical argument is random, which indicates a lack of consistency or lack of robustness in support of any given theoretical argument across studies. To diminish type II errors – a false rejection of existing effects – our analysis only takes into account the direction of the effects without considering their significance level (for more details, see Hunter and Schmidt [2004]).¹⁵

TRANSPPOSITION, INFRINGEMENT AND APPLICATION OF DIRECTIVES

According to Hartlapp and Falkner (2009: 285), ‘one of the reasons for disparate findings on compliance in the EU lies in the differing and [...] often unclear conceptualization of the dependent variable: compliance’. Even if compliance studies focus on directives, they investigate different phenomena, namely infringement proceedings, application and transposition records. In our view, this variation in the dependent variable does not necessarily explain the mixed findings, but it draws our attention to the data generating process: data on infringement proceedings reflect the Commission’s evaluation of a suspect non-complying member state, while transposition records, on the other hand, are notifications of a member state’s transposition activities without indicating whether these activities have correctly and completely implemented a directive (Luetgert and Dannwolf 2009). This suggests that statistics on infringement and transposition cases reflect the two sides of the same (strategic) coin of (non-)compliance, which we include in our analysis.

For quantitative transposition studies, a common source of information is the Commission’s database on the activities of the member states, which must notify their measures for transposing directives into national law.¹⁶ According to Article 288 (249) EC Treaty member states must transpose directives correctly and on time. Many transposition studies separate timely transposition from correct and only investigate timeliness with respect to the deadline for notification.¹⁷ Other studies investigate practical application of directives

relying primarily on annual Commission reports, interviews and secondary sources of information about application. Finally, data from infringement proceedings of the Commission distinguishes between (three) different types of non-compliance: non-transposition; non-conformity; and bad application.¹⁸

Instead of separating each (sub-)type of compliance analysis, our sample includes these various views on compliance ranging from activities on timely to correct transposition to the application of the directives' policy goals and their enforcement by infringement proceedings. More than half of the 37 studies of our sample analyse transposition (56.8 per cent), followed by application (29.7 per cent) and infringement proceedings (13.5 per cent). Among the 21 transposition studies, many investigate the duration of the transposition process, and four out of five studies which analyse correct transposition are of a qualitative nature. Four studies define duration by notification of the first transposition instrument, whereas five studies concentrate on delay. The five infringement studies often use annual statistics on infringement proceedings. A few qualitative studies gathered information on application by expert interviews (Duina and Blithe 1999; Falkner and Treib 2008; Falkner *et al.* 2004) or applied content analysis to Commission documents (Hille and Knill 2006; Jensen 2007).

THEORETICAL ARGUMENTS AND OPERATIONALIZATION(S)

The 37 studies of our sample conducted 209 empirical tests of various theoretical arguments. To evaluate the empirical implications of these theoretical accounts under control of research design, we relate their findings to 12 main arguments, which were operationalized using 82 variables.¹⁹ Table 1 provides an overview of the number of studies that test these arguments, further broken down by method and dependent variable.

One of the most prominent explanatory variables is *goodness-of-fit*, according to which non-compliance increases with the difference between the domestic status quo and the policy goals of a directive. Our sample shows that 11 out of 17 qualitative and 9 out of 20 quantitative studies evaluate the goodness-of-fit argument. These nine quantitative studies focus exclusively on explaining transposition decisions by goodness-of-fit, while qualitative studies also investigate the impact of goodness-of-fit on application. Empirically, goodness-of-fit is approximated by using very different variables, such as a directive's correspondence either with national policy legacies (Duina and Blithe 1999), interest group structures (Duina 1997), policies (Börzel 2000) or transposition-related financial consequences (Falkner *et al.* 2005).

Preferences and *monitoring* are main arguments on rational compliance from the enforcement school of thought (Tallberg 2002). To mitigate rational non-compliance by a member state, the Commission and/or national agencies attempt to detect non-compliance and eventually establish sanctioning costs to a non-complying member state by starting an infringement proceeding. The role of *preferences* often refers to the support of a member state for the adoption of a directive, which is conceived as a necessary condition for compliance.

Table 1 Explanatory factors for measures of (non-)compliance by qualitative and quantitative studies

<i>Theoretical argument</i>	<i>Number of qualitative studies (C)</i>				<i>Number of quantitative studies (Q)</i>				<i>C and Q total</i>
	<i>Transposition</i>	<i>Infringements</i>	<i>Application</i>	<i>Total</i>	<i>Transposition</i>	<i>Infringements</i>	<i>Application</i>	<i>Total</i>	
Goodness of fit (GOF)	4	–	7	11	9	–	–	9	20
Preference fit (PREF)	2	1	4	7	10	2	1	13	20
Supranational monitoring and enforcement (SME)	1	–	3	4	4	1	–	5	9
National monitoring and enforcement (NME)	–	–	4	4	–	1	–	1	5
Power (EU) (POW)	–	–	–	–	–	4	–	4	4
Low complexity of EU laws (COMPL)	–	1	–	1	10	–	–	10	11
Institutional decision-making capacity (IDMC)	4	–	2	6	13	4	1	18	24
Few interministerial co-ordination problems (COORD)	2	1	1	4	5	–	–	5	9
Administrative efficiency (AE)	2	1	4	7	8	2	1	11	18
Learning (LEARN)	–	–	1	1	1	2	–	3	4
Favourable culture (CULT)	2	–	–	2	6	2	–	8	10
Positive public support (PS)	–	–	2	2	2	1	–	3	5
Control variable	–	–	–	–	1	1	1	3	3
No appropriate category	1	–	1	2	3	1	–	4	6
Total	6	1	10	17	15	4	1	20	37

Out of 13 quantitative studies, 10 investigate the impact of preferences on transposition, while four out of seven qualitative studies also evaluate its effect on application.

Compared to preferences, *monitoring* has received less empirical attention, whereby supranational monitoring should be distinguished from national activities. (Supranational) Monitoring by the Commission is a crucial part of the centralized compliance system in the EU (Tallberg 2002: 614), which is studied by nine articles. In contrast, national monitoring points to fire-alarm mechanisms that alleviate compliance, such as access to courts (Börzel 2006) and interest group activities (Börzel 2000). There are only five studies – four of which are case studies – which analyse this argument and explain application.

Furthermore, compliance studies frequently point to the *complexity* of directives, which is measured by various directive-specific characteristics – including text length, discretion and time allotted for transposition. Kaeding (2008) and Toshkov (2008) find that complexity has a negative impact on compliance – owing to either discretion or an increased workload. About a quarter of all studies and a half of the large N studies from our sample analyse and emphasize this argument using transposition as the explanandum.

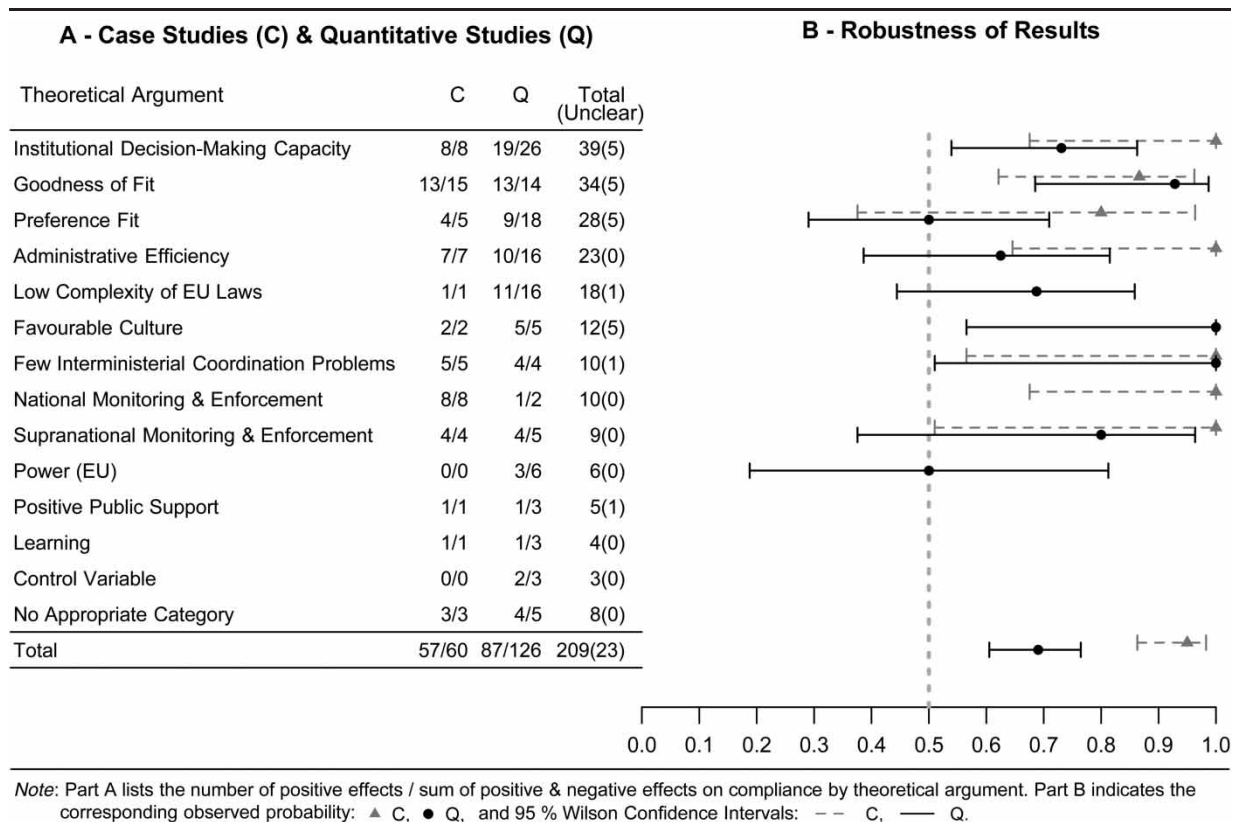
According to Mastenbroek and Kaeding (2006) recent research pays more attention to domestic politics and *institutional decision-making constraints*. The number of veto players, federalism and the effective number of parties measure institutional decision-making capacity in 24 articles, including almost all quantitative and one-third of qualitative studies. Furthermore, van den Bossche (1996) and Mastenbroek (2003) argue that *inter-ministerial co-ordination* can delay transposition, which is corroborated by a quarter of quantitative and qualitative studies. In the same vein, Berglund *et al.* (2006) and Mbaye (2001) point to the impact of *administrative efficiency* arguing that inefficient bureaucracies are more prone to private interests (Borghetto *et al.* 2006). Nearly half of all selected studies – predominantly quantitative – consider this argument.

Scholars also emphasize the importance of *learning* effects, which are expected to enhance the capacities of member states to comply with directives. For example, Berglund *et al.* (2006: 701) argue that a rule system needs time to develop, and Dimitrakopoulos (2001: 605) states that domestic administrations learn over time by interacting with the Commission. Similarly, Sedelmeier (2008) refers to a socialization of élites with increasing acceptance of European norms over time. These arguments are investigated only in four studies – three quantitative and one qualitative.

Another relevant variable in this literature is *culture* in terms of satisfaction with democracy, common norms, rule of law (Berglund *et al.* 2006) and cultural approaches on conflict resolution (Sverdrup 2004). Cultural effects are investigated in 10 studies, including Falkner *et al.*'s (2005) typology of three worlds of compliance that they also apply to Eastern countries in a more recent study (Falkner and Treib 2008).

Table 2 summarizes the effects of these arguments for qualitative and quantitative studies, indicating the number of positive effects on compliance out of

Table 2 Number of positive effects on compliance by theoretical argument



all reported effects. Table 2A lists the number of confirming findings for qualitative and quantitative studies and the total number of tests, while table 2B shows the respective point estimator of the share of confirming findings along with their confidence intervals.²⁰ When the rejection of the null hypothesis is not possible, we conduct a power test to check whether our sample is large enough to generate statistical evidence.²¹

Overall, quantitative studies find evidence for 87 out of 126 reported tests, while case studies confirm their theoretical expectations in 57 out of 60. Despite their large variation and in contrast to Toshkov *et al.* (2010), we find robust findings for two theoretical accounts of compliance with directives – institutional decision-making capacity of governments and goodness-of-fit – which reveal robustness across both qualitative and quantitative studies. Inter-ministerial co-ordination effects are also robust for all studies, but their small number leaves us cautious with the interpretation of the empirical implications of this argument. The same is true for culture, although the expected effect is confirmed in all qualitative and quantitative studies.²²

The overall confirmation of the positive impact of institutional decision-making capacity and goodness-of-fit suggests that these effects hold across different operationalization. A good fit facilitates member states' compliance with directives. Implementation problems can thus be handled with capacity enhancing strategies. Similarly, fewer veto players with diverse policy preferences constrain to a lesser extent decision-makers in the smooth transposition and application of directives. However, the institutional decision-making capacity of a country is often persistent, which makes it difficult to promote compliance. Another capacity-related argument – administrative efficiency – is robust only across case studies, where the expected positive effect on compliance is always confirmed. Similarly, the effect of national and supranational monitoring and enforcement is robust only for case studies. In contrast, we find mixed results for the same arguments in quantitative studies. The power test indicates that additional empirical tests are needed if we want to draw any further conclusions about the effect of administrative efficiency and enforcement-related variables across studies.

The results regarding the effects of preference fit, complexity and power at the European level are mixed in quantitative studies. For these arguments, case studies either find mixed results with a power test below 0.8 or a robustness check was not possible given the low number of reported tests. Whether, how and when member states' policy preferences – and thus willingness to comply – matter for the transposition and application processes therefore requires further empirical investigation. The same holds true for the impact of member states' power at the European level, public support and learning on compliance, which were investigated less than six times either qualitatively or quantitatively.

Although some theoretical arguments have robust effects on compliance across studies, we still need to control for the research design of those studies, i.e., whether and to what extent they are 'representative' for policy fields and countries. The control for research design is also necessary, because the

likelihood for compliance might have changed over time owing to the accession of countries and legislative productivity of directives in different policy fields, which were integrated by several treaty revisions. Across both qualitative and quantitative compliance studies, institutional decision-making capacity of governments and goodness-of-fit have shown robust effects. Administrative efficiency, as well as national monitoring and enforcement, are robust only across qualitative studies. Because these findings are derived from studies which not only differ in methodology but also in research design, we continue with a close inspection of their coverage of policies and countries.

SELECTING POLICY FIELDS AND COUNTRIES

In the past 20 years the membership size has more than doubled, going from 12 to 27 countries, and an increasing number of policy competences over distinct policy fields have been transferred to the EU level by treaty revisions. The empirical effect of this expansion in scope and scale is indicated by the number of directives that entered into force between 1989 and 2005 in each policy field, which old and new member states were obliged to transpose into domestic law. Table 3 shows the field-specific distribution of directives, the corresponding potential and effective level of non-compliance measured by the number of infringement proceedings.²³ It also indicates the number of qualitative and quantitative compliance studies that investigated these policy fields using four widely tested (and robust) arguments.

First and foremost, it is worth noting that directives are not uniformly distributed across policy fields. Almost a third (31.5 per cent) of directives deals with agricultural policies, followed by a similar share (30.9 per cent) in the field of industrial & internal market policies, which means that about two-thirds of all directives are devoted to only two policy fields. These are followed by transport (9.1 per cent) and environment policies (7 per cent). The two dominant fields show an average rate of non-compliance by non-notification. However, only 8 out of 37 studies consider one or both of these major policy fields in their compliance analyses. By contrast, 22 (or 60 per cent) of compliance studies focus on the fields of environmental and/or social policies, which only account for 6.4 per cent of all directives. Compared to the average rate of non-compliance in the two dominant fields, environmental and social policies reveal one of the highest levels of non-compliance.

This focus on environmental and social policies, which are characterized by a low number of directives with many infringement proceedings, is mainly caused by case studies, while most quantitative studies do not differentiate between policy fields (indicated by 'no differentiation' in Table 3). This selection bias on few directives with high non-compliance rates implies a very field-specific interpretation of their findings. Apart from this field-specific selection of case studies, we generally note that compliance studies rarely include policy fields with a high share of directives and infringement proceedings, such as the right of establishment and freedom to provide services, transport, agricultural

Table 3 Legislative output and infringements of directives/number of studies analysing robust factors by policy field

EU Policy Areas	N directives (%)	Non-notification %	Bad application %	Violative opportunity	Number of case studies (C)					Number of quantitative studies (Q)					C and Q All
					IDMC	GOF	AE	NME	All	IDMC	GOF	AE	NME	All	
AGRICULTURE	528 (31.5)	15.7	0.4	7272	–	1	–	–	1	3	2	2	–	4	5
INTERNAL MARKET	518 (30.9)	14.2	0.7	7272	1	1	–	–	2	2	2	1	–	3	5
TRANSPORT	153 (9.1)	18.9	1.3	2166	–	–	–	–	–	4	2	1	–	5	5
ENVIRONMENT	118 (7.0)	22.4	6.1	1653	3	7	2	3	8	3	2	1	–	4	12
SERVICE	108 (6.4)	22.4	4.9	1482	–	–	1	–	1	2	1	2	–	2	3
LABOUR & SOCIAL	87 (5.2)	23.5	2.5	1215	2	4	3	1	6	8	6	5	1	9	15
TAXATION	62 (3.7)	4.7	2.0	852	–	–	–	–	–	1	–	–	–	1	1
ENERGY	27 (1.6)	14.9	1.9	375	–	–	–	–	–	2	1	–	–	3	3
UNDERTAKINGS	19 (1.1)	30.6	3.5	255	–	–	–	–	–	2	1	1	–	2	2
JUSTICE	17 (1.0)	33.7	0.0	255	–	–	–	–	–	1	2	1	–	2	2
COMPETITION	15 (0.9)	11.9	3.5	201	–	–	–	–	–	–	–	–	–	–	–
CUSTOMS	11 (0.7)	12.4	0.0	153	–	–	–	–	–	1	–	–	–	1	1
FINANCE	5 (0.3)	17.4	1.4	69	–	–	–	–	–	–	1	–	–	1	1
ECONOMY	3 (0.2)	2.4	0.0	42	–	–	–	–	–	–	1	–	–	1	1
PEOPLE'S EUROPE	3 (0.2)	16.7	19.4	36	–	–	–	–	–	–	–	–	–	–	–
EXTERNAL RELATIONS	1 (0.06)	6.7	0.0	15	–	–	–	–	–	–	–	–	–	–	–
NO DIFFERENTIATION					1	–	1	–	1	8	3	5	–	9	10
Average rates	104.7	16.8	3	1454	0.38	0.81	0.38	0.25	1.13	1.81	1.25	0.86	0.06	2.31	3.44
Total	1675 (100)	16.7	1.5		6	11	7	4	17	18	9	11	1	20	37

Notes: EU Policy Areas: AGRICULTURE – Agriculture; Fisheries; INTERNAL MARKET – Industrial Policy and Internal Market; TRANSPORT – Transport Policy; ENVIRONMENT – Environment, Consumers and Health Protection; SERVICES – Right of Establishment and Freedom to Provide Services; LABOUR & SOCIAL – Freedom of Movement for Workers and Social Policy; TAXATION – Taxation; ENERGY – Energy; UNDERTAKINGS – Law Relating to Undertakings; JUSTICE – Area of Freedom, Security and Justice; COMPETITION – Competition Policy; CUSTOMS – Customs Union and Free Movement of Goods; FINANCE – General, Financial and Institutional Matters; ECONOMY – Economic and Monetary Policy and free Movement of Capital; PEOPLE'S EUROPE – People's Europe (Freedom of movement of people, European citizenship); EXTERNAL RELATIONS – External Relations; NO DIFFERENTIATION – no selection of specific policy fields; these studies usually use infringement statistics which might, but not necessarily, cover all policy fields. IDMC – Institutional decision-making capacity; GOF – Goodness of fit; AE – Administrative Efficiency; NME – National Monitoring and Enforcement; All – All 12 theoretical arguments listed in Table 2.

Table 4 Infringements of directives and number of studies analysing robust factors by country

Country	Transposition	Application	Total	Number of Qualitative Studies (C)					Number of Quantitative Studies (Q)					C & Q	
	% RO	% RO	Total Directives	IDMC	GOF	AE	NME	All	IDMC	GOF	AE	NME	All	Total	Total Coverage*
Greece	28.1	1.6	1675	–	–	1	–	1	5	2	3	–	6	7	19
Italy	27.5	3.2	1675	–	2	2	–	2	4	2	2	–	5	7	19
Portugal	26.1	1.9	1675	–	–	–	–	–	2	–	1	–	2	2	14
Luxembourg	21.8	1.4	1675	–	–	–	–	–	–	1	–	–	1	1	13
Belgium	19.5	1.8	1675	1	1	–	–	1	2	1	1	–	3	4	16
France	18.9	2.7	1675	–	1	1	–	1	3	1	1	–	4	5	17
Germany	17.4	1.7	1675	2	5	1	3	6	5	2	3	–	6	12	24
Ireland	17.1	1.6	1675	–	1	–	1	1	3	–	1	–	3	4	16
Austria	15.2	0.5	1056	–	–	–	–	–	–	–	–	–	–	–	11
UK	13.0	1.4	1675	2	6	2	1	6	5	1	3	–	5	11	23
Spain	12.1	1.9	1675	–	2	2	2	3	5	2	3	–	6	9	21
Netherlands	9.9	0.8	1675	2	2	–	–	3	6	3	3	–	7	10	22
Finland	7.0	0.7	1056	–	–	–	–	–	–	–	–	–	–	–	11
Sweden	6.9	0.3	1056	–	–	–	–	–	1	–	–	–	1	1	12
Denmark	2.3	0.4	1675	1	1	–	–	1	2	1	1	–	3	4	16
EU12	–	–	–	–	–	–	–	–	1	1	1	–	1	1	/
EU15	–	–	–	1	1	2	–	2	7	3	3	1	8	10	/
EU27	–	–	–	–	–	–	–	–	1	1	1	–	1	1	/
NMS12	–	–	–	–	–	–	–	–	1	–	1	–	1	1	/
Total	16.7	1.5	23268	6	11	7	4	17	18	9	11	1	20	37	/

Notes: *Table 4 reports the number of studies, which included a given country in their analysis. It is differentiated whether a study considered all EU countries (EU12, EU15, EU27) together, or selected one or a group of countries. Column 'Total Coverage' indicates the total number of studies which included a given country in their analysis, independently whether they considered all EU countries (EU12/EU15/EU27) together or selected a sample of countries.

NMS12: 12 European member states from central, eastern and southern Europe which acceded to the EU in 2004 and 2007; RO – Reasoned Opinion; IDMC – Institutional decision-making capacity; GOF – Goodness of fit; AE – Administrative Efficiency; NME – National Monitoring and Enforcement; All – All 12 theoretical arguments listed in Table 2.

and internal market policies, and fields that experience surprisingly few compliance problems such as taxation.

This finding also refers to a second important component in the research design of compliance studies that concerns the choice of countries. Formally, the potential for non-compliance differs by membership duration and the number of country-specific exemptions from implementation. A typical example for such an exemption is Luxembourg for the implementation of fisheries directives. We only show patterns for the EU-15 because the countries acceded between 2004 and 2007 do not yet provide for a large transposition record, given a conventional transposition deadline of two years after the adoption of a directive. Table 4 lists effective non-compliance of EU-15 by the number of received reasoned opinions from the Commission.²⁴

Accordingly, Scandinavian states rarely receive a reasoned opinion from the Commission, whereas countries like Greece, Italy, Portugal and Luxembourg often reach this official stage of an infringement proceeding. Portugal stands out with an apparently high share of non-compliance cases by non-notification in addition to a comparatively high share of bad application. Since most quantitative studies do not select on countries, this is more relevant for small N studies. A basic coverage of most countries by all case studies exists, though only two studies by Falkner *et al.* (2004, 2007) cover the total set of EU-15 countries.

Upon closer examination, Germany, the United Kingdom, the Netherlands and Spain are well covered by case studies on compliance in contrast to Portugal, Luxembourg, Austria, Finland and Sweden. With respect to the number of reported tests by theoretical arguments, it turns out that goodness-of-fit (GOF) and administrative efficiency (AE) have been examined in (other) countries such as Greece, Italy and France as well. For other important findings, such as institutional decision-making capacity of governments and national monitoring and enforcement, Southern countries – which are among the member states with the highest infringement rates – have only been subject to a few/limited number of case studies.

In sum, the concentration of qualitative compliance studies on social and environmental policies recommends a cautious interpretation of their field-specific findings. In some studies compliance scholars commonly investigated a few prominent directives, which may overemphasize their findings (e.g., Börzel 2000, 2002, 2006; Knill and Lenschow 1998).²⁵ This is also true for Falkner *et al.* (2005) on six labour and social policy directives and Thomson *et al.* (2006) on about 20 directives, which are used in multiple publications (Falkner and Treib 2008; Falkner *et al.* 2004; Thomson 2007, 2009; Thomson *et al.* 2007; Toshkov 2007; Zhelyazkova and Torenvlied 2009). This risks a multiplication of similar or identical findings in the field of compliance research. Furthermore, some countries such as Germany, the United Kingdom, the Netherlands and Spain seem to dominate qualitative insights about compliance. With the exception of three policy fields (competition, people's Europe and external relations), quantitative studies usually cover all policy fields and therefore beg a lower risk for selection bias.

CONCLUSION

This study has set out to present a research synthesis on the current state of compliance research. Moving beyond literature reviews of compliance studies, we classified and coded the findings of a large number of explanatory factors on compliance, statistically analysed the robustness of their effects and evaluated their representativeness by research design. In spite of variation in methodology and research design, two theoretical arguments stand out with robust results: the goodness-of-fit and the institutional decision-making capacity argument. Administrative efficiency, as well as national monitoring and enforcement, are robust only across qualitative studies. However, our research synthesis also reveals that case studies have a strong tendency to confirm their theoretical arguments – an observation that recommends a cautious interpretation of their findings. Furthermore, the small number of quantitative tests impedes stronger interpretation of the findings for several arguments, such as public support, learning and monitoring, which play an important role in the theoretical discussion on compliance.

The selection of policy fields and countries with only a certain level of (non-)compliance constitutes a selection risk on the dependent variable. When the two dominant policy fields – covering more than 60 per cent of all policies – are hardly represented in compliance research, which instead focuses on policy fields with about 6 per cent coverage of all policies, the true variation of (non-)compliant behaviour across policy fields is truncated. Similarly, ignoring countries with specific compliance records reduces the full range of variation of (non-)compliant behaviour that exists in the real world of the EU, which results in biased inferences and underestimated effects of the explanatory variables (King *et al.* 1994: 130). This might be one explanation for the mixed findings of administrative efficiency- and preference-related arguments, which were also tested in quantitative studies more often in policy fields and countries with a specific range of non-compliant behaviour.

This non-representative selection of policy fields and countries is also problematic for the generalizability of the reported findings on compliance. Policy fields and countries differ with respect to their own characteristics, which might be important for member states' ability and willingness to comply. For example, compared to agricultural policies, environmental and social directives frequently receive more public attention and may therefore raise higher audience costs for the implementing member state government. Similarly, Euroscepticism often varies across countries for the same policy (e.g., Lubbers and Scheepers 2005) and, according to Falkner and Treib (2008), countries have different cultures of compliance, which can explain variation in their behaviour. This means that we should also carefully interpret robust effects when they have been tested in those policy fields and countries where we expect to find these effects. We therefore remain cautious with the interpretation of results on administrative efficiency and national monitoring and enforcement.

Our research synthesis shows that administrative efficiency is consistent only across case studies, which analyse this argument in Greece, Italy, France, Germany, the United Kingdom (UK) and Spain. According to the typology of Falkner and Treib (2008), these countries belong either to the 'world of domestic politics' or 'world of neglect'. Case studies, therefore, miss a single country from the 'world of law observance', where smooth transposition is suggested to be less dependent upon administrative efficiency. Another example is national/supranational monitoring and enforcement, which is expected to work in less-ambiguous, less-complex and easy to monitor policies like environment, economy, agriculture and transport.²⁶ We find that monitoring- and enforcement-related arguments are predominantly tested for environmental policies. Also here, case studies mostly concentrate on countries like Germany, Ireland, UK and Spain, which belong to the 'world of domestic politics'. This constitutes a favourable context for the theoretical test because monitoring and enforcement are central factors in the cost-benefit calculations of this group of countries.

The case of institutional decision-making capacity and goodness-of-fit arguments is different.

Institutional decision-making capacity arguments are tested by case studies predominantly for policies with high audience costs (labour, social and environment) and in countries from the 'world of domestic politics' where we would expect to find this effect, because governments are concerned about their re-election chances. They seek to increase their popularity and minimize their audience costs. Therefore, one would expect that exactly in policies with high audience costs and in countries where domestic concerns prevail upon compliance, policy preferences and the number of veto players are highly important for smooth transposition. However, given that quantitative studies provide a broad coverage of nearly all policy fields and countries, where the effect remains consistent, we are confident about its generalizability.

We are also more convinced from the effect of goodness-of-fit, which is tested by case studies predominantly in policies (environment, social and labour) and countries ('world of domestic politics') where we would not expect goodness-of-fit to be of primer importance for transposition. In spite of this, the reported effects are consistent across studies. Additionally, in quantitative studies this argument holds across the majority of policy fields and all countries. Hence, while the reported results for administrative efficiency and national/supranational monitoring and enforcement are robust across case studies, the selection of policy fields and countries suggests further research before generalizing findings. Following the recommendations by mixed methods approaches (e.g., Lieberman 2005), future compliance research could draw on quantitative findings across all countries and policy fields for an informed selection of cases to fill these gaps.

Another finding of our study concerns the gap between theoretically relevant hypotheses on compliance identified in the literature and actually published research that evaluates them empirically. Accordingly, several theoretically relevant questions, such as the effect of public support on compliance, seem to require more attention by the empirical community. Finally, for future

compliance research, our analysis suggests that scholars should pay more attention to the selection of countries and policy fields. Otherwise, compliance research risks inducing selection bias, when the best-complying Scandinavian states and the bad compliers, such as Portugal and Greece, remain excluded from the analysis. Similarly, it is important to note that policy fields with most legislative activities, such as agriculture, internal market and transport, receive little attention from compliance scholars, while the policy fields of environment and social policies, which comprise substantially fewer legislative acts and are characterized by a high degree of non-compliance, are well-researched. Hence, a more careful consideration of not only the dependent variable but also of the countries and policy fields under scrutiny is warranted.

Biographical notes: Mariyana Angelova is PhD candidate at the Graduate School of Economic and Social Sciences and research assistant at the Collaborative Research Center SFB 884 'Political Economy of Reforms', University of Mannheim, Germany. Tanja Dannwolf is PhD candidate at the University of Mannheim. Thomas König is Professor of International Relations and Spokesperson of the Collaborative Research Center SFB 884 'Political Economy of Reforms' at the University of Mannheim.

Address for correspondence: Mariyana Angelova, Collaborative Research Center SFB 884 'Political Economy of Reforms', University of Mannheim, 68131 Mannheim, Germany. email: angelova@uni-mannheim.de/Tanja Dannwolf, Department of Political Science, Chair of Political Science III, University of Mannheim, 68131 Mannheim, Germany. email: dannwolf@uni-mannheim.de/Thomas König, Department of Political Science, Chair of Political Science II, University of Mannheim, 68131 Mannheim, Germany. email: koenig@uni-mannheim.de

ACKNOWLEDGEMENTS

The authors' names on this paper are given in alphabetical order; all authors contributed equally to all of the work. The codebook, database and all replication files are available online at <http://www.sowi.uni-mannheim.de/lspol2/08downloads01.html>. We are very grateful to James Lo, Brooke Luetgert, Santiago Olivella, Nikoleta Yordanova and several anonymous reviewers for their valuable comments and constructive criticism.

NOTES

- 1 Additionally, some studies report that member states considerably delay implementation despite a good policy fit (Bugdahn 2005: 192; Falkner *et al.* 2005: 261–2; Knill and Lenschow 1998: 600–2), while others find that they comply properly despite a bad fit (Falkner *et al.* 2005; Haverland 2000).
- 2 Other examples for contradictory findings include the effect of (in-)efficient bureaucracies (Duina 1997: 174; Thomson 2007: 1000), the impact of Council

voting power (Jensen 2007: 466, Perkins and Neumayer 2007), the relevance of governmental (Hille and Knill 2006: 546; Linos 2007: 558) and public support of the EU (Mbaye 2001: 273; Sverdrup 2004).

- 3 Despite their different methodologies and research design, compliance scholars evaluate similar theoretical arguments, which they measure using a number of operational variables. For example, one prominent hypothesis is that restrictions on institutional capacity have a negative impact on compliance. This concept of institutional capacity limitations is measured through various indicators, including the existence of a second chamber, the number of government parties, the degree of centralization and involvement of the national parliament in the transposition process. Although each indicator is an independent empirical phenomenon *per se*, they are used to operationalize the same latent concept of capacity restrictions and to evaluate the corresponding theoretical argument.
- 4 For experimental studies Lipsey and Wilson (2000: 2) recommend including only quantitative analyses when these use similar research designs and their samples do not overlap. However, in the field of compliance with directives many (quantitative) studies neither apply a similar research design nor exclude an overlapping of their samples.
- 5 Toshkov (2010) reports the results for veto players and the results for federalism, regionalism and corporatism independently, although they represent the same theoretical argument that additional actors with decision-making powers can constraint the transposition decisions of the government.
- 6 For an introduction to research synthesis and meta-analysis see Borenstein *et al.* (2009), Lipsey and Wilson (2000), Cooper (2010) and Cooper *et al.* (2009).
- 7 See the Thomson Reuters (ISI) Web of Knowledge, Social Sciences Citation Index, available at <http://isiknowledge.com/> (accessed 3 July 2012).
- 8 Because the peer review process for books and chapters for edited volumes is not as well-established and standardized as it is for journal articles, we refrain from including book chapters in our study.
- 9 For this reason we excluded the following four studies: Mendrinou (1996); Börzel (2001); Demmke (2001); and Tallberg (2002).
- 10 Interaction effects provide a more realistic, but also more complex, picture of the reasons for non-compliance. The assessment of interaction effects, however, does not fit into our study synthesis approach to count the effects of one theoretical argument at a time.
- 11 For this reason our sample excludes the work by Zhelyazkova and Torenvlied (2009).
- 12 Whereas the coding scheme for the theoretical reasoning was developed deductively, the operationalization categories for the independent variables were naturally less fixed and the coding scheme was updated when necessary.
- 13 Orwin (1994: 152) recommends a Kappa in this range as acceptable.
- 14 For the estimation of the 95 per cent confidence intervals of binomial proportions, we compared three other recommended intervals – Agresti–Coull, Jeffreys and Clopper–Pearson intervals (Brown *et al.* 2001). The results are only marginally different from the Wilson intervals and do not change our results.
- 15 For interested readers, significant levels can be found in our database, available online at <http://www.sowi.uni-mannheim.de/lspol2/08downloads01.html>
- 16 This database may be found on the European Union's website *EUR-Lex – Access to European Union Law*, available free of charge at <http://eur-lex.europa.eu/en/index.htm> (last accessed 4 July 2012), more specifically in sector 7.
- 17 A strong definition of compliance requires formal timely and correct implementation, practical application, as well as monitoring and enforcement (Börzel 2003: 60).
- 18 Non-notification captures the act of transposition of European directives into national law by the adoption and notification of legal instruments to the European Union. In

- contrast, non-conformity relates to the legal correctness of the transposition instruments, and bad application refers to the actual application on the ground. A fourth category, enforcement on the national level, has been suggested (Falkner *et al.* 2005: 12). Actual goal achievement has not really been of interest in this literature so far.
- 19 Note that multiple variables may inspect the same theoretical argument in one study.
 - 20 These confidence intervals represent a hypothesis test, where the null hypothesis is that the proportion of positive effects is equal to 0.5 – the threshold where both positive and negative findings are equally likely and hence the reported effect might be random. Randomness represents inconsistency. Accordingly, we consider the results as robust when the confidence intervals do not include 0.5.
 - 21 Power tests are particularly important when we try to establish whether a type II error has occurred – that is, whether we have wrongfully failed to reject a null hypothesis. This may occur in cases when n is small (see Agresti and Finlay 2009: 168–9). Statistical power indicates the probability of rejecting the null hypothesis given that it is actually false, and it is equal to 1 minus the probability of making a type II error. Therefore, low powers (usually below 0.80) are indicators for a high probability of making a type II error. For the purposes of this article we calculated the powers for a sign test (see Cohen 1988).
 - 22 We disregard significance levels, thereby adopting a conservative approach in that we minimize type 2 errors. However this decision comes at a price – the inclusion of not significant results might introduce a ‘noise’ in the meta-analysis. To control for this possibility we conducted a separate analysis only for statistically significant results from quantitative studies. The observable probability of finding a positive significant effect is either identical to the values reported in Table 2 or is slightly different. The differences in proportions are statistically insignificant with an average difference equal to 0.08. The results of the robustness test are also identical for the goodness-of-fit argument. We cannot draw any conclusions as to whether this also holds for the robustness of other theoretical arguments because the breaking of the results into significant and not significant effects considerably lowers the number of observations in each of the groups and thus results in low, in particular less than 0.8, statistical power across the board.
 - 23 Table 3 considers directives which entered into force between 1989 and 2005, and infringements which were issued between 1989 and 2009. Violative opportunities are calculated for each policy field by multiplying the number of adopted directives with the number of member states. Note that differences exist because some countries became members after the adoption date of a directive. Since none of the studies base the selection of policy fields on theoretical grounds, to identify policy fields we followed the policy division in the directory codes of EURLEX as of October 2010. Because some studies analyse more than one policy field, the total number of studies is not the mere sum of all studies per individual policy field. Instead it indicates the number of studies which tested a given theoretical argument at least once.
 - 24 Considering a two-year transposition period for directives, we coded whether a reasoned opinion for non-notification or bad application was issued for each directive and member state.
 - 25 Council Directive 80/778/EEC of 15 July 1980 relating to the quality of water intended for human consumption, Council Directive 85/337/EEC of 27 June 1985 on the assessment of the effects of certain public and private projects on the environment.
 - 26 In line with existing literature we use the length of directive deadlines as a proxy for complexity of directives in terms of work that needs to be done (Mastenbroek 2003). Highly detailed and complex directives are claimed to be transposed slowly (Ciavarini Azzi 2000: 56) and therefore allow more time for transposition.

We find that directives dealing with environment, economy, transport and agriculture have the lowest mean values of deadline length. We are highly thankful for this descriptive data to Nikoleta Yordanova and Asya Zhelyazkova (2011).

REFERENCES

(All studies included in our sample are marked with an asterisk.)

- Agresti, A. and Finlay, B. (2009) *Statistical Methods for the Social Sciences*, Upper Saddle River, NJ: Pearson Prentice Hall.
- *Bailey, I. (2002) 'National adaptation to European integration: institutional vetoes and goodness-of-fit', *Journal of European Public Policy* 9(5): 791–811.
- *Berglund, S., Gange, I. and Van Waarden, F. (2006) 'Mass production of law. Routinization in the transposition of European directives: a sociological institutional account', *Journal of European Public Policy* 13(5): 692–716.
- Borenstein, M., Hedges, L.V., Higgins, J.P.T. and Rothstein, H.R. (2009) *Introduction to Meta-Analysis*, Chichester: John Wiley.
- *Borghetto, E., Franchino, F. and Giannetti, D. (2006) 'Complying with the transposition deadlines of EU directives. Evidence from Italy', *Rivista Italiana di Politiche Pubbliche* 1(1): 7–38.
- Börzel, T. (2001) 'Non-compliance in the European Union: pathology or statistical artefact?' *Journal of European Public Policy* 8(5): 803–24.
- *Börzel, T.A. (2000) 'Why there is no "southern problem". On environmental leaders and laggards in the European Union', *Journal of European Public Policy* 7(1): 141–62.
- Börzel, T.A. (2002) 'Pace-setting, foot-dragging, and fence-sitting: member state responses to Europeanization', *Journal of Common Market Studies* 40(2): 193–214.
- Börzel, T.A. (2003) *Environmental Leaders and Laggards in Europe. Why there is (not) a 'Southern Problem'*, Aldershot: Ashgate.
- *Börzel, T.A. (2006) 'Participation through law enforcement – the case of the European Union', *Comparative Political Studies* 39(1): 128–52.
- Brown, L.D., Cai, T.T. and DasGupta, A. (2001) 'Interval estimation for a binomial proportion', *Statistical Science* 16(2): 101–33.
- *Bugdahn, S. (2005) 'Of Europeanization and domestication: the implementation of the environmental information directive in Ireland, Great Britain and Germany', *Journal of European Public Policy* 12(1): 177–99.
- *Bursens, P. (2002) 'Why Denmark and Belgium have different implementation records: on transposition laggards and leaders in the EU', *Scandinavian Political Studies* 25(2): 173–95.
- Bushman, B.J. and Wang, M.C. (2009) 'Vote-counting procedures in meta-analysis', in H. Cooper, L.V. Hedges and J.C. Valentine (eds), *The Handbook of Research Synthesis and Meta-Analysis*, 2nd edn, New York: Russell Sage Foundation, pp. 207–20.
- Ciavarini Azzi, G. (2000) 'The slow march of European legislation: the implementation of directives', in K. Neunreither and A. Wiener (eds), *European Integration after Amsterdam: Institutional Dynamics and Prospects for Democracy*, Oxford: Oxford University Press, pp. 52–67.
- Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum.
- Cooper, H. (2010) *Research Synthesis and Meta-Analysis. A Step by Step Approach*, Los Angeles, CA: Sage.
- Cooper, H., Hedges, L.V. and Valentine, J.C. (eds) (2009) *Handbook of Research Synthesis and Meta-Analysis*, New York: Sage.

- Deeks, J.J., Higgins, J.P.T. and Altman, D.G. (2008) 'Analysing data and undertaking meta-analyses', in J.P.T. Higgins and S. Green (eds), *Cochrane Handbook for Systematic Reviews of Interventions*, Chichester: Wiley-Blackwell, pp. 243–96.
- Demmke, C. (2001) *Towards Effective Environmental Regulation: Innovative Approaches in Implementing and Enforcing European Environmental Law and Policy*, Cambridge, MA: Harvard Law School.
- *Dimitrakopoulos, D. (2001) 'Learning and steering: changing implementation patterns and the Greek central government', *Journal of European Public Policy* 8(4): 604–22.
- *Dimitrova, A. and Rhinard, M. (2005) 'The power of norms in the transposition of EU directives', *European Integration Online Papers* 9(16), available at <http://eiop.or.at/eiop/texte/2005-016a.htm> (accessed 4 July 2012).
- *Duina, F. (1997) 'Explaining legal implementation in the European Union', *International Journal of the Sociology of Law* 25(2): 155–79.
- *Duina, F. and Blithe, F. (1999) 'Nation-states and common markets: the institutional conditions for acceptance', *Review of International Political Economy* 6(4): 494–530.
- *Falkner, G., Hartlapp, M., Leiber, S. and Treib, O. (2004) 'Non-compliance with EU directives in the member states: opposition through the backdoor', *West European Politics* 27(3): 452–73.
- *Falkner, G., Hartlapp, M. and Treib, O. (2007) 'Worlds of compliance: why leading approaches to EU implementation are only sometimes-true theories', *European Journal of Political Research* 46(3): 395–416.
- *Falkner, G. and Treib, O. (2008) 'Three worlds of compliance or four? The EU-15 compared to new member states', *Journal of Common Market Studies* 46(2): 293–313.
- Falkner, G., Treib, O., Hartlapp, M. and Leiber, S. (2005) *Complying with Europe. EU Harmonisation and Soft Law in the Member States*, Cambridge: Cambridge University Press.
- *Giuliani, M. (2003) 'Europeanization in comparative perspective: institutional fit and national adaptation', in K. Featherstone and C. Radaelli (eds), *The Politics of Europeanization*, Oxford: Oxford University Press, pp. 134–55.
- Haas, P.M. (1998) 'Compliance with EU directives: insights from international relations and comparative politics', *Journal of European Public Policy* 5(1): 17–37.
- Hartlapp, M. and Falkner, G. (2009) 'Problems of operationalization and data in EU compliance research', *European Union Politics* 10(2): 281–304.
- *Haverland, M. (2000) 'National adaptation to European integration: the importance of institutional veto points', *Journal of Public Policies* 20(1): 83–103.
- *Haverland, M., Steunenberg, B. and Waarden, F.V. (2008) 'Sectors at different speeds: analyzing transposition deficits in the European Union', *ECPR Joint Sessions of Workshops*, Rennes, 11–16 April.
- Héritier, A. et al. (2001) *Differential Europe. The European Union Impact on National Policymaking*, Lanham, MD: Rowman & Littlefield.
- *Hille, P. and Knill, C. (2006) "It's the bureaucracy, stupid." The implementation of the *acquis communautaire* in EU candidate countries, 1999–2003', *European Union Politics* 7(4): 531–52.
- Hunter, J.E. and Schmidt, F.L. (2004) *Methods of Meta-Analysis. Correcting Error and Bias in Research Findings*, Thousand Oaks, CA: Sage.
- *Jensen, C.B. (2007) 'Implementing Europe: a question of oversight', *European Union Politics* 8(4): 451–77.
- *Kaeding, M. (2008) 'Lost in translation or full steam ahead – the transposition of EU transport directives across member states', *European Union Politics* 9(1): 115–43.
- King, G., Keohane, R. and Verba, S. (1994) *Designing Social Inquiry. Scientific Inference in Qualitative Research*, Princeton, NJ: Princeton University Press.

- *Knill, C. and Lenschow, A. (1998) 'Coping with Europe: the impact of British and German administrations on the implementation of EU environmental policy', *Journal of European Public Policy* 5(4): 595–614.
- *König, T. and 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): 163–94.
- Krippendorff, K. (2004) *Content Analysis. An introduction to its methodology*, London: Sage Publications.
- Krislov, S., Ehlermann, C.-D. and Weiler, J.H. (1986) *The Political Organs and the Decision-Making Process in the United States and the European Community*, New York: de Gruyter.
- *Lampinen, R. and UUsikylä, P. (1998) 'Implementation deficit – why member states do not comply with EU directives?', *Scandinavian Political Studies* 21(3): 231–51.
- Lieberman, E.V. (2005) 'Nested analysis as a mixed-method strategy for comparative research', *American Political Science Review* 99(3): 435–52.
- *Linos, K. (2007) 'How can international organizations shape national welfare states? Evidence from compliance with European Union directives', *Comparative Political Studies* 40(5): 547–70.
- Lipsey, M.W. and Wilson, D.B. (2000) *Practical Meta-Analysis*, Thousand Oaks, CA: Sage.
- Lubbers, M. and Scheepers, P. (2005) 'Political versus instrumental Euro-scepticism: mapping scepticism in European countries and regions', *European Union Politics* 6(2): 223–42.
- *Luetgert, B. and Dannwolf, T. (2009) 'Mixing methods: an analysis of EU member state transposition patterns', *European Union Politics* 10(3): 307–34.
- *Mastenbroek, E. (2003) 'Surviving the deadline: the transposition of EU directives in the Netherlands', *European Union Politics* 4(4): 371–95.
- Mastenbroek, E. (2005) 'EU compliance: still a "black hole"?', *Journal of European Public Policy* 12(6): 1103–20.
- Mastenbroek, E. and Kaeding, M. (2006) 'Europeanization beyond the goodness of fit: domestic politics in the forefront', *Comparative European Politics* 4(4): 331–54.
- *Mastenbroek, E. and Van Keulen, M. (2006) 'Beyond the goodness of fit: a preference-based account of europeanization', in R. Holzhaecker and M. Haverland (eds), *European Research Reloaded: Cooperation and Integration among Europeanized States*, Dordrecht: Springer, pp. 19–42.
- *Mbaye, H.A.D. (2001) 'Why national states comply with supranational law. Explaining implementation infringements in the European Union, 1972–1993', *European Union Politics* 2(3): 259–81.
- Mendrinou, M. (1996) 'Non-compliance and the European Commission's role of integration', *Journal of European Public Policy* 3(1): 1–22.
- Orwin, R.G. (1994) 'Evaluating coding decisions', in H. Cooper and L.V. Hedges (eds), *The Handbook of Research Synthesis*, New York: Russell Sage Foundation, pp. 149–62.
- *Perkins, R. and Neumayer, E. (2007) 'Do membership benefits buy regulatory compliance? An empirical analysis of EU directives 1978–1999', *European Union Politics* 8(2): 180–206.
- Sedelmeier, U. (2008) 'After conditionality: post-accession compliance with EU – law in East Central Europe', *Journal of European Public Policy* 15(6): 806–25.
- Siedentopf, H. and Ziller, J. (1988) *Making European Policies Work: The Implementation of Community Legislation in the Member States. Vol. 1: Comparative Syntheses*, London: Sage.
- *Steunenbergh, B. (2006) 'Turning swift policy-making into deadlock and delay. National policy coordination and the transposition of EU directives', *European Union Politics* 7(3): 293–319.

- *Steunenberg, B. and Toshkov, D. (2009) 'Comparing transposition in the 27 member states of the EU: the impact of discretion and legal fit', *Journal of European Public Policy* 16(7): 951–70.
- Sverdrup, U. (2004) 'Compliance and conflict management in the European Union: Nordic exceptionalism', *Scandinavian Political Studies* 27(1): 23–43.
- Tallberg, J. (2002) 'Paths to compliance: enforcement, management, and the European Union', *International Organization* 56(3): 609–43.
- *Thomson, R. (2007) 'Time to comply: national responses to six EU labour market directives revisited', *West European Politics* 30(5): 987–1008.
- *Thomson, R. (2009) 'Same effects in different worlds: the transposition of EU directives', *Journal of European Public Policy* 16(1): 1–18.
- Thomson, R., Stokman, F.N., Achen, C.H. and König, T. (2006) *The European Union Decides. The Empirical Relevance of Policy Making Models*, Cambridge: Cambridge University Press.
- *Thomson, R., Torenvlied, R. and Arregui, J. (2007) 'The paradox of compliance: infringements and delays in transposing European Union directives', *British Journal of Political Science* 37(4): 685–709.
- *Toshkov, D. (2007) 'In search of the worlds of compliance: culture and transposition performance in the European Union', *Journal of European Public Policy* 14(6): 933–59.
- *Toshkov, D. (2008) 'Embracing European law: compliance with EU directives in Central and Eastern Europe', *European Union Politics* 9(3): 379–402.
- Toshkov, D. (2010) 'Taking stock: a review of quantitative studies of transposition and implementation of EU law', *Vienna Institute for European Integration Research. Working Paper Series*.
- Toshkov, D., Knoll, M. and Wewerka, L. (2010) 'Connecting the dots: case studies and EU implementation research', *Vienna Institute for European Integration Research Working Paper Series*.
- Treib, O. (2003) 'Die Umsetzung von EU-Richtlinien im Zeichen der Parteipolitik: Eine akteurszentrierte Antwort auf die Misfit-These', *Politische Vierteljahresschrift* 44(4): 506–28.
- Treib, O. (2008) 'Implementing and complying with EU governance outputs', *Living Reviews in European Governance* 3(5), available at <http://europeangovernance.livingreviews.org/Articles/lreg-2008-5/> (accessed 4 July 2012).
- Van den Bossche, P. (1996) 'In search for remedies for non-compliance: the experience of the European Communities', *Maastricht Journal of European and Comparative Law* 3(4): 371–99.
- Weiler, J.H.H. (1991) 'Problems of legitimacy in post 1992 Europe', *Außenwirtschaft* 46(3/4): 411–37.
- Wilson, D.B. (2009) 'Systematic coding', in H. Cooper, L.V. Hedges and J.C. Valentine (eds), *The Handbook of Research Synthesis and Meta-Analysis*, New York: Russell Sage Foundation, pp. 159–75.
- Yordanova, N. and Zhelyazkova, A. (2011). 'Reversed economic left–right party ideology? Transposition of EU legislation in Eastern vs. Western Europe', Paper presented at the *EUSA Twelfth Biennial International Conference*, Boston, MA, 3–5 March.
- Zhelyazkova, A. and Torenvlied, R. (2009) 'The time-dependent effect of conflict in the Council on delays in the transposition of EU directives', *European Union Politics* 10(1): 35–62.
- *Zubek, R. (2005) 'Complying with transposition commitments in Poland: collective dilemmas, core executive and legislative outcomes', *West European Politics* 28(3): 592–619.