"The EU Multi-level System and the Europeanization of Domestic Blame Games"
Tim Heinkelmann-Wild, Lisa Kriegmair and Berthold Rittberger

ONLINE APPENDIX

1. Data selection and coding procedure

We coded blame attributions of national MPs in a two-step process. In a first step, potential blame attributions are identified by *automatically* coding text segments including direct references to the specific policies under investigation and keywords from a blame-related dictionary. We used the MaxQDA function "keyword-in-context" and coded the three sentences before and seven sentences after a case-specific keyword. We then used the MaxQDA function "quantitative content analysis" and coded the three sentences before and three sentences after a keyword from a previously constructed dictionary. Figure A.1 summarizes our coding process.

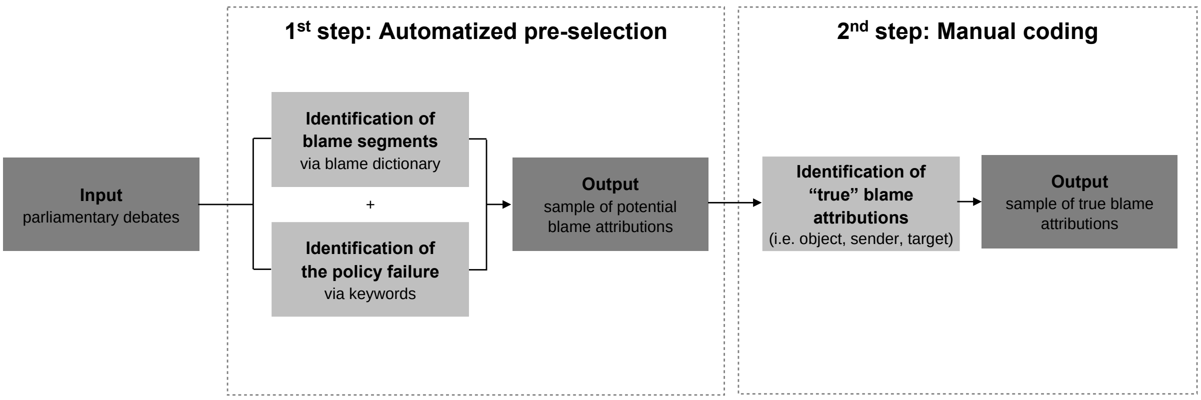

Figure A.1. Two-step coding process

Our *blame dictionary* combines keywords from two sources. First, we build on a pre-existing blame dictionary (Müller, Porcaro, & von Nordheim, 2018). This dictionary contains 177 German words which denote an attribution of responsibility, for instance to an actor or an institution, and was developed to measure the extent of scapegoating in the euro area in the context of the financial crisis. Second, to account for the context of migration policies, we extended this dictionary by drawing on manually coded blame attributions in migration policymaking (Heinkelmann-Wild, & Zangl, 2019). From the manually coded data, we selected the lemmatized versions of those words that were frequently recurring in the coded blame statements and had a negative connotation.

We refined our blame dictionary through multiple trial runs, using a limited set of articles. This allowed us to eliminate keywords from our dictionary that did not help identifying "true" blame statements. For instance, the term "falle" was eliminated from the dictionary, as the lemma is part of a number of German words that are not related to responsibility, such as "ein*falle*n", and thus lead to a large number of statements falsely identified as potential blame statements. All in all, our combined "blame dictionary" comprises 169 words such as "anklag*", "mangel*", "schuld*", and "verantwort*". Applying our dictionary, we identified 881 potential blame segments (see Table A.1).[1]

The second step constituted in the *manual coding* of the relevant blame attributions in the new sample of pre-selected segments by three coders. The following provides an example:

> 'It is a fact that all Frontex-coordinated rescue operations of Operation Triton take place directly off the Libyan coast. And this despite the fact that the operation would have to take place at the European external border. […] An EU-wide action must not play into the hands of the machinations of inhuman smuggling gangs and fuel the deadly risk.' (Krings 2015, p. 9474; translation by the authors)

The blame sender is Günter Krings, from the German governing party CDU. He attributes responsibility to a blame target, Frontex (i.e. an EU-level actor). The EU policy failure constituting the blame object is that Frontex, by operating in proximity of the Libyan coast, is said to facilitate human trafficking and to thus increase the risk of refugees drowning in the Mediterranean.

In our automatically pre-coded samples, the three coders manually identified 558 blame attributions in 390 debates (see Table A.1).[2]

Table A.1. Overview of the results of the two-step coding process.

Case	Country	Period	Search keys	Debates	Potential blame	True blame statements
Border control	Austria	01/01/2004-13/08/2018	"Frontex"	47	105	35
	Germany	01/01/2004-13/08/2018	"Frontex"	122	284	81
Asylum system	Austria	01/01/2003-13/08/2018	"Dublin"	84	183	152
	Germany	01/01/2003-13/08/2018	"Dublin"	137	309	181

Note: The number of blame statements by sender (i.e. 449) and the total number of blame attributions to a specific target (i.e. 558) vary as one sender often blames two (or more) targets, which were coded separately. Since the dictionary cannot differentiate between single and multiple targets, we used blame attributions by sender for this table.

1 The blame dictionary is available at: https://data.ub.uni-muenchen.de/175/.

2 The coded data is available at: https://data.ub.uni-muenchen.de/174/.

2. Reliability of the dictionary

The two-step coding process allows the human coder to decide whether a pre-selected segment contains relevant blame attributions. It is thus unproblematic for the purpose of this paper that the number of pre-selected segments exceeds the number of relevant blame statements (see Table A.1.). Yet, it is important to test whether the dictionary systematically omits relevant blame statements. To check for omitted blame statements, we asked our three coders to identify blame attributions in a sample of 24 randomly selected debates without applying the blame dictionary first. There is moderate agreement between the dictionary-based pre-selection of blame-related paragraphs and those paragraphs identified by human coders. The dictionary-based pre-selection led to the omission of 50 out of 93 blame-related paragraphs that were identified in test coding conducted by the coders without the dictionary (54%). At the same time, the coders omitted eight blame-related segments they identified in the original coding with the help of the dictionary.

However, when we turn to the exact blame attributions that were coded in the pre-selected segments and compare them with the blame attributions in the test coding (see Table A.2), we find that the dictionary does not systematically omit blame attributions from specific senders or to specific targets. As the omitted codes are equally distributed across the four combinations of blame senders and targets, we can thus exclude a systematic bias of the dictionary-based pre-selection.

Table A.2: Concordance of human coders with and without dictionary-based pre-selection

	Original coding (with pre-selection)	**First test coding** (without pre-selection)	**Agreement**
total	**67**	**145**	**50**
Gov blame to “external” EU actors	16	50	11
Gov blame to domestic actors	6	6	5
Opp blame to “external” EU actors	14	37	10
Opp blame to domestic actors	31	52	24

What is more, we find a similar (dis-)agreement when we asked a fourth coder to code the same randomly selected sample of 24 debates without dictionary-based pre-selection and compared it with the first test coding (see Table A.3). The fourth coder agreed with 73 out of the 168 blame attributions coded in the first test coding (43%). While the fourth coder omitted 72 statements found in the first test coding, she identified 23 statements that were omitted by the other coders. This overall agreement corresponds to comparable studies of responsibility attributions (Gerhards, Offerhaus, & Roose, 2007, p. 117; Schwarzenbeck, 2017, pp. 305–307).

Table A.3. Concordance of human coders without dictionary-based pre-selection

	First test coding (without pre-selection)	**Second test coding** (without pre-selection)	**Agreement**
total	**145**	**96**	**73**
Gov blame to “external” EU actors	50	47	30
Gov blame to domestic actors	6	0	0
Opp blame to “external” EU actors	37	22	18
Opp blame to domestic actors	52	27	25

In sum, these tests corroborate our confidence that the two-step process to pre-select and code blame attributions suggested in this paper is an efficient and reliable method to analyze politicians' blame attribution behavior. While the automatic, dictionary-based pre-selection led to a comparable share of omitted blame statements as purely human coding, the automatized pre-selection allows for the reliable reproduction of the pre-selection process without a systematic bias. In addition, the dictionary-based pre-selection facilitates the coding process considerably. While the three original coders on average needed 20 minutes to code the pre-selected segments within a debate, without dictionary, it took them almost five times as long to read an entire debate.

3. Statistical tests of the hypotheses

The empirically observed patterns of blame attributions for contested EU policies corroborate our expectations about the direction and frequency of national politicians' blame. To further strengthen our confidence in the empirical results, we conducted a series of statistical tests.

First, we assessed whether our data on the *direction of blame attributions* deviates from a random distribution via cross tabulation. In other words, we tested the null hypothesis that there is no relationship between the independent variables and the direction of blame. Tables A.4-7 compare our observations with the expected values for a random distribution (in brackets) for each independent variable (i.e. the institutional position of the blame sender and the policy-specific authority structure), while holding the other independent variable constant. The expected value for each cell is calculated by multiplying the row total by the column total, then dividing by the grand total.

We find that the observed and expected values differ considerably. In the EU asylum system case, the relationship between the blame sender belonging to a government or opposition party and the direction of blame comes with a *chi-square value of 66.4*. This implies that the null hypothesis can be rejected at the 0.01 level of significance (99% confidence level) (Table A.4). Moreover, we find a *chi-square value of 57.59* for the relationship between the policy-specific authority structure and the direction of opposition parties' blame. This also indicates that the null hypothesis can be rejected at the 0.01 level of significance (99% confidence level) (Table A.7). While two of the sub groups do not exhibit chi-square values that allow for the rejection of the null-hypothesis, the lower blame statement numbers for government and opposition parties in the EU border control case (Table A.5) and the overall blame attributions by government parties (Table A.6) arguably impeded a similar level of significance. Since we can reject the null hypothesis for both independent variables in the two subgroups with higher blame statement numbers, we are however confident that the observed patterns in our sample are not random.

Table A.4. Observed values vs. expected values for a random distribution (in brackets) in the EU asylum system case.

	Government	Opposition	Row totals
Blame to "external" EU actors	55 (36)	95 (114)	150
Blame to domestic actors	36 (55)	190 (171)	226
Column totals	91	285	

Table A.5. Observed values vs. expected values for a random distribution (in brackets) in the EU border control case.

	Government	Opposition	Row totals
Blame to "external" EU actors	6 (5)	122 (123)	128
Blame to domestic actors	1 (2)	53 (52)	54
Column totals	7	175	182

Table A.6. Observed values vs. expected values for a random distribution (in brackets) for government parties.

	Asylum system	Border control	Row totals
Blame to "external" EU actors	55 (57)	6 (4)	61
Blame to domestic actors	36 (34)	1 (3)	37
Column totals	91	7	98

Table A.7. Observed values vs. expected values for a random distribution (in brackets) for opposition parties.

	Asylum system	Border control	Row totals
Blame to "external" EU actors	95 (134)	122 (83)	217
Blame to domestic actors	190 (151)	53 (92)	243
Column totals	285	175	460

We also checked if our hypothesis about the direction of blame holds on the national level. For this purpose, we divided the two subgroups with sufficient statement numbers (Table A.4 and A.7) into an Austrian and a German subgroup and again tested the null hypothesis that there is no relationship between the dependent and the independent variables.

The deviation between the observed and expected values represented in Tables A.8-11 largely remains the same. In the EU asylum system case, the relationship between the blame sender belonging to a government or opposition party and the direction of blame comes with a *chi-square value of 10.22* for Austria (Table A.8) and a *chi-square value of 11.76* for Germany (Table A.9). This implies that the null hypothesis can be rejected at the 0.01 level of significance (99% confidence level). In addition, for the relationship between the policy-specific authority structure and the direction of blame, we obtained a *chi-square value of 18.14* for Austrian opposition parties (Table A.10) and a *chi-square value of 25.45* for German opposition parties (Table A.11). This indicates that the null hypothesis can be rejected at the 0.01 level of significance (99% confidence level). Overall, we are thus confident that our model holds for both Austrian and German politicians.

Table A.8. Observed values vs. expected values for a random distribution (in brackets) in the EU asylum system case in Austria.

	Austrian government	Austrian opposition	Row totals
Blame to "external" EU actors	21 (13)	31 (39)	52
Blame to domestic actors	20 (28)	95 (87)	115
Column totals	41	126	167

Table A.9. Observed values vs. expected values for a random distribution (in brackets) in the EU asylum system case in Germany.

	German government	German opposition	Row totals
Blame to "external" EU actors	34 (23)	64 (75)	98
Blame to domestic actors	16 (27)	95 (84)	111
Column totals	50	159	209

Table A.10. Observed values vs. expected values for a random distribution (in brackets) for Austrian opposition parties.

	Asylum system	Border control	Row totals
Blame to "external" EU actors	31 (46)	31 (16)	62
Blame to domestic actors	95 (80)	14 (29)	109
Column totals	126	45	171

Table A.11. Observed values vs. expected values for a random distribution (in brackets) for German opposition parties.

	Asylum system	Border control	Row totals
Blame to "external" EU actors	64 (85)	91 (70)	155
Blame to domestic actors	95 (74)	39 (60)	134
Column totals	159	130	289

Second, we conducted a Wilcoxon rank-sum test (also known as Mann-Whitney two-sample statistic) to assess whether there is a statistically significant relationship between the independent variables and the *frequency of blame*. This test is a suitable choice, since our dependent variable (i.e. the absolute number of blame attributions per debate) is not distributed normally. It tests the null hypothesis that two independent samples are from populations with the same distribution. We thus calculated the rank-sum test for each independent variable while holding the other independent variable constant (see Table A.12). For three out of four comparisons, we obtain values that allow us to reject the null hypothesis that there is no relationship between the variables at the 0.01 level of significance (99% confidence level). We are thus confident that the observed distributions of blame attributions in our sample are not random.

Table A.12. Wilcoxon rank-sum test for the two independent variables.

	Rank sum		Difference	Total observations
(1) Government blame	Border	Asylum		
	29432	46814	-5.368***	
Observations (column totals)	169	221		390
(2) Opposition blame	Border	Asylum		
	33221	43024	0.181	
Observations (column totals)	169	221		390
(3) Blame for asylum policy	Government	Opposition		
	7971	16561	-12.905***	
Observations (column totals)	123	98		221
(4) Blame for border policy	Government	Opposition		
	4278	10087	-12.275***	
Observations (column totals)	92	77		169

Note: *p<.1; **p<.05; ***p<.01.

We also checked if our hypotheses about the frequency of blame hold on the national level. For that purpose, we divided our sample into subgroups by the senders' nationality (see Table A.13) and employed the Wilcoxon rank-sum test to assess whether there is a relationship between the dependent and the two independent variables. The results in Table A.13 indicate that we can reject the null hypothesis that there is no relationship between the variables for all eight possible constellations at least at the 0.1 level of significance (90% confidence level). We are thus very confident that our model holds for the blame attribution behaviour of both Austrian and German politicians.

Table A.13. Wilcoxon rank-sum test for the two independent variables in national samples.

	Rank sum		Difference	Total observations
(1) Government blame (Austria)	Border	Asylum		
	2705	5941	-2.928***	
Observations (column totals)	47	84		131
(2) Government blame (Germany)	Border	Asylum		
	14295	19376	-4.437***	
Observations (column totals)	122	137		259
(3) Opposition blame (Austria)	Border	Asylum		
	2714	5933	-2.058**	
Observations (column totals)	47	84		131
(4) Opposition blame (Germany)	Border	Asylum		
	16842	16828	1.799*	
Observations (column totals)	122	137		259
(5) Blame for asylum policy (Austria)	Government	Opposition		
	887	2684	-7.608***	
Observations (column totals)	40	44		84
(6) Blame for asylum policy (Germany)	Government	Opposition		
	3595	5858	-10.319***	
Observations (column totals)	83	54		137
(7) Blame for border policy (Austria)	Government	Opposition		
	561	567	-6.649***	
Observations (column totals)	33	14		47
(8) Blame for border policy (Germany)	Government	Opposition		
	1770	5733	-10.204***	
Observations (column totals)	59	63		122

Note: *p<.1; **p<.05; ***p<.01.

References

Gerhards, J., Offerhaus, A., & Roose, J. (2007). Die öffentliche Zuschreibung von Verantwortung: Zur Entwicklung eines inhaltsanalytischen Instrumentariums. [The public attribution of responsibility: Regarding to the development of a content-analytical toolkit]. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *59*, 105–124. https://doi.org/10.1007/s11577-007-0005-2

Heinkelmann-Wild, T., & Zangl, B. (2019). Multi-level blame games: Blame-shifting in the European Union. *Governance*, 1–17. https://doi.org/10.1111/gove.12459

Krings, G. (2015): Annex 38. Parliamentary State Secretary Dr Günter Krings's reply to the question by Luise Amtsberg (MP). Deutscher Bundestag, 18th legislative period, 99th session, 22.04.2015, 9474.

Müller, H., Porcaro, G., & von Nordheim, G. (2018). Tales from a crisis: diverging narratives of the euro area. *Bruegel Policy Contribution*. (3), 1–12.

Schwarzenbeck, H. (2017). *Politische Verantwortung: Der Einfluss von Governance-Design auf die öffentliche Zuweisung von Verantwortung [Political responsibility: The impact of governance design on public responsibility attribution].* Baden-Baden: Nomos.