Online Appendix

Setting the Tone:

The Diffusion of Moral and Moral-Emotional Appeals in Political and Public Discourse

Table of Contents

Online Appendix A: Information on Twitter data

Twitter data have been collected using the R-package *academictwitteR* (Barrie and Chun-ting Ho [2021] 2021). The "politics" sample include all available Twitter accounts of Members of Parliament (MPs) from the 19th legislative period and official party accounts for all German political parties represented in the Bundestag. Overall, after filtering by topic (described further below), the politics sample includes 29,073 tweets from 492 different Twitter accounts. Parties included are the six major parties represented in parliament: *Die Linke* (The Left), the Greens, the social democratic SPD, the liberal party FDP, the conservative party CDU/CSU, and the radical-right party AfD. As described in the main text, to test H2, we split the politics sample in two, distinguishing the challenger party AfD from all other parties. As a replication (see Online Appendix G), we also categorize parties into party families as suggested by other research (Rooduijn et al. 2023; Lehmann et al. 2023): *Die Linke* as radical-left, the Greens and the SPD as center-left, the FDP and the CDU/CSU as center-right, and the AfD as radical right.

To incorporate citizen data, we followed a similar strategy as Barberá and co-authors (2019). We created an "attentive public" sample by randomly sampling Twitter users that follow two of the largest German newspapers on Twitter (a left-leaning newspaper, *Süddeutsche Zeitung*, and a conservative newspaper, *Die Welt*). As outlined by Barberá and co-authors (2019), the idea is that most citizens do not follow politics on a daily basis and do not have clear policy preferences (Converse 2006). However, the ones that are politically interested – the attentive public – should not only pay more attention to politicians' messages but also potentially be able to shape the behavior/communication of elites (Katz and Lazarsfeld 1955).

From our attentive public sample, we filtered out inactive ones by only including accounts who have posted 100 tweets or more. Finally, as Barberá and colleagues, we took a random sample of

10,000 users from whom we scraped all available tweets. Again, we filtered all scraped tweets by topic (using the same keyword list as applied on the politics sample, described below) to only retain tweets addressing the topic of immigration. The remaining attentive public dataset consists of 10,634 tweets from 1,153 users during the period of study.

Finally, we included media perspectives in our analysis. We did so because it may well be that both public and political discourses are led by the mass media. We therefore include newspapers' communication as a control variable in our time series model. To do so, we scraped the Twitter accounts of the same left-leaning newspaper (*Süddeutsche Zeitung*) and right-leaning newspaper (*Die Welt*), as well as the largest tabloid newspaper (*Bild*). After we filtered by topic, the newspaper sample consists of 13,278 tweets from three different Twitter accounts.

In our selection of publications for the attentive public sampling process, we exclusively focused on quality newspapers, thus excluding *Bild*. The choice to not include *Bild* in our definition of the attentive public but nonetheless include it in our media sample stems from the newspaper's status as a tabloid publication. It often faces public criticism for its sensationalist news coverage and occasionally controversial journalistic practices. Tabloid newspapers' intended audience are typically not politically interested citizens; rather, like *Bild*, they are aimed at readers who enjoy celebrity news stories. Including *Bild* in the sampling process would therefore likely bias our sample of the attentive public in the direction of less politically interested citizens. In addition, the conservative leaning of the newspaper might have introduced an ideological imbalance in the attentive public sample. At the same time as we refrain from basing our sample definition of the attentive public on followership of *Bild,* we acknowledge *Bild*'s potential influence on public and political discourse, which motivated us to included it in our newspaper sample. Online Appendix H replicates the analyses excluding *Bild* from the newspaper sample with similar results as those reported in the main text.

Finally, to ensure that any detected effects truly capture the diffusion of rhetorical style (i.e., taking over the same kind of language and incorporating it in one's own communication), we removed all retweets from our data sets. This is in line with the approach suggested by Barberá et al. (2019) to alleviate concerns of artificially enhanced correlations between actor groups. This issue could be particularly problematic in the dynamic between the attentive public and the newspaper sample, since we derived the former from the followers of the latter.

Online Appendix B: Computational text analysis

Online Appendix B1: German keyword string to identify the topic of immigration

"immigr", "migrat","migrant", "migrier", "einwander", "zuwander", "zugewander","eingewander", "flüchtling", "asyl", "geflücht","gastarbeit", "ausländ", "schutzsuch", "vertrieben", "balkanroute", "integration", "integrier", "assimil", "multikult", "syrer", "syrien", "afghan", "irak", "nahosten", "naher osten", "nahen osten", "aussengrenz", "außengrenz", "abschieb", "herkunftsstaat", "herkunftsland", "herkunftsländ", "zurückweisung", "rückführung", "lesbos", "zuzug", "zugezog", "islam", "muslim", "moslem"

Translated keyword string to English:
immigra, migr, refugee, asylum, guest worker, foreign, displaced, balkan route, integration, integrate, assimilate, multicultural, syria, afghan, iraq, middle east, deportation, country of origin, homeland, repatriation, lesbos, influx, newcomer, islam, muslim

Online Appendix B2: Transformer Model

Transformer models revolutionized natural language processing by introducing a novel architecture that eschews traditional recurrent layers in favor of self-attention mechanisms and positionally encoded inputs. This design allows the model to process entire sequences of data simultaneously, a stark departure from the sequential processing of older architectures. The self-attention mechanism is the cornerstone of the transformer, enabling the model to dynamically weigh the significance of different parts of the input data in relation to each other. For instance, in a sentence, the model can assess the importance and relationship of each word to every other word, capturing nuances of context, syntax, and semantics more effectively.

The transformer architecture is composed of an encoder and a decoder, each consisting of multiple layers that perform specific functions. The encoder layers work to encode the input text into a high-dimensional space, a representation that captures the essence and context of the input text. On the other side, the decoder follows a similar structure but is designed for generating output sequentially. This architecture enables transformers to perform a wide range of tasks with remarkable efficiency and accuracy, from understanding the sentiment of a text to translating languages and generating human-like text.

Fine-tuning a pre-trained transformer model is an efficient method widely used in natural language processing (NLP) for classification tasks. This approach starts with a model that has been pre-trained on a large dataset of text, often in a general context, to learn a wide range of features or patterns. The pre-trained model serves then as a starting point, containing a rich representation of the data it was trained on. By fine-tuning a pre-trained transformer model on a human-annotated dataset, which includes various forms of political communication annotated for moral and emotions appeals, Simonsen and Widmann (2023) tailored the model's capabilities to identify and measure these appeals within new, unseen texts.

The transformer model used in this study is a fine-tuned multilingual transformer model, specifically a mDeBERTa (He, Gao, and Chen 2023) model, which has been pre-trained on text data sourced from 100 languages (Conneau et al. 2020), allowing the model to classify text in a variety of different languages.

For the fine-tuning process, Simonsen and Widmann (2023b) utilized a dataset comprising over 20,000 sentences from political communications in six languages, which have been annotated for moral and emotional appeals by crowd-coders. Each sentence received evaluations from a minimum of five individuals, ensuring a robust assessment of its moral and emotional implications. Sentences have been classified based on a majority decision, i.e., at least three out of five coders needed to agree in their assessment of a given sentence. The model's effectiveness was then measured using a separate test set, employing metrics the F1 score, precision, and recall—key indicators for the performance of classification algorithms in machine learning. Precision is the ratio of correctly predicted sentences to the total predicted sentences, indicating the number of false positives. Recall is the ratio of correctly predicted sentences to the total number of true sentences, thereby indicating the number of false negatives. F1 score is the harmonic mean of precision and recall, which provides an overall assessment of the algorithm's accuracy in identifying both positive and negative instances. Table 1 reports the different performance metrics.

Table B1. Performance Metrics for German for the Fine-Tuned Transformer Model as reported by Simonsen and Widmann (2023b)

	Macro Precision	**Macro Recall**	**Macro F1**	***N* sentences**
Moral	0.71	0.70	0.70	496
Emotion	0.71	0.68	0.69	496
Positive	0.78	0.79	0.79	496
Negative	0.77	0.74	0.75	496

As explained in the main text, we use the model classification to create our variables of interest. For instance, positive moral appeals are tweets identified as moral and positive, without concurrent emotional appeals. Following the same logic, negative moral-emotional appeals are tweets identified as moral, emotional, and negative.

Online Appendix C: Optimal lag length

We use the VARselect function from the vars package (Pfaff 2008) to determine optimal lag length. Tables C1 and C2 presents the results for our different vector autoregression models, both for the simple (Figure 1) and the complex set up (Figure 2). For this test, we estimate each model (using different rhetorical appeals) with up to 7 lags and report the results below. The output suggests several criteria for determining the optimal number of lags.

AIC (Akaike Information Criterion) is a measure used to compare models while considering both the goodness of fit and the simplicity of the model. It penalizes models for having too many lags (parameters), thus helping to avoid overfitting. A lower AIC value indicates a better model fit. SC (Schwarz Criterion), also known as BIC (Bayesian Information Criterion), is another model selection criterion that heavily penalizes the number of parameters in the model, more so than AIC. This results in a preference for simpler models unless the additional lags significantly improve the model fit.

As can be seen, for most models, the recommended lag structure is $p = 1$. We therefore use a 1 lag structure in the main analysis and replicate our findings also with a 4-lag structure (see Online Appendix E).

Table C1: Optimal lag structure for simple VAR models

	AIC	SC
Positive Moral	1	1
Negative Moral	1	1
Negative Moral-Emotional	4	1
Positive Moral-Emotional	4	1

Table C2: Optimal lag structure for complex VAR models

	AIC	SC
Positive Moral	1	1
Negative Moral	1	1
Negative Moral-Emotional	1	1
Positive Moral-Emotional	1	1

Online Appendix D: Descriptives

For each rhetorical style, Figures D1 to D4 displays the mean proportion of immigration tweets for each party family and the public. As can be seen, the radical right displays the highest level of negative moral-emotional appeals, whereas for other rhetorical styles other party families show higher average levels. This confirms our theorization that the radical right challenger party AfD might engage in rhetorical innovation by combining moral appeals with negative emotional appeals. This finding is generally in line with previous research showing that radical challenger parties are more negative in their political communication than established mainstream parties (Widmann 2021). We add nuance to this research by showing that the negativity of the AfD only holds for negative moral-emotional rhetoric, not for negative moral rhetoric devoid of emotional appeals (compare Figures D2 and D4). Here, the radical right is on a comparable level to all other party families. Thus, the moral-emotional rhetoric seems uniquely connected to radical-right communication.

Figure D1: Levels of positive moral appeals across groups

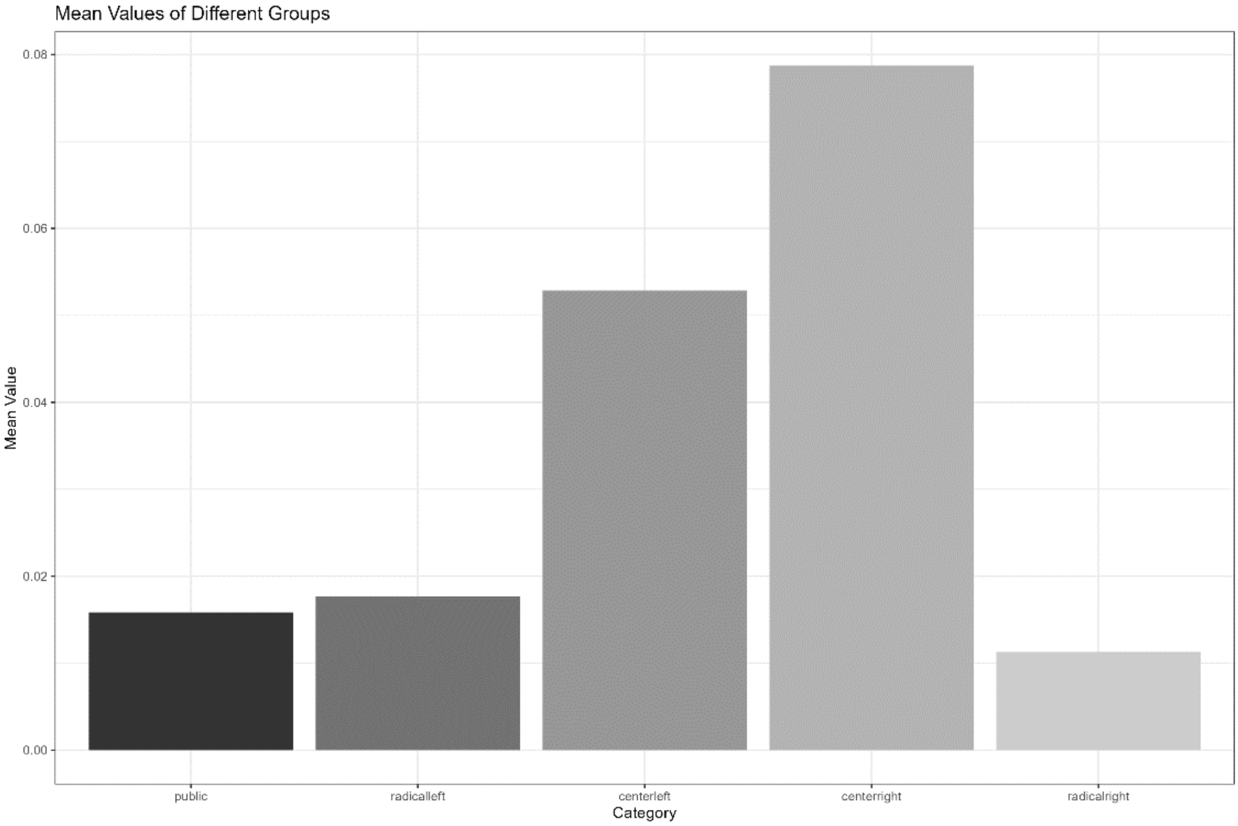


Figure D2: Levels of negative moral appeals across groups

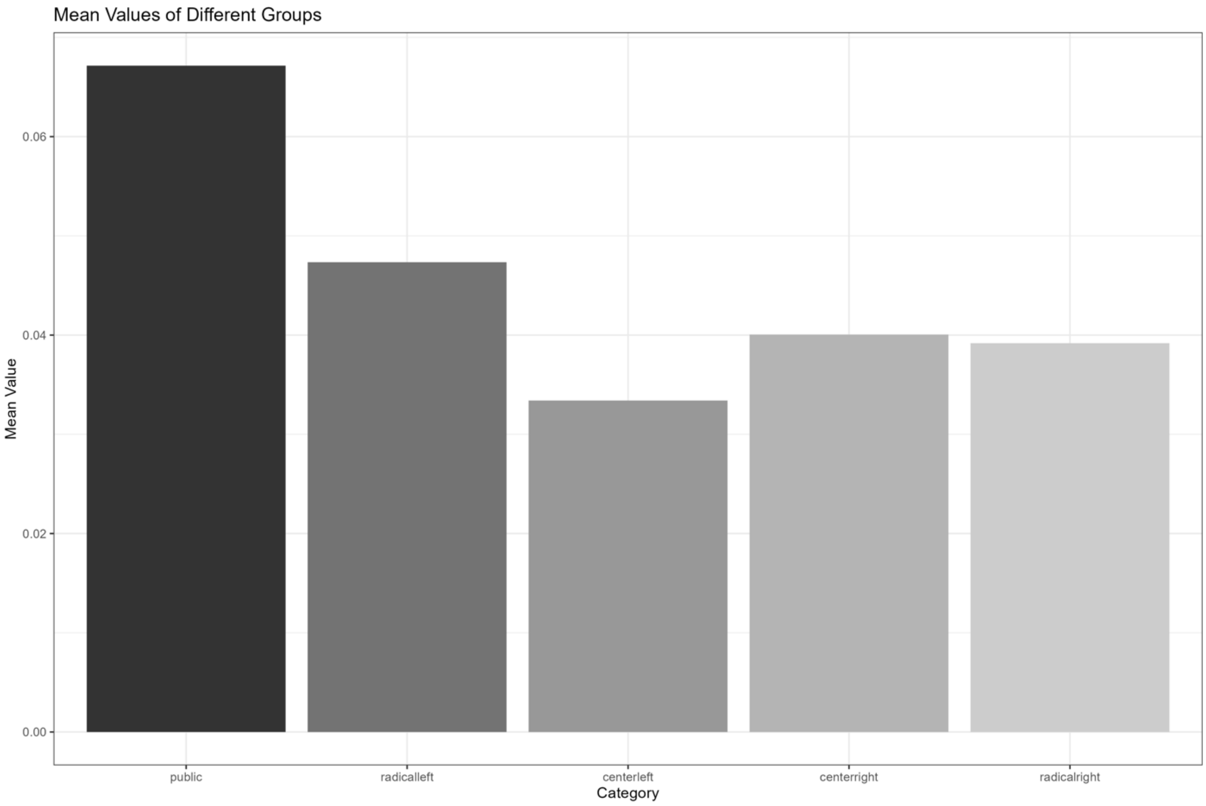


Figure D3: Levels of positive moral-emotional appeals across groups

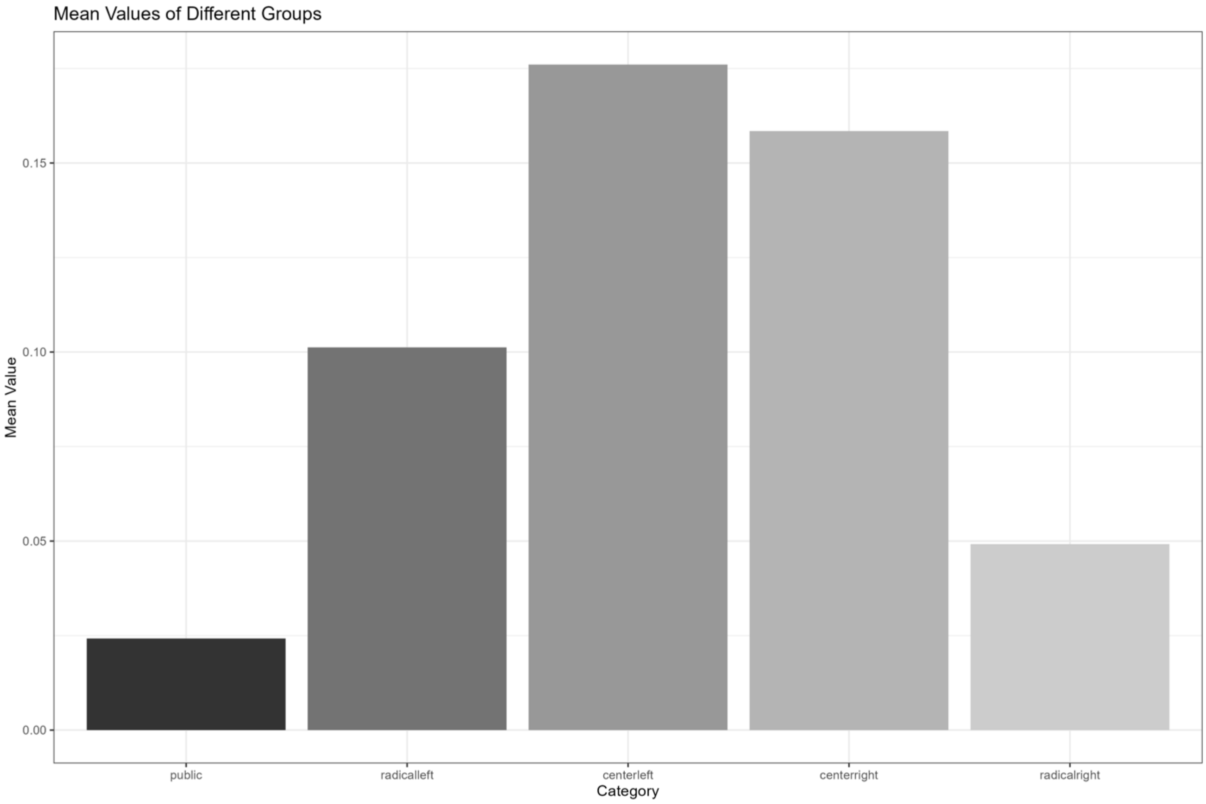
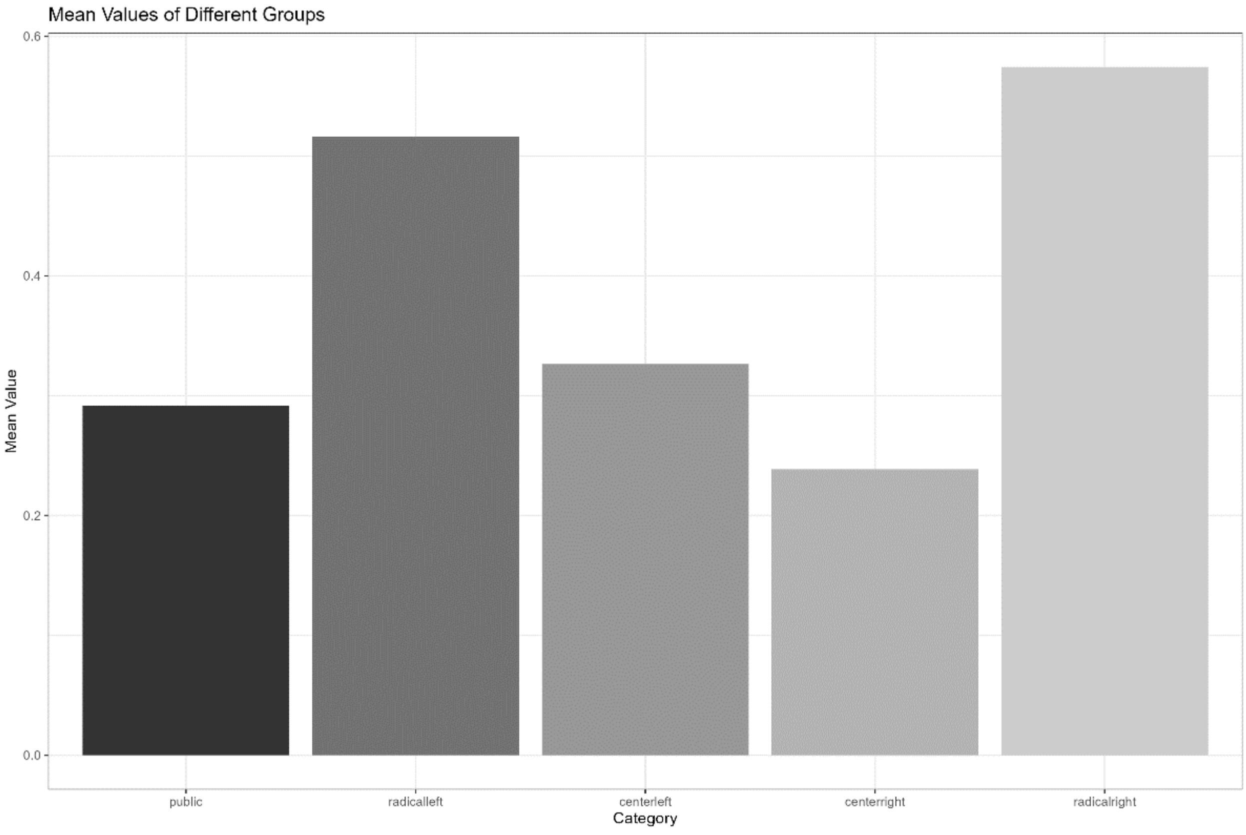


Figure D4: Levels of negative moral-emotional appeals across groups



Offering insight into the overall presence and variability in our rhetorical phenomena, Table D1 presents mean and standard deviations for the different rhetorical styles by actor group, and Figures D5 and D6 display corresponding histograms for each rhetorical style for the politics and public groups, respectively.

Table D1: Mean and standard deviations (SD) by actor group

	M Positive	M Negative	ME Positive	ME Negative
Mean Public	0.016	0.067	0.024	0.292
SD Public	0.024	0.047	0.028	0.103
Mean Politics	0.035	0.039	0.110	0.430
SD Politics	0.027	0.026	0.053	0.100

In line with the SVAR analysis, which looks at the response over the first 7 days after the shock, these histograms display the proportion of immigration tweets containing each rhetorical style per week. As can be seen, there is considerable variation within these variables, indicating that the usage of various moral and moral-emotional appeals is not stable but fluctuating.

Figure D5: Histograms for the politics group

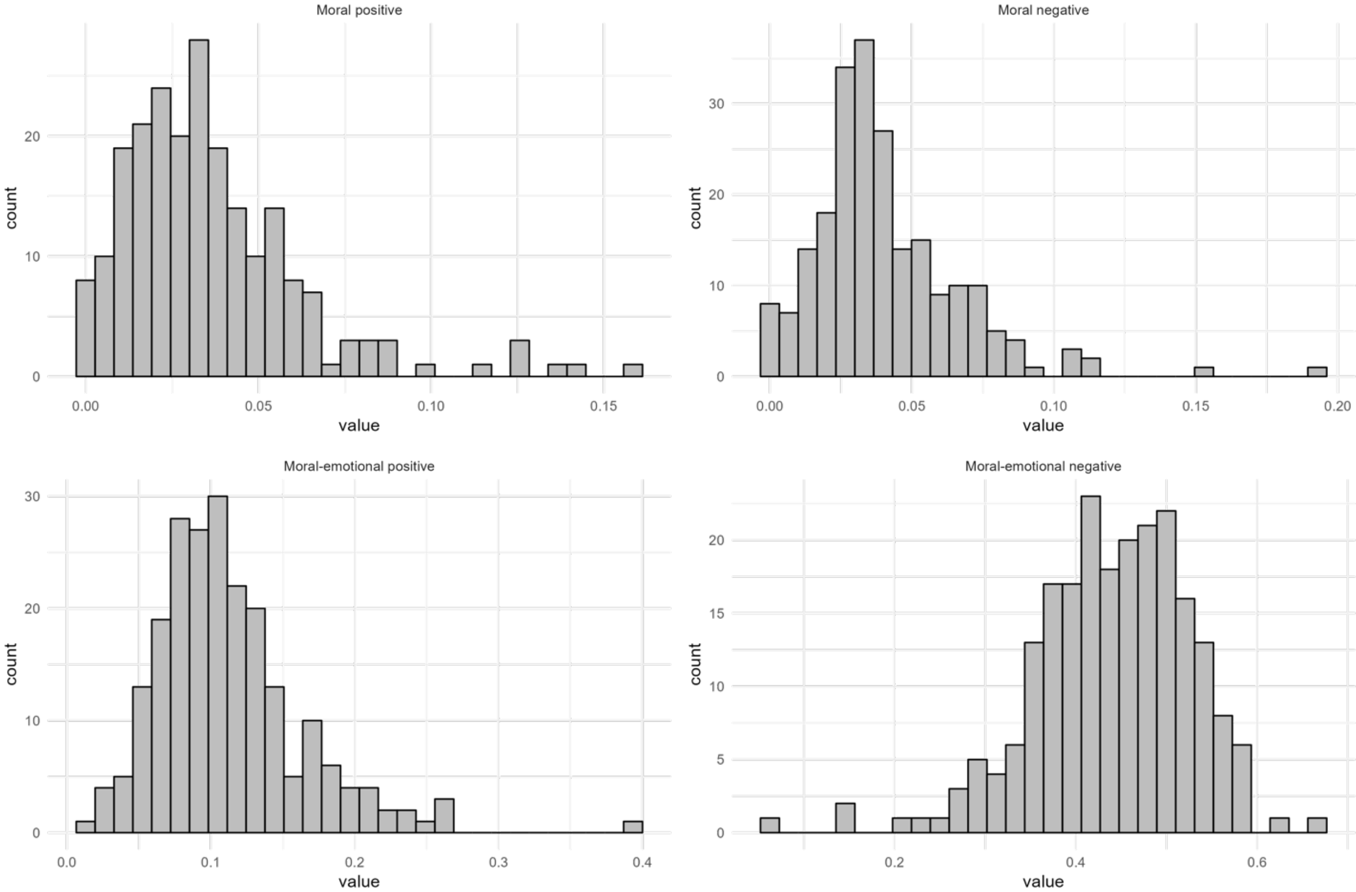
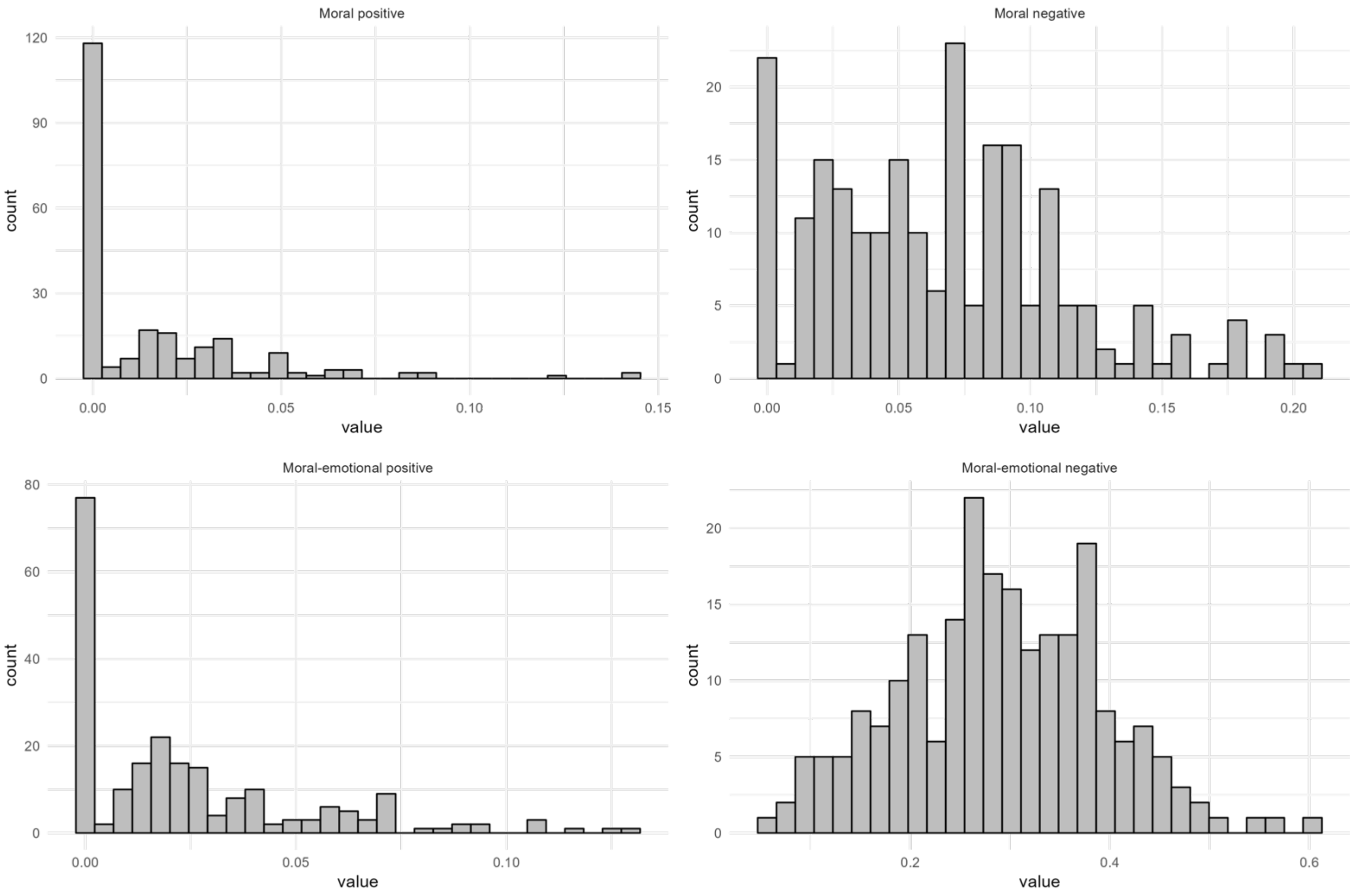


Figure D6: Histograms for the public group



Online Appendix E: Replication Using Constant Lag Order of 4

The lag order of the structural vector autoregression models presented in the main text was determined by the VARselect function of the vars package (Pfaff 2008), taking into considerations different criteria (as laid out in Online Appendix C). This procedure resulted in an optimal lag structure of p = 1, for all models in Figure 1 (main text). In Figure E1, we display the same estimates for a constant setting of 4 lags in order to examine their robustness for varying lag specifications. The results are consistent with the evidence reported in the paper. The effect size for Politics -> Public for Moral-Emotional Negative are even larger than reported in the main paper.

Figure E1: Cumulative IRFs: Predicted rhetorical appeals across groups (p = 4)

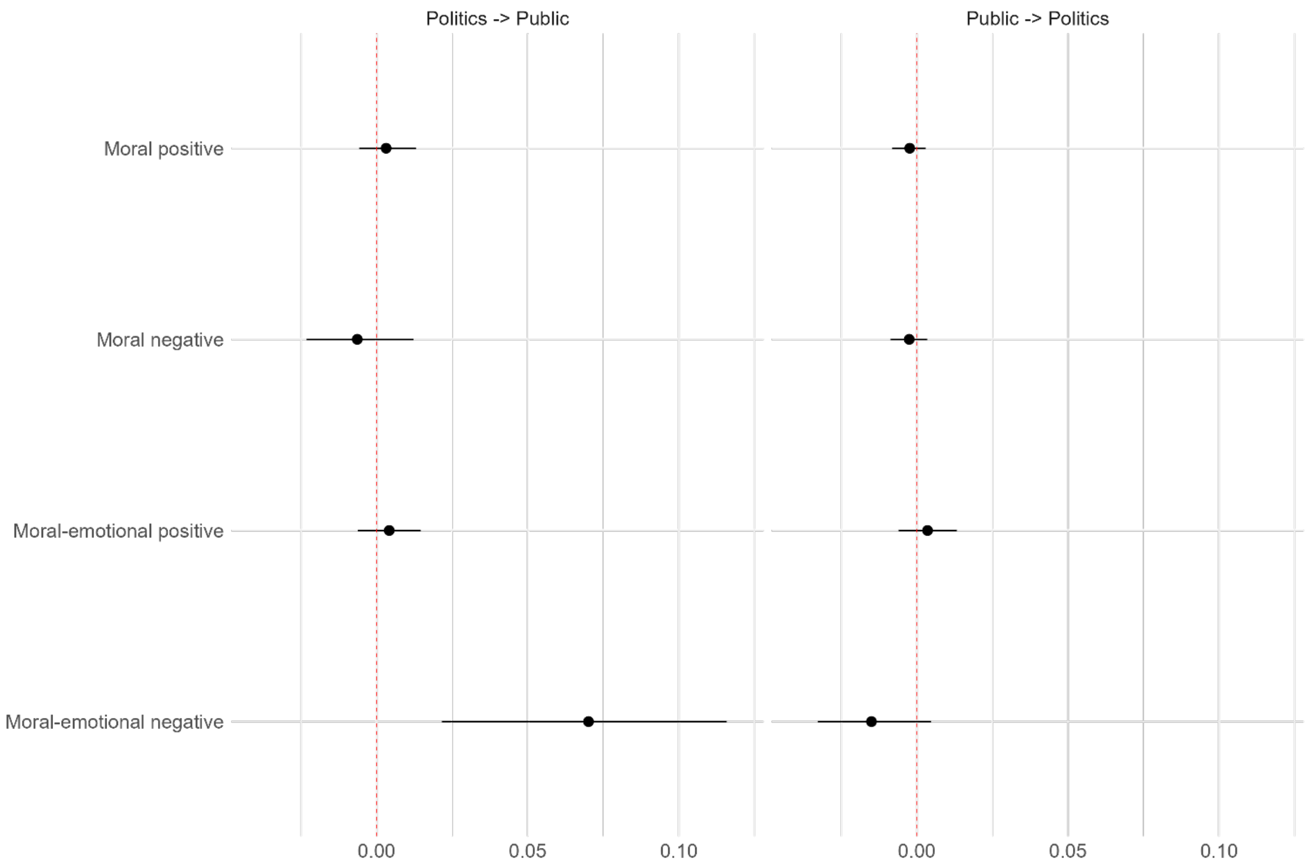


Figure E1: IRFs (with 1000 simulation runs) are based on four SVAR models (one for each rhetorical style) with a 4-day lag structure, illustrating the predicted, cumulative change over the seven subsequent days in public discourse after a shock on day 0 in political discourse (left panel) and in political discourse after a shock on day 0 in public discourse (right panel).

We also replicate the complex SVAR models (Figure 2) from the main text using a constant setting of 4 lags. Figure E2 displays the results of this exercise. Again, the results are largely consistent with the evidence reported in the main text. Interestingly, the negative coefficient for negative moral appeals, which did not correspond to our theoretical expectations, does not reach conventional levels of statistical significance in this replication. Moreover, the effect of radical-right challenger party's use of negative moral-emotional appeals is larger than in the analysis (p = 1) reported in main text. These patterns confirm our confidence in the conclusion that H2 finds support in the data.

Figure E2: Cumulative IRFs: Predicted rhetorical appeals across groups (p = 4)

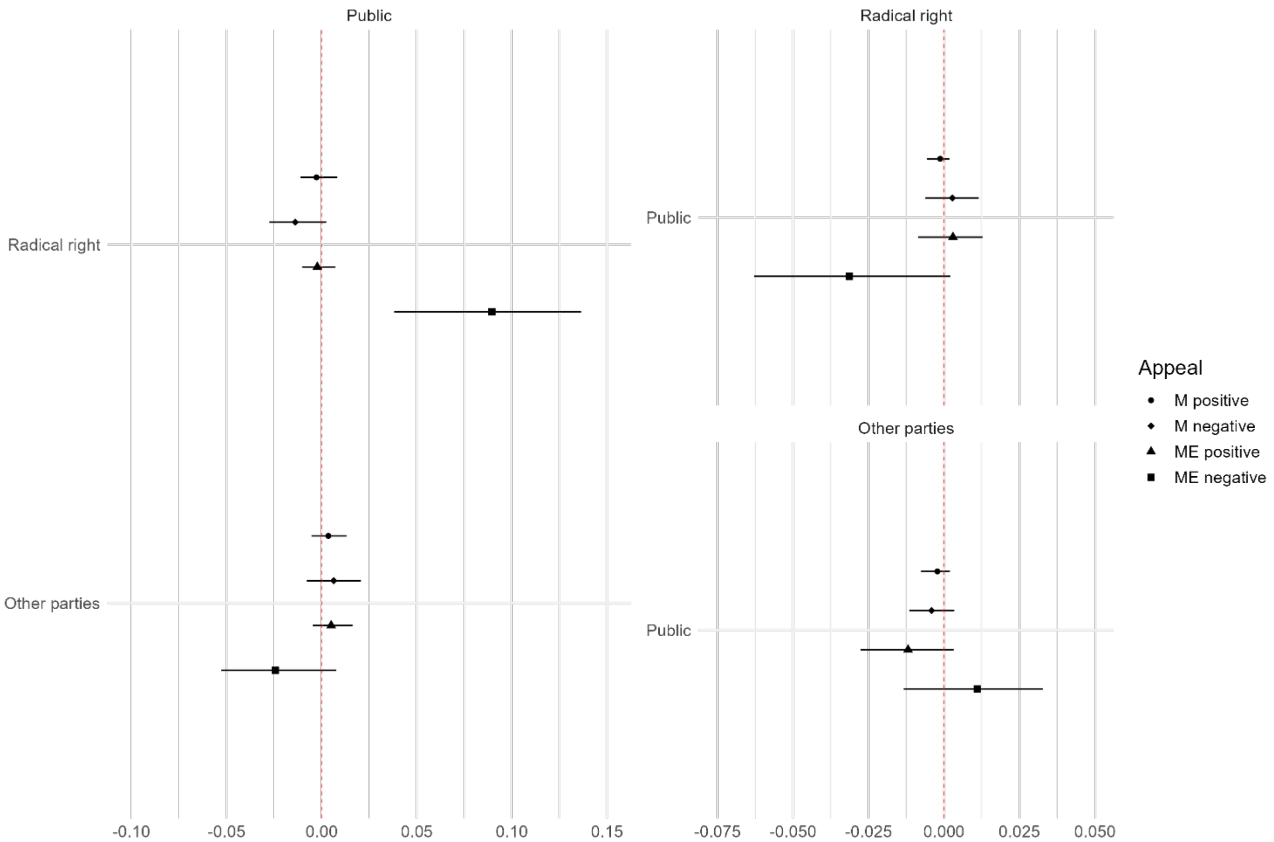


Figure E2: IRFs (with 1000 simulation runs) are based on four SVAR models (one for each rhetorical style), illustrating the predicted, cumulative change over the seven subsequent days in public discourse after a shock on day 0 in different party families' discourse (left panel) and in party families' discourse after a shock on day 0 in public discourse (right panel), i.e. y-axis denotes the impulse group and the title denotes the response group.

Online Appendix F: Equation balance

We follow the recommendation by Pickup and Kellstedt (2023) and show that our vector autoregression models including different rhetorical styles in different actor groups result in balanced equations. To do so, we follow previous work with similar research designs (e.g. Kraft and Newman 2023) and firstly show that the time series under consideration (i.e., different rhetorical styles for different actor groups) are stationary. The Augmented Dickey-Fuller (ADF) test is a statistical test used to determine whether a time series is stationary by testing for the presence of a unit root. A significant p-value (typically <0.05) indicates strong evidence against the null hypothesis of a unit root, suggesting that the series is stationary and does not depend on time. Table F1 presents the results of the ADF test for the different time series for the simple VAR models, including three actor groups and the two different lag structures. Table F2 presents the same for the complex VAR models with more actor groups.

Table F1: ADF test results for simple VAR models including 3 actor groups

Appeal	Actor Group	Lags	Statistic	*p*-Value
Positive Moral	Politics	1	-25.096	< 0.01
	Politics	4	-16.103	< 0.01
	News media	1	-27.023	< 0.01
	News media	4	-17.672	< 0.01
	Attentive Public	1	-28.283	< 0.01
	Attentive Public	4	-18.759	< 0.01
Negative Moral	Politics	1	-26.687	< 0.01
	Politics	4	-16.584	< 0.01
	News media	1	-26.526	< 0.01

	News media	4	-17.328	< 0.01
	Attentive Public	1	-28.314	< 0.01
	Attentive Public	4	-17.782	< 0.01
Positive Moral-Emotional	Politics	1	-26.473	< 0.01
	Politics	4	-15.473	< 0.01
	News media	1	-26.857	< 0.01
	News media	4	-16.695	< 0.01
	Attentive Public	1	-27.329	< 0.01
	Attentive Public	4	-16.759	< 0.01
Negative Moral-Emotional	Politics	1	-21.459	< 0.01
	Politics	4	-12.351	< 0.01
	News media	1	-26.857	< 0.01
	News media	4	-17.533	< 0.01
	Attentive Public	1	-26.104	< 0.01
	Attentive Public	4	-16.153	< 0.01

Table F2: ADF test results for complex VAR models

Appeal	Actor Group	Lags	Statistic	p-Value
Positive Moral	AfD	1	-26.831	< 0.01
	AfD	4	-15.861	< 0.01
	Other Parties	1	-26.88	< 0.01
	Other Parties	4	-16.074	< 0.01
Negative Moral	AfD	1	-26.831	< 0.01
	AfD	4	-15.861	< 0.01
	Other Parties	1	-26.88	< 0.01
	Other Parties	4	-16.074	< 0.01
Positive Moral-Emotional	AfD	1	-26.831	< 0.01
	AfD	4	-15.861	< 0.01
	Other Parties	1	-26.88	< 0.01
	Other Parties	4	-16.074	< 0.01
Negative Moral-Emotional	AfD	1	-26.831	< 0.01
	AfD	4	-15.861	< 0.01
	Other Parties	1	-26.88	< 0.01
	Other Parties	4	-16.074	< 0.01

These results suggest that the model specifications employed in both the main text and the Online Appendix (OA) are balanced, with all variables in the vector autoregressions demonstrating stationarity (Pickup and Kellstedt 2023). Consequently, our models are expected to have stable error terms, allowing for the use of conventional test statistics for inference purposes.

Secondly, similarly to Kraft and Newman (2023), we demonstrate that our model specification is balanced by presenting the time series of residuals for the main results presented in Figures 1 and 2 of the main text. Balanced equations should result in white noise residuals (Pickup and Kellstedt, 2023) and Figures F1 and F2 confirm that the residuals of each autoregression equation are stationary (Figure F1 for the simple VAR models and Figure F2 for the complex VAR models).

Figure F1: Time series of vector autoregression residuals for daily changes in rhetorical styles for different actor groups, based on the main model presented in Figure 1

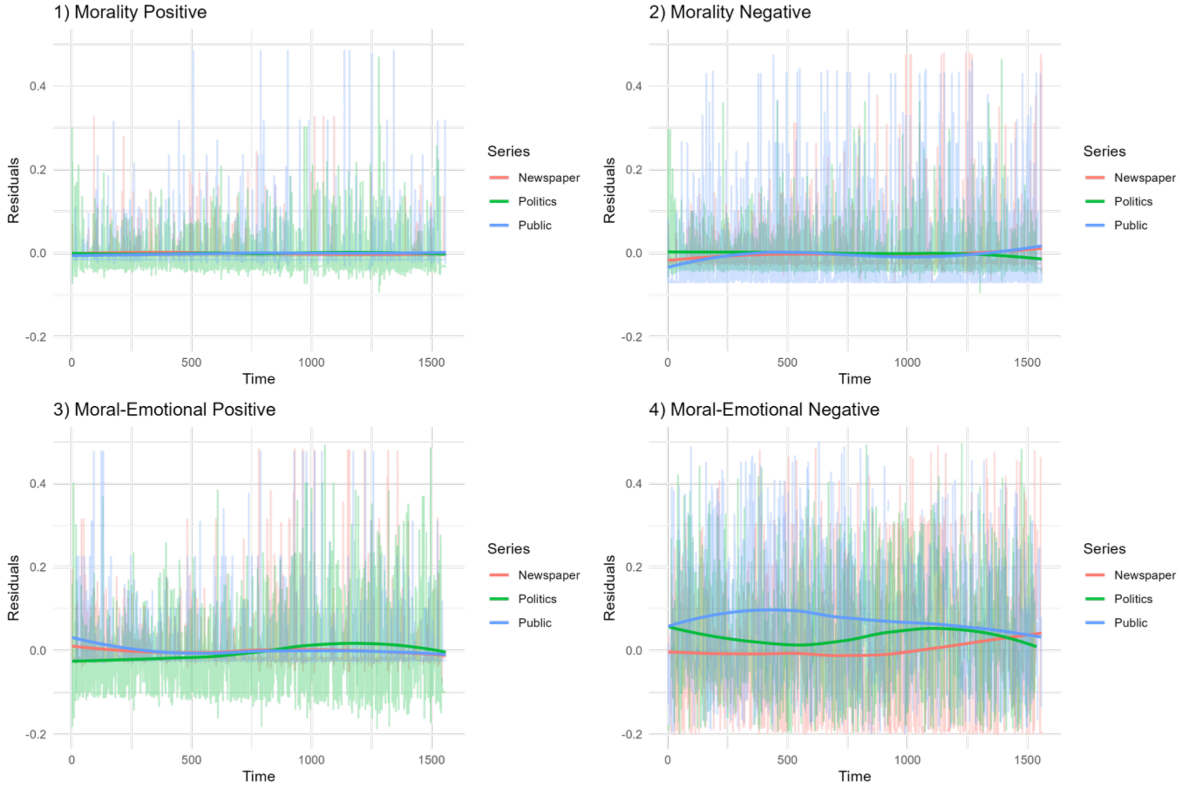
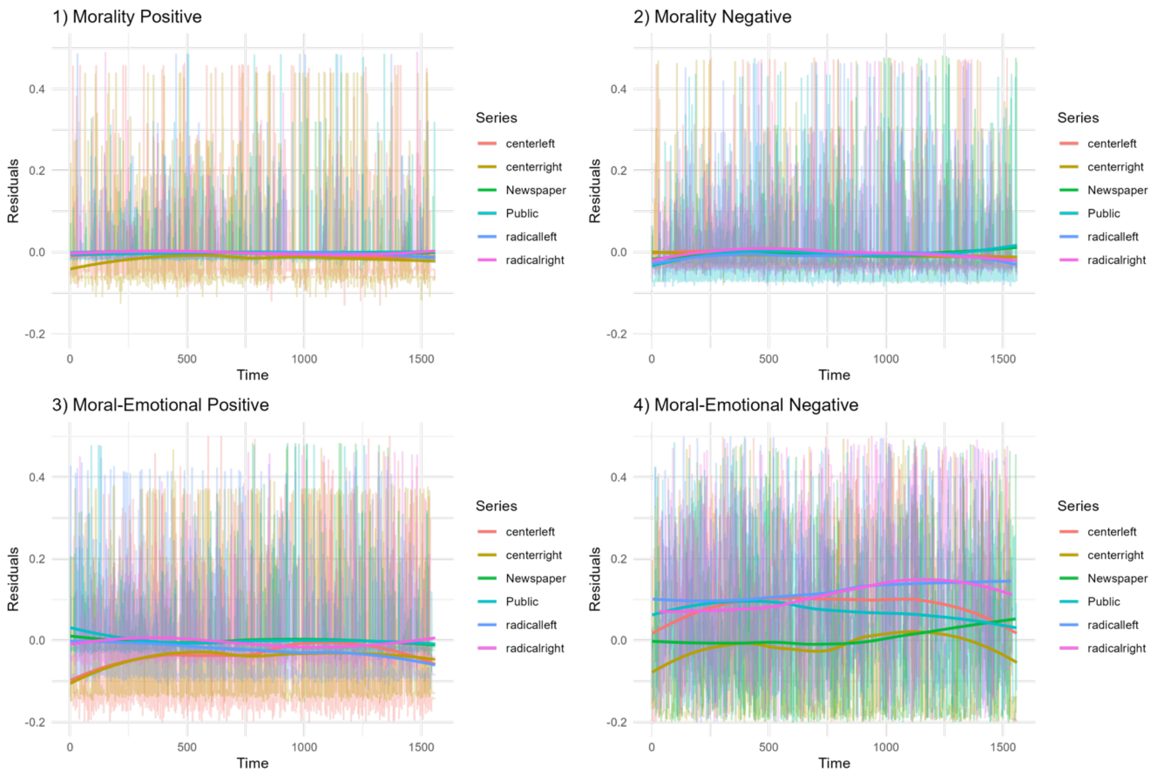


Figure F2: Time series of vector autoregression residuals for daily changes in rhetorical styles for different actor groups, based on the main model presented in Figure 2



Figures F3 to F6 (based on the simple model presented in Figure 1) further show that no significant autocorrelation is present, and the residuals can therefore be characterized as white noise. In sum, these results establish that our estimation strategy is based on balanced equations that allow for valid inferences regarding the short-term relationship between daily changes in rhetorical styles between actor groups.

Figure F.3: Autocorrelations of vector autoregression residuals for daily changes in moral positive appeals by actor groups, based on the main model presented in Figure 1 of the main text

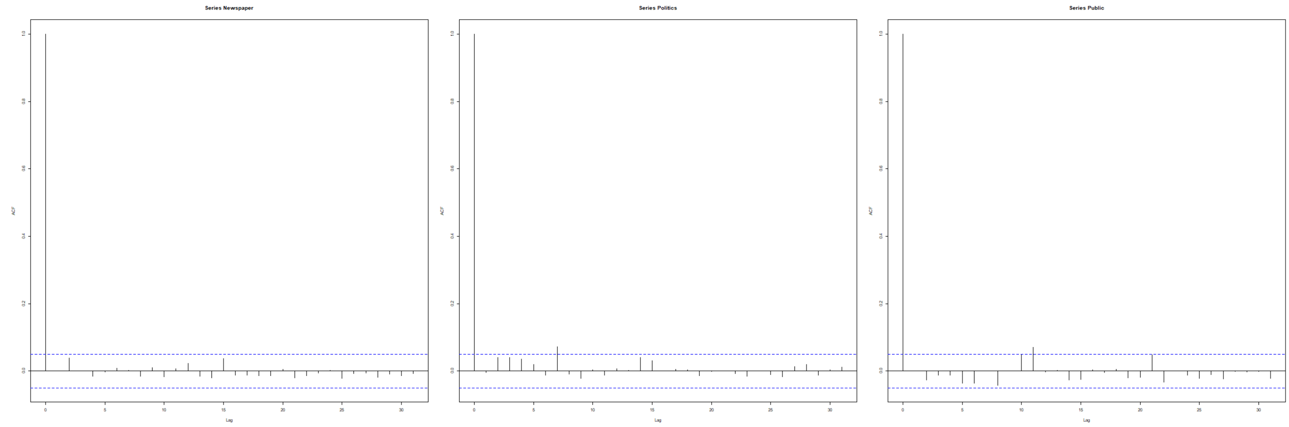


Figure F.4: Autocorrelations of vector autoregression residuals for daily changes in moral negative appeals by actor groups, based on the main model presented in Figure 1 of the main text

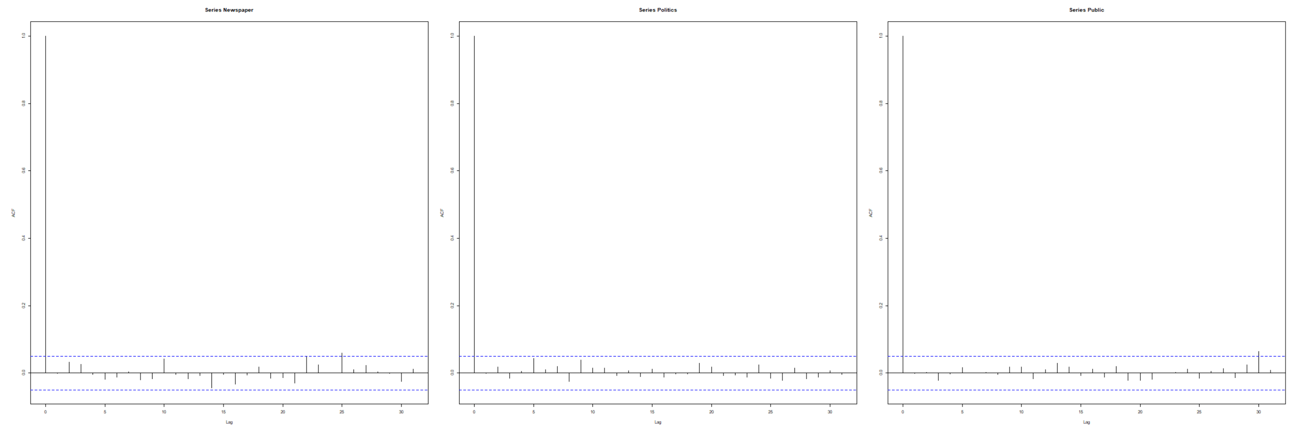


Figure F.5: Autocorrelations of vector autoregression residuals for daily changes in moral-emotional positive appeals by actor groups, based on the main model presented in Figure 1 of the main text

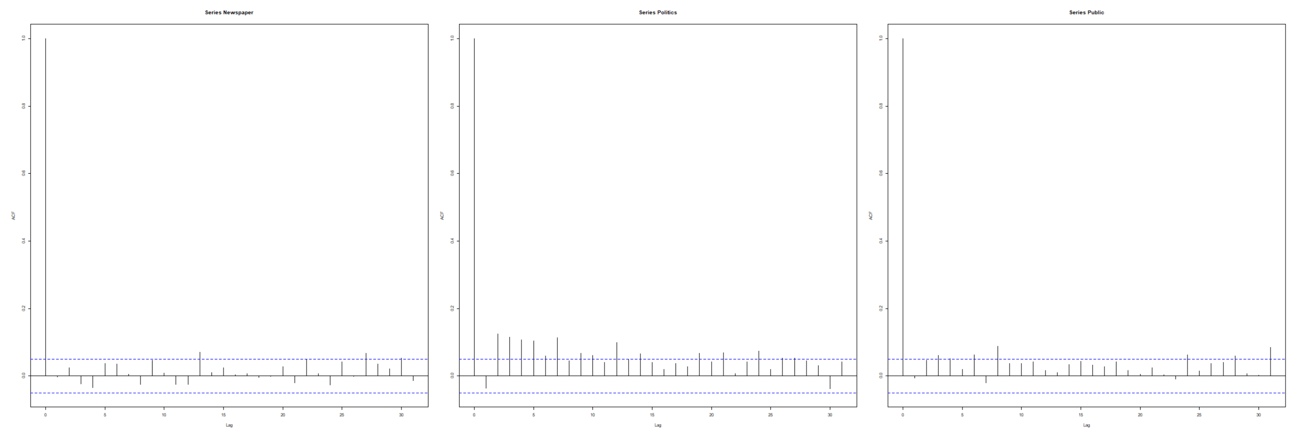
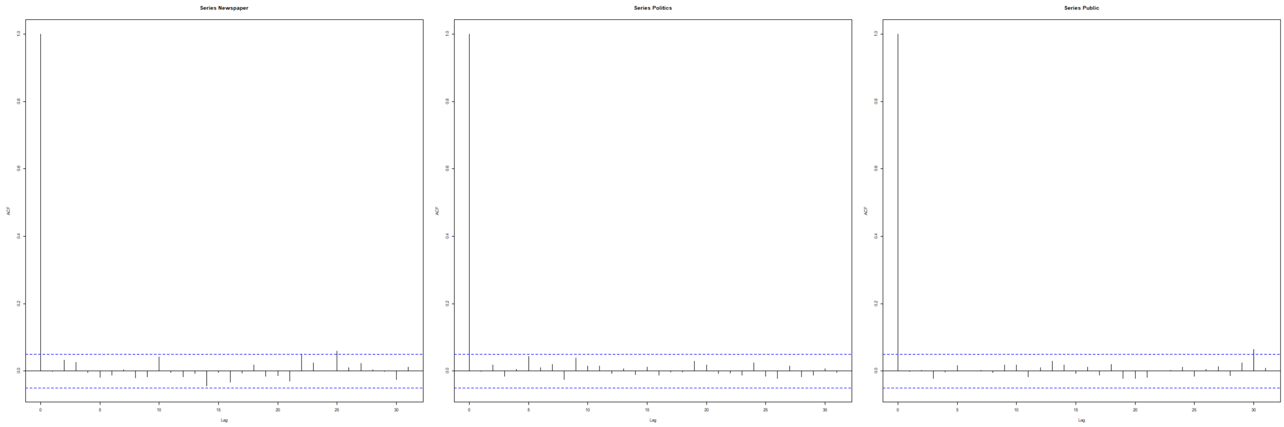


Figure F.6: Autocorrelations of vector autoregression residuals for daily changes in moral-emotional negative appeals by actor groups, based on the main model presented in Figure 1 of the main text



Lastly, Figures F7 to F10 show the same for the complex models.

Figure F.7: Autocorrelations of vector autoregression residuals for daily changes in moral positive appeals by actor groups, based on the main model presented in Figure 2 of the main text

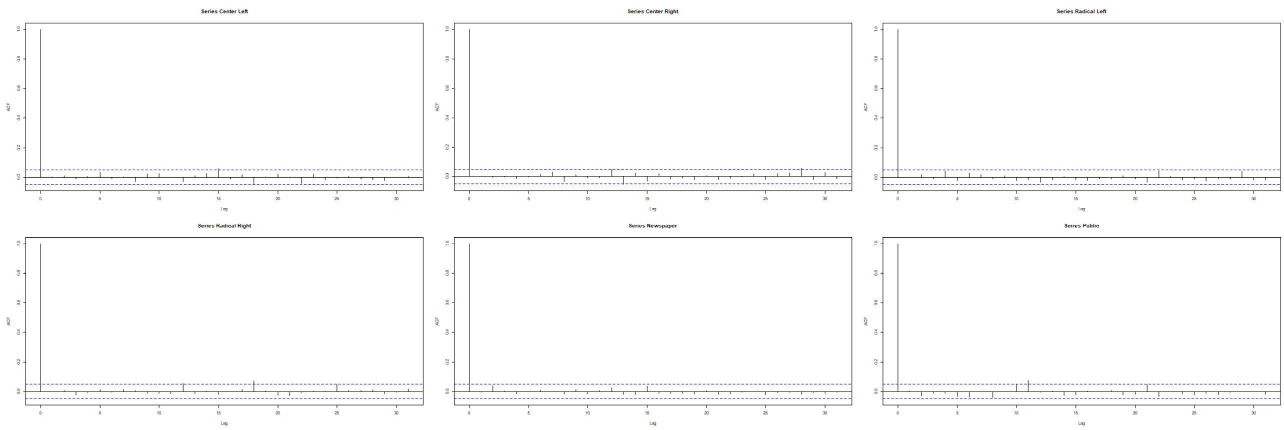


Figure F.8: Autocorrelations of vector autoregression residuals for daily changes in moral negative appeals by actor groups, based on the main model presented in Figure 2 of the main text

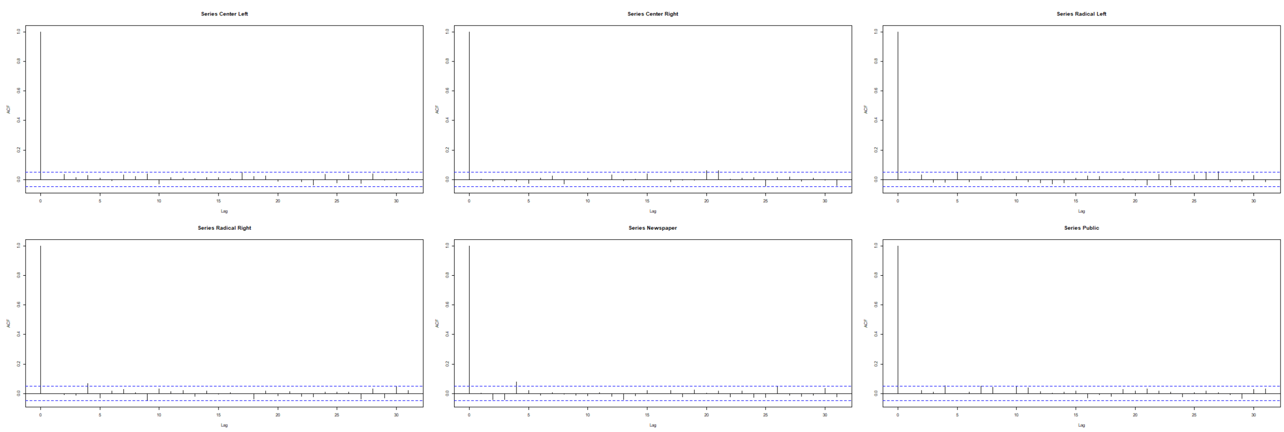


Figure F.9: Autocorrelations of vector autoregression residuals for daily changes in moral-emotional positive appeals by actor groups, based on the main model presented in Figure 2 of the main text

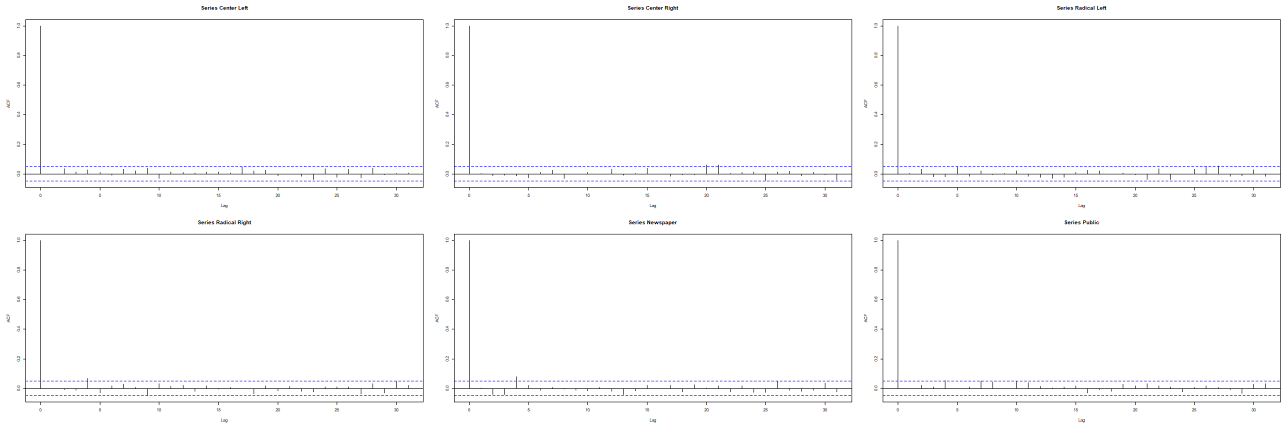
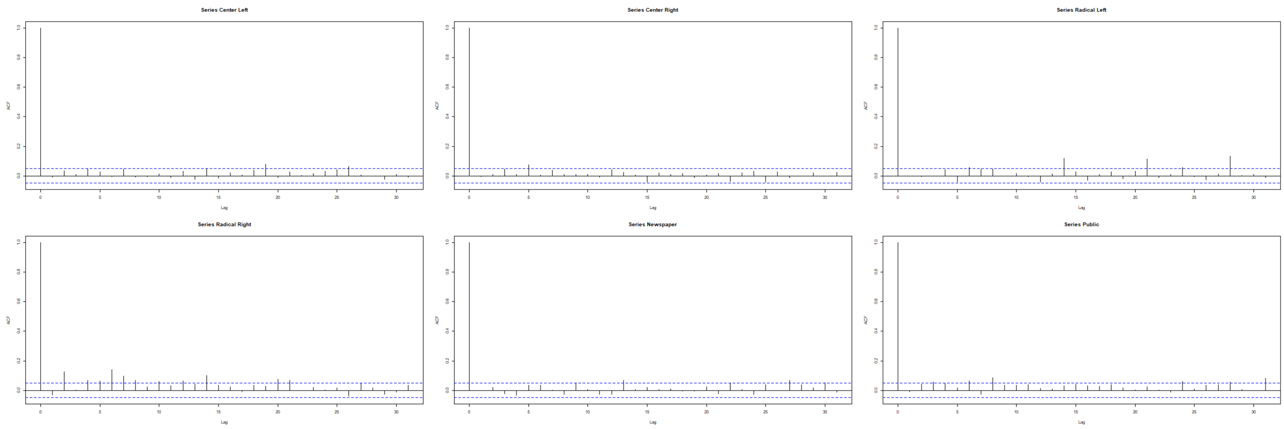


Figure F.10: Autocorrelations of vector autoregression residuals for daily changes in moral-emotional negative appeals by actor groups, based on the main model presented in Figure 2 of the main text



Online Appendix G: Full SVAR models (inter-party effects)

Figure G1 replicates Figure 2 in the main text but using party families instead of a binary split between the AfD and all other parties. Thus, Figure G1 shows the effects of radical-left, center-left, center-right, and radical-right parties on the public, and vice versa.

Figure G1: Cumulative IRFs: Predicted rhetorical appeals across party families

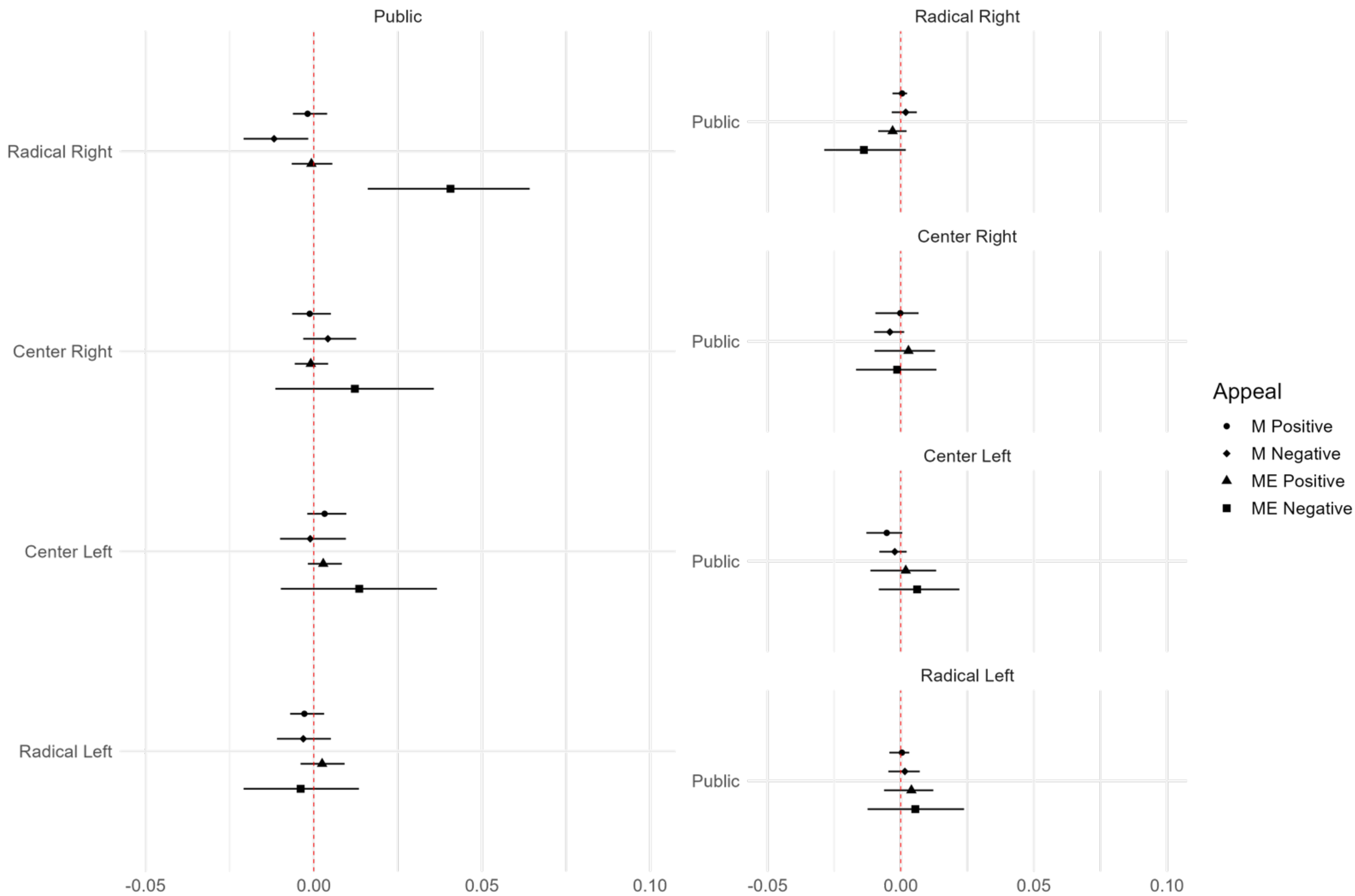


Figure G1: IRFs (with 1000 simulation runs) are based on four SVAR models (one for each rhetorical style) with a 1-day lag-structure, illustrating the predicted, cumulative change over the seven subsequent days in public discourse after a shock on day 0 in different party families' discourse (left panel) and in party families' discourse after a shock on day 0 in public discourse (right panel), i.e. y-axis denotes the impulse group and the title denotes the response group.

As can be seen, the results remain the same with no other coefficient reaching statistical significance besides the radical-right challenger party's influence on the public.

Figure G2 shows additional relationships between all included actor groups, displaying all possible interdependencies between different party families and the public. As can be seen in the panel furthest to the right, the radical right challenger party remains the only party (family) exerting influence on the rhetorical style of the attentive public. However, party families also influence one another in their way of discussing immigration.

Figure G2: Cumulative IRFs: Predicted rhetorical appeals across party families showing inter-party effects

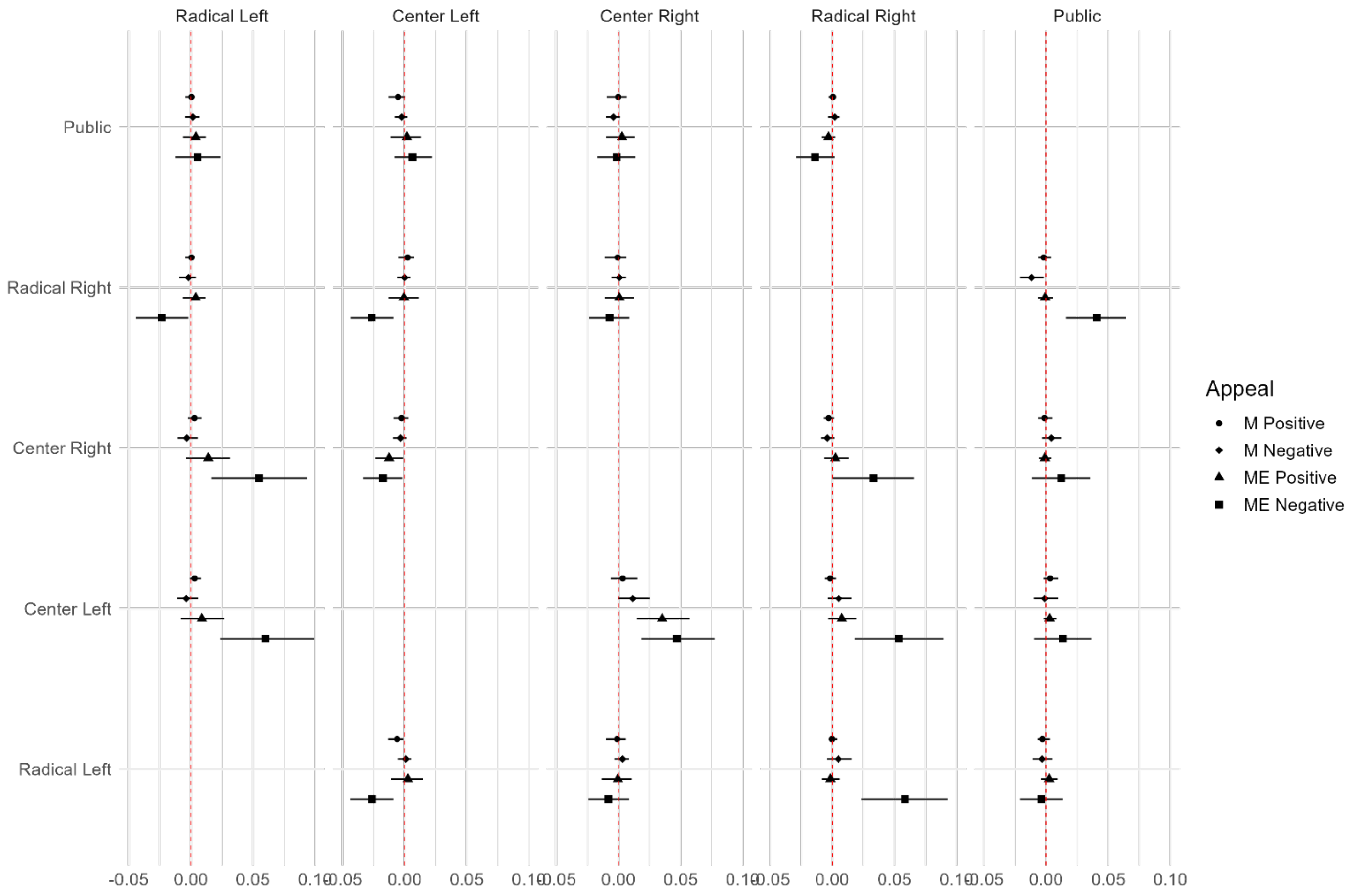


Figure G3: IRFs (with 1000 simulation runs) are based on four SVAR models (one for each rhetorical style) with a 1-day lag-structure, illustrating the predicted, cumulative change over the seven subsequent days in public discourse after a shock on day 0 in different party families' discourse (left panel) and in party families' discourse after a shock on day 0 in public discourse (right panel), i.e. y-axis denotes the impulse group and the title denotes the response group.

Online Appendix H: Replication without *Bild* newspaper

To alleviate concerns about inconsistencies between newspapers included in the media sample and the set of newspapers used to sample the attentive public, we replicate the main analysis here without including *Bild* newspaper in the media sample. Hence, we are including only *Die Welt* and *Süddeutsche Zeitung* in the media sample, the same newspapers we used to sample the attentive public. Figures H1 and H2 replicate Figures 1 and 2 from the main text. As can be seen, the results remain the same.

Figure H1: Cumulative IRFs: Predicted rhetorical appeals across public and politics group, without including BILD newspaper in the media sample

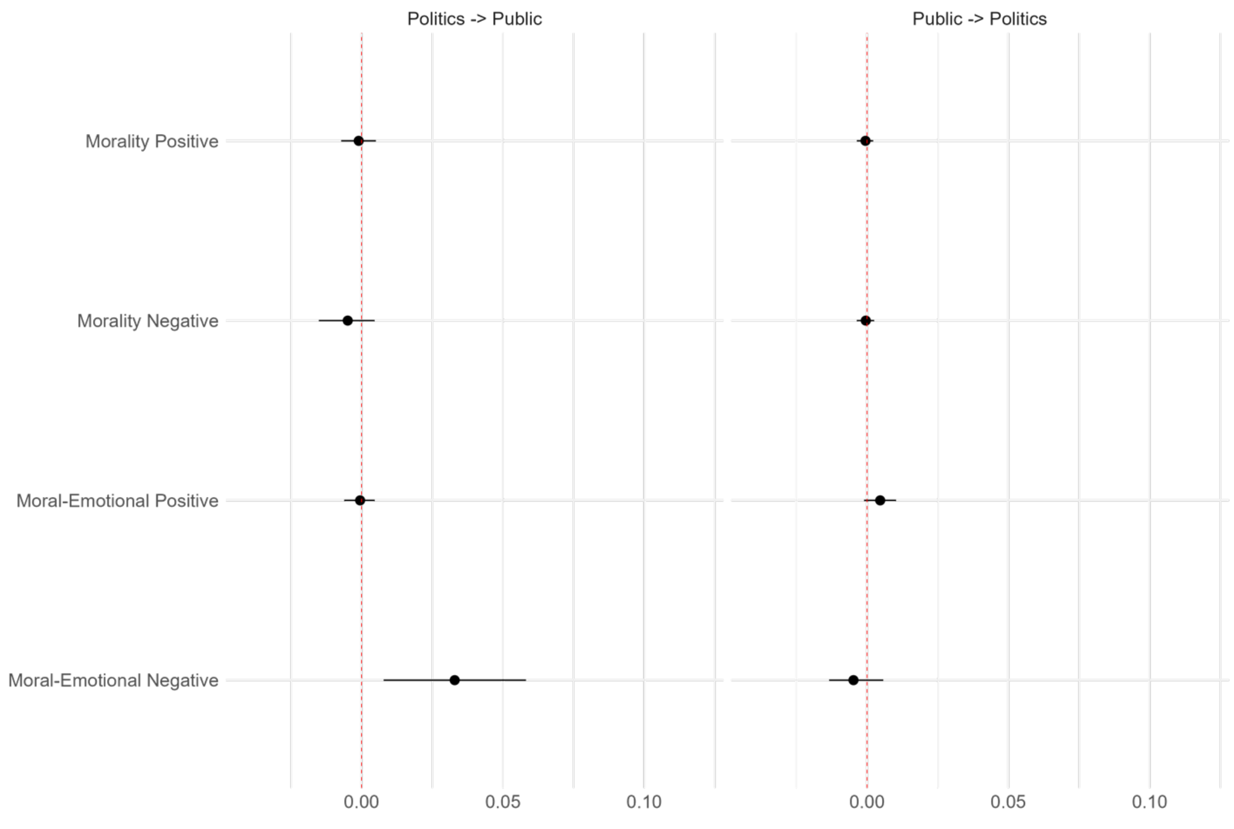


Figure H4: IRFs (with 1000 simulation runs) are based on four SVAR models (one for each rhetorical style) with a 1-day lag structure, illustrating the predicted, cumulative change over the seven subsequent days in public discourse after a shock on day 0 in political discourse (left panel) and in political discourse after a shock on day 0 in public discourse (right panel).

Figure H2: Cumulative IRFs: Predicted rhetorical appeals across different actor groups, without including BILD newspaper in the media sample

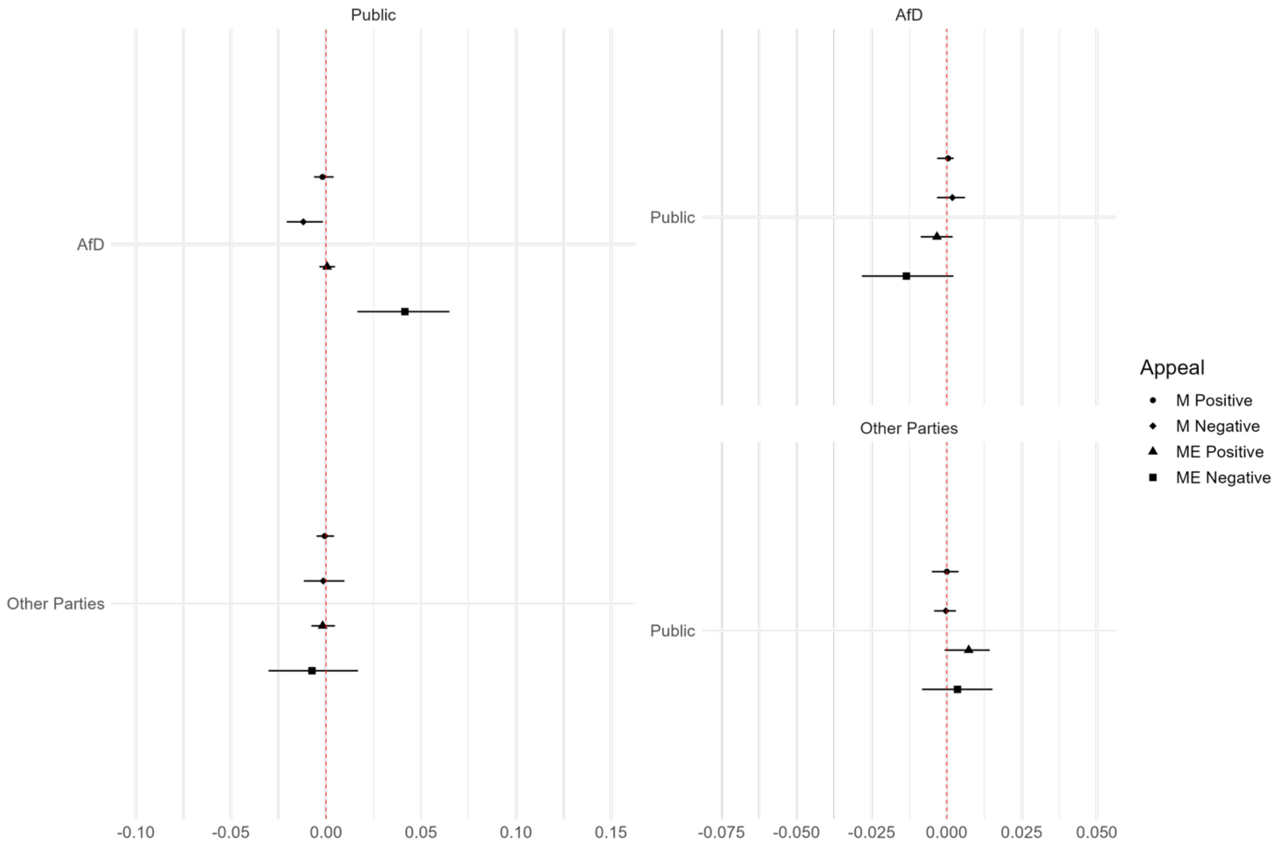


Figure 5: IRFs (with 1000 simulation runs) are based on four SVAR models (one for each rhetorical style) with a 1-day lag-structure, illustrating the predicted, cumulative change over the seven subsequent days in public discourse after a shock on day 0 in different party families' discourse (left panel) and in party families' discourse after a shock on day 0 in public discourse (right panel), i.e. y-axis denotes the impulse group and the title denotes the response group.

References Appendix

Barberá, Pablo, Andreu Casas, Jonathan Nagler, Patrick J. Egan, Richard Bonneau, John T. Jost, and Joshua A. Tucker. 2019. “Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data.” *American Political Science Review* 113 (4): 883–901. https://doi.org/10.1017/S0003055419000352.

Barrie, Christopher, and Justin Chun-ting Ho. (2021) 2021. “academictwitteR: An R Package to Access the Twitter Academic Research Product Track v2 API Endpoint.” R. https://github.com/cjbarrie/academictwitteR.

Kraft, Patrick W., and Benjamin J. Newman. 2023. “Complaints about Police Misconduct Have Adverse Effects for Black Civilians.” *Political Science Research and Methods*, October, 1–24. https://doi.org/10.1017/psrm.2023.49.

Pfaff, Bernhard. 2008. “VAR, SVAR and SVEC Models: Implementation Within *R* Package **Vars**.” *Journal of Statistical Software* 27 (4). https://doi.org/10.18637/jss.v027.i04.

Pickup, Mark, and Paul M. Kellstedt. 2023. “Balance as a Pre-Estimation Test for Time Series Analysis.” *Political Analysis* 31 (2): 295–304. https://doi.org/10.1017/pan.2022.4.

Simonsen, Kristina Bakkær, and Tobias Widmann. 2023. “The Politics of Right and Wrong: Moral Appeals in Political Communication over Six Decades in Ten Western Democracies.” https://doi.org/10.31219/osf.io/m6qkg.

Widmann, Tobias. 2021. “How Emotional Are Populists Really? Factors Explaining Emotional Appeals in the Communication of Political Parties.” *Political Psychology* 42 (1): 163–81. https://doi.org/10.1111/pops.12693.