# What Drives Legislative Anger? Multimodal Evidence from South Korea's National Assembly*

Haejo Kang† Songpo Yang‡

**Abstract**

Legislative anger is often understood as expressive representation, a signal that legislators passionately defend constituent interests. We argue that partisan blame underpins these displays, especially when parties attack one another over alleged misconduct. Using deep-learning models, we measure vocal and facial anger across 355 hours of plenary debate in South Korea's 20th National Assembly (2016–2020), covering 11,060 speech segments by 247 legislators. We find that anger peaks when legislators discuss scandal, at twice the level of policy debate. Moreover, a quasi-experiment around the impeachment of President Park Geun-hye shows that the same legislators become angrier in opposition and calmer in government. These findings challenge accounts that treat legislative anger primarily as constituency signaling and highlight how partisan blame and institutional position shape emotional expression in legislatures.

*Kang led the theoretical design, empirical analysis, and writing; Yang led the construction of the multimodal dataset and robustness analysis. Author order is alphabetical.

†Ph.D candidate, Department of Political Science, Columbia University. Email: hk3125@columbia.edu.

‡Boya Postdoctoral Fellow, School of International Studies, Peking University. Email: yangsp@pku.edu.cn.

# 1 Introduction

Anger is the most visible emotion in legislative politics. Legislators raise their voices on the floor, accuse one another in hearings, and express outrage for the cameras. These displays are consequential. They shape how voters evaluate politicians (Boussalis et al., 2021), and exposure to angry elites makes citizens themselves angrier (Stapleton and Dawkins, 2022). That anger feeds a mass politics in which hostility toward the opposing party has become a defining feature of partisanship (Iyengar et al., 2012; Mason, 2018). The anger of elites themselves, however, has been studied mainly from the angle of representation, as a display of passion for the voters watching. Yet anger, unlike enthusiasm or pride, takes a target, someone to hold responsible. When a legislator stands up and sounds angry, what is that anger for, and at whom is it aimed?

The representation literature builds on classic accounts of constituency influence and the electoral connection, in which the legislator acts as the district's agent (Miller and Stokes, 1963; Mayhew, 1974; Fenno, 1978). Recent work extends the logic to emotion, showing that emotive rhetoric intensifies when more voters are watching and that electoral incentives reward displays of passion (Stout et al., 2025; Osnabrügge et al., 2021; Park, 2021a). On this reading, an angry speech is passionate advocacy, the legislator giving voice to what the district feels. Yet the same speech admits a second reading. The partisan competition literature shows legislators acting as members of party teams locked in conflict, attacking the other side in votes, obstruction, and rhetoric (Lee, 2009; Grimmer and King, 2011; Theriault, 2013). On this reading, an angry speech is an attack, aimed at the party across the aisle rather than delivered for the district. This second reading has not been put to the test: the contest has been measured; its emotions have not. And the two readings are hard to pry apart by observation alone, because an angry speech—often televised and clipped for social media—sends messages to the public and to fellow legislators at once. Whom those messages serve is where they come apart.

This article argues that anger is an instrument of the partisan contest, aimed at fellow legislators rather than displayed for voters. Legislators express anger when the contest with the opposing party gives them grounds for attack. Scandal supplies the clearest grounds. When misconduct,

criminal or not, attaches to identifiable political actors, the wrongdoing itself justifies the anger.[1] Institutional position supplies grounds of a second source of partisan ammunition. Opposition status provides grounds for attacking the incumbent party, whereas governing parties must defend their own record. If anger serves partisan contest rather than constituency signaling, the same legislator should become angrier when moving into opposition and less angry when moving into government.

The 20th National Assembly of South Korea (2016–2020) provides an ideal setting for testing this argument. It featured intense polarization between two rival blocs, a legislative term marked by repeated scandals, and an institutional shock that reversed governing status without a legislative election(Cheong and Haggard, 2023; Park, 2021b). The impeachment of President Park Geun-hye moved the Democratic Party from opposition into government and the conservative bloc in the opposite direction, allowing us to examine whether legislators' emotional expression changed when their partisan roles changed. The mixed-member electoral system provides a second source of leverage by placing district and non-district legislators side by side within the same legislature, allowing us to distinguish partisan incentives from constituency-based incentives.

Testing this argument requires measuring anger in a way that existing approaches cannot. Most work on legislative emotion measures the text of speeches, with some relying on survey responses. Both approaches are indirect: surveys capture what respondents report feeling rather than what they express, while text captures the script prepared in advance, the most easily managed layer of expression. Text cannot reveal whether the same written line is delivered calmly or with visible anger. We therefore measure delivery directly. Building on recent work that measures emotion in politicians' voices and faces (Dietrich et al., 2019; Boussalis et al., 2021), we apply deep learning models to 355 hours of plenary video from South Korea's 20th National Assembly, covering 11,060 speech segments from 247 legislators between 2016 and 2020. We score anger independently in two channels—voice and face—and combine them into a composite measure of delivered anger.

[1] Throughout the paper, scandal refers to floor discussion of corruption, abuse of power, or other misconduct by political actors. In our period this includes the Choi Soon-sil affair and the impeachment of President Park Geun-hye, the Cho Guk controversy, and numerous cases of bribery, abuse of power, and misuse of public funds involving legislators and officials from both major party blocs.

Consistent with our argument, we find that anger peaks where blame can be pinned on the opposing party. Discussing scandal raises a legislator's vocal anger by 0.35 standard deviations relative to routine parliamentary business, roughly two and a half times the effect of policy debate, where district interests are most directly at stake. On our composite measure of delivered anger, which averages the voice and the face, the scandal effect is 0.21 standard deviations, roughly twice the policy effect. The pattern holds within legislator and within session, and it is statistically indistinguishable between legislators with and without districts.

The impeachment quasi-experiment provides the sharpest test of our argument. When the impeachment reversed the roles of the two blocs, legislators entering government became measurably calmer and legislators entering opposition angrier, a within-person shift of 0.35 standard deviations in vocal anger and 0.24 on the composite. The shift arrives when the roles formally change rather than when the scandal breaks, and the shift is specific to adversarial emotions. Governing legislators sound less angry and less contemptuous but more joyful and calmer, which is what a change in institutional role predicts and what a general shock does not.

These findings contribute first to the study of partisan competition. This literature has documented elite conflict in roll calls, obstruction, and rhetoric (Lee, 2009; Grimmer and King, 2011; Theriault, 2013; Tuttnauer, 2018), but has rarely observed the emotions with which that conflict is waged. We show that anger behaves as this literature would expect a weapon to behave. It concentrates where the opposing party can be blamed, it is taken up by whichever bloc sits in opposition, and it is set down by whichever bloc governs, while remaining unmoved by the constituencies legislators represent. In this we join recent work treating nonverbal expression as an instrument of partisan conflict (Rask and Hjorth, 2025) and text-based evidence that opposition status darkens the tone of parliamentary speech (Rheault et al., 2016). The finding also carries an implication for the study of affective polarization. If elite anger drives citizen anger (Stapleton and Dawkins, 2022), and elite anger is governed by institutional position, then part of the emotional temperature of mass politics is set inside the legislature, by the structure of the contest rather than the feelings of the contestants.

Second, the paper makes a methodological contribution to the growing use of audio and video as data (Dietrich et al., 2019; Boussalis et al., 2021; Rask and Hjorth, 2025). We provide the first measurement of a discrete emotion in both the voice and the face of every speaker across an entire legislature, and combine the channels into a single composite measure of delivered anger. The design separates three layers of expression that existing work conflates, namely what legislators write, what they say, and how they deliver it. Because the framework requires only video archives and is standardized within each channel, it travels to any legislature with recorded proceedings.

## 2 The Targets of Legislative Anger: Constituents or Colleagues?

In the representation literature, the legislator acts on behalf of the constituents of the district. Constituency preferences discipline legislative behavior (Miller and Stokes, 1963; Pitkin, 1967), the electoral connection organizes activity around advertising, credit claiming, and position taking (Mayhew, 1974), and legislators cultivate a presentation of self that binds them to the people they represent (Fenno, 1978). On this view, communication is itself a core act of representation. What legislators say, and how they say it, constitutes the account they give to their principals (Grimmer, 2013). Recent work extends the argument to emotion. Legislators do not merely transmit constituent interests. They perform constituent feeling, outrage on behalf of the aggrieved and compassion on behalf of the suffering, so that emotional expression becomes a form of representation in its own right (Stout et al., 2025). Consistent with this view, emotive rhetoric rises in the high-profile debates that attract the largest audiences (Osnabrügge et al., 2021), and legislators grandstand when electoral incentives sharpen (Park, 2021a).

The partisan competition literature reads the same behavior differently. Legislative behavior is organized as much by competition between party teams as by ideology or district interest. Parties manufacture conflict on issues without ideological stakes because weakening the other side is itself the objective (Lee, 2009, 2016), partisan taunting constitutes a measurable and sizable share of elite communication (Grimmer and King, 2011), and inter-party combat has become a defining

feature of contemporary legislatures (Theriault, 2013). The demand side reinforces this supply of conflict. Partisan voters increasingly dislike and are angered by the opposing party (Iyengar et al., 2012; Mason, 2018; Abramowitz and Webster, 2016; Webster, 2020), supplying an audience that welcomes combativeness, and elites can perform hostility they do not feel (Arceneaux et al., 2024). Moreover, the adversarial role is institutionalized. Opposing the government is the opposition's constitutional job, and opposition parties allocate substantial effort to confrontation (Lijphart, 2012; Tuttnauer, 2018; Proksch and Slapin, 2012). This literature, however, has studied the contest in votes, obstruction, and words rather than in the emotions with which it is waged. We argue that anger belongs to the same arsenal, an emotional weapon of the contest for power.

Among emotions, anger occupies a privileged position in politics. At the mass level, it is the mobilizing emotion, driving participation and turnout in ways that anxiety and enthusiasm do not (Valentino et al., 2011; Brader, 2006). This mass-level anger is measured by asking. National surveys carry emotion batteries in which respondents report whether a candidate or party has made them feel angry, and experimental work induces anger directly (Valentino et al., 2011; Webster, 2020). Appraisal theory explains what makes anger distinct. Anger is the blame emotion. Unlike fear, sadness, or anxiety, it arises from the appraisal that an identifiable agent is responsible for a wrong, and it motivates confrontation and punishment of that agent (Lazarus, 1991; Smith and Ellsworth, 1985). An angry floor speech, therefore, accuses, and in a two-bloc legislature, the accused is, with rare exceptions, the opposing party. Anger is not one emotional flavor among many. It is the emotion whose expression constitutes an attack, which makes it the natural terrain for adjudicating whom legislative expression serves.

Far less attention has gone to the production of legislative anger, what makes legislators angry in the first place and at whom the anger is aimed, than to its consequences for voter evaluations, citizen emotion, and electoral outcomes (Boussalis et al., 2021; Stapleton and Dawkins, 2022). Part of the reason is the difficulty of measurement. Legislators do not answer emotion batteries, so their anger must be read off what they say and do. The largest literature scores the text of speeches (Grimmer and Stewart, 2013; Rheault et al., 2016; Gennaro and Ash, 2022; Osnabrügge

et al., 2021; Yildirim et al., 2025), and a newer body of work measures one nonverbal channel at a time, tracking shifts in vocal pitch from floor audio (Dietrich et al., 2019; Rask and Hjorth, 2025) or classifying facial expressions frame by frame in televised debates with computer vision (Boussalis et al., 2021). Advances in deep learning now allow more, and we use them to recover anger directly from both nonverbal channels for every speaker in a legislature.

The deeper reason is that the two accounts, anger as service to constituents and anger as a weapon in the contest for power, are ordinarily difficult to separate in expressive behavior, because a fiery speech serves both purposes at once. Anger delivered on the floor can be read as passionate advocacy for the district or as an attack on the opposing party, and most observable implications of the two readings coincide. Both predict more emotion on contested topics, more emotion from electorally exposed legislators, and more emotion when cameras are on. The field has therefore been unable to say whom legislative anger is for. Separating the accounts requires what the observational record rarely supplies, namely variation in a legislator's competitive position that leaves every constituency untouched, legislators with no constituency at all to serve as the comparison, and a measure of the emotion legislators actually deliver rather than the script they prepare.

# 3 Anger in the South Korean National Assembly

South Korea's 20th National Assembly (2016–2020) supplies the variation this test requires. It is also a setting in which existing scholarship divides along the same line examined in this paper. One line of work emphasizes the district basis of Korean legislative politics. Regional loyalties structure electoral competition, and district legislators cultivate intensive constituency service (Park, 1988; Horiuchi and Lee, 2008). Another emphasizes the partisan contest. Rivalry between the conservative and progressive blocs intensified over precisely this period, with polarization documented among both elites and the mass public (Cheong and Haggard, 2023; Park, 2021b). The Assembly of these years is therefore not merely a convenient laboratory. It is a case in which both accounts of legislative behavior have deep empirical roots.

The period also provides substantial variation in scandal exposure. The Choi Soon-sil affair, the impeachment of President Park Geun-hye, the Cho Guk controversy, and repeated corruption allegations involving politicians from both major blocs kept accusation on the floor throughout the legislative term. Scandal-heavy periods are not unique to Korea. The United States Congress experienced two presidential impeachment proceedings within a single term, and Brazil's Lava Jato investigations shaped congressional politics for years. Few legislatures, however, experienced such concentrated variation in blame and accusation within a four-year period.

South Korea's mixed-member electoral system provides additional leverage by placing two types of legislators side by side within the same chamber. Of the National Assembly's 300 members, 253 are elected from single-member districts and represent geographically defined constituencies, while 47 enter through party lists and have no individual district. This arrangement allows us to compare legislators with and without constituency ties while holding the broader partisan environment constant.

The impeachment of President Park Geun-hye provides an unusual opportunity to observe the same legislators on both sides of the partisan contest. The scandal broke in October 2016, the Assembly voted to impeach that December, the Constitutional Court removed the president in March 2017, and the inauguration of President Moon Jae-in that May moved the conservative bloc from government to opposition and the progressive bloc from opposition to government. Governing status reversed while every constituency stayed in place. Because no legislative election occurred between the two periods, the same legislators faced the same districts before and after the reversal, so any change in their anger cannot be traced to a change in whom they represent.

Taken together, this setting creates a direct test between the representation and partisan-contest accounts of legislative anger. If anger reflects passionate representation, it should be concentrated in policy debates where district interests are most directly implicated, stronger among district legislators, and unaffected by the impeachment-induced change in governing status because constituencies remained constant. If anger serves the partisan contest, the pattern should differ. Anger should be highest during scandal debates, where blame attaches to opposing political actors. It should not

depend on whether legislators hold districts. And it should follow partisan position, rising when legislators enter opposition and falling when they enter government. Testing these expectations requires a measure of anger that captures what legislators actually deliver rather than only the scripts they prepare.

# 4 Data and Methods

## 4.1 Data

Our analysis focuses on nationally televised plenary sessions (*bonhoeui*), where legislators deliver speeches before the full chamber.[2] Two features of plenary procedure are central to our design. First, the agenda is set by the speaker in consultation with the House Steering Committee rather than by individual legislators, so the topic a legislator addresses is largely predetermined rather than chosen in response to how they feel at the moment of speaking. Second, floor speech is scripted: interpellation questions must be filed forty-eight hours in advance, and even five-minute free speeches require a written summary submitted before the session opens.[3] What a legislator says is thus substantially fixed in writing beforehand, while the delivery, the voice and the face, is produced live in the chamber.

We collected video recordings of all plenary sessions archived by the National Assembly's official broadcasting archive. The corpus comprises 156 plenary meetings totaling approximately 21,300 minutes (355 hours) of recorded proceedings. We segmented these recordings into individual speech segments, with each segment corresponding to one legislator's continuous speaking turn within a session. The full dataset contains 11,462 segments from 250 unique legislators. After ex-

[2]We exclude parliamentary audit sessions (*gukjeonggamsa*), confirmation hearings (*insacheongmunhoe*), standing committee meetings, and special budget committee sessions, which differ substantially from plenary sessions in their level of confrontation, audience composition, and media exposure. Restricting to plenary sessions ensures that variation in anger reflects topic and institutional role rather than the setting itself.

[3]National Assembly Act, Arts. 76 and 77 (agenda setting; modification requires either a motion signed by at least 20 members followed by a plenary vote or the agreement of the Speaker and floor leaders), Art. 122-2 (interpellation of the government, *daejeongjilmun*, which is also organized into thematically designated days), and Art. 105 (five-minute free speech, *5-bun jayu baleon*).

cluding 402 segments without a valid topic label (see below), the analytic sample comprises 11,060 segments from 247 legislators.[4] Each segment carries three parallel streams—video, audio, and transcript—which we analyze separately below. To link facial expressions to individual legislators, we obtained official portrait photographs of all members serving in the 20th National Assembly and used face recognition to identify the speaker associated with each segment. The resulting linkage enables us to attach individual-level characteristics, including party affiliation, ideological bloc, gender, seniority, and reelection status, to each segment's emotion measurements.

## 4.2 Topic Classification

We classified each segment into one of five mutually exclusive topic categories using a large language model (Claude Haiku, Anthropic): SCANDAL (discussion of corruption, abuse of power, or other moral transgressions by political actors; $n = 1{,}727$), EVENT (discussion of major national events and crises, such as the Sewol ferry disaster; $n = 2{,}114$), POLICY (substantive debate over legislation, budgets, or governance; $n = 4{,}051$), REFORM (institutional or electoral reform; $n = 364$), and PROCEDURAL (routine parliamentary business; $n = 2{,}804$). An additional 402 segments (3.5%) received labels that did not correspond to any of the five predefined categories and are excluded from all analyses.[5] The classification prompt and category definitions are described in Appendix E.

To assess classification reliability, two Korean-speaking research assistants independently coded a stratified random sample of 200 segments; agreement was substantial (pairwise Cohen's $\kappa$ = 0.76–0.78 across all coder pairs; Krippendorff's $\alpha$ = 0.80; see Appendix J for full inter-coder results including confusion matrices). Note that the classifier identifies the subject under discussion rather than the speaker's position on that issue. Thus, a governing-party defense of the administration and an opposition attack during the same scandal debate receive the same topic label.

[4]The sample includes 202 male and 48 female legislators, drawn from conservative (92), progressive (101), and minor party or independent (57) blocs. Appendix A reports full descriptives.

[5]Procedural segments serve as the omitted baseline in all regressions. The main results are substantively unchanged when policy or reform is used as the reference category.

## 4.3 Measuring Delivered Anger

Our measurement strategy follows directly from the institutional features described above. The text of a floor speech records the script, fixed in advance, while the voice and the face are produced live in the chamber. We therefore measure anger in two delivery channels—vocal prosody and facial expression—and combine them into a composite measure of delivered anger. To maintain a direct comparison with text-based approaches, we also score every transcript using KOTE, a transformer-based Korean text-emotion model (Jeon et al., 2022).[7]

We measure delivered anger independently across two channels using Hume AI's expression measurement platform, which produces continuous scores (0–1) for 48 emotion dimensions grounded in the semantic space theory of emotion (Cowen and Keltner, 2021; Keltner et al., 2023).[8] The facial model processes video frame by frame, whereas the prosody model analyzes pitch, rhythm, timbre, and intensity in the audio signal. Importantly, the two models rely on separate input streams—silent video for facial expression and audio alone for vocal expression—so the two channels are estimated independently rather than derived from a single multimodal score.

Each channel yields a segment-level mean anger score.[9] We $z$-standardize each score within

[6] This distinction is intentional. The topic variable captures the presence of a potential blame target in the discussion rather than the rhetorical posture of an individual speaker.

[7] We aggregate KOTE's anger-related labels into a segment-level text-anger score. Because KOTE is a Korean language model applied to the same segments and designed to measure the same construct, differences between text-based and delivery-based measures are less likely to reflect differences in model applicability alone and may instead capture differences between expressive layers. Three segments whose ASR transcripts contain only a two- or three-character fragment do not contain sufficient text for scoring, leaving $N = 11{,}057$ observations for text-based measures; these segments are retained in the voice and face channels, which directly analyze audio and video streams.

[8] Semantic space theory holds that emotional expression is better described as a continuous, high-dimensional space of blended states than as a small set of discrete basic emotions. Large-scale studies of vocal and facial expression recover dozens of reliably distinguished dimensions (Cowen et al., 2020; Cowen and Keltner, 2020, 2021). The practical implication for our design is that the models score anger as one dimension among many rather than assigning each segment to a single category. We use Hume as the primary measure because its models are trained on large-scale naturalistic expression data, which better matches the subtle and regulated emotional displays common in parliamentary settings; see Brooks et al. (2023) for a technical description and validation of the API-based models. Appendix C shows that our results are not sensitive to this measurement choice.

[9] The mean is taken across the segment's full duration. Peak scores, capturing brief spikes that may reflect momentary departures from emotional regulation, are analyzed in Appendix G (Table 12) and Appendix O.

its channel and define composite delivered anger as the average of the two standardized scores.[10] Averaging is motivated by the fact that the two channels capture complementary aspects of the same underlying construct, despite only modest correlation ($r = 0.09$). The composite records whether delivered anger rose overall, while the channel-specific measures reveal the mode through which it rose.

## 4.4 Diagnostics and Validation

Because voice, face, and text are scored on the same segments, the design permits three direct diagnostics. The first asks whether the text, the layer on which most existing measurement of legislative emotion relies, can stand in for the delivery. It cannot. Text anger correlates with voice anger at $r = 0.34$—the emotional tone of the script explains about 12% of the variance in the anger of the voice that reads it—and with face anger at $r = -0.05$. Figure 1 shows the disagreement segment by segment. Taking the top quartile of each measure as a flag for anger, 57% of the segments the text measure flags are not delivered with an angry voice, and 77% are not delivered angrily on the composite. Conversely, many segments delivered with high levels of anger are associated with transcripts that receive low text-anger scores.[11]

[10] Standardization removes scale differences between the two models, whose raw distributions may reflect calibration as well as behavior. Because the two channels are nearly uncorrelated, the composite has a standard deviation of approximately 0.73 rather than 1; we report composite coefficients in channel-standard-deviation units throughout, and re-standardizing the composite itself rescales them by a factor of roughly 1.36 without changing any test statistic. As a nonparametric alternative, we also compute within-channel percentile ranks, placing all three measures—including text-based anger—on a common 0–1 scale; results are substantively identical

under either approach (Appendix B).

[11] 75% of text-flagged segments are not delivered with an angry face. The disagreement grows with the stringency of the flag: at the 90th percentile, 76% of text-flagged segments lack an angry voice and 89% lack an angry face.

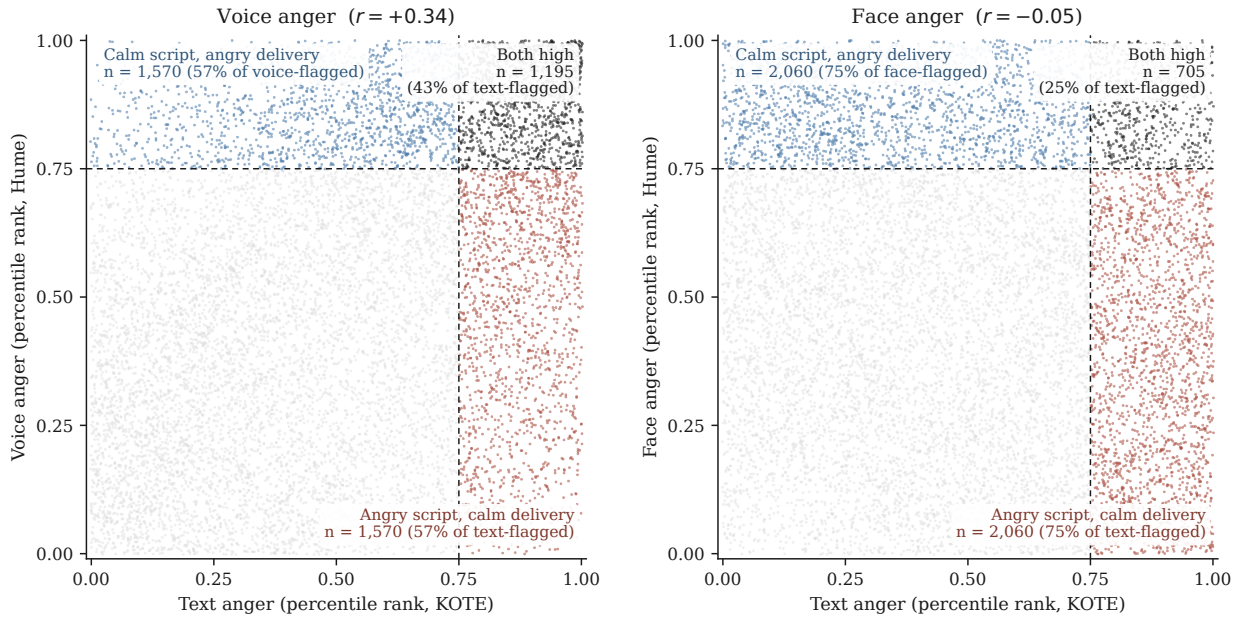


Figure 1: Text anger versus delivered anger. Each point is one speech segment ($N = 11{,}057$), plotted by its percentile rank on the KOTE text-anger measure (horizontal axis) and the Hume delivery measure (vertical axis; voice left, face right). Dashed lines mark the 75th percentile. Segments in the lower-right quadrant are flagged as angry by the text but not delivered angrily; segments in the upper-left are delivered angrily on scripts the text reads as calm.

The gap between script and delivery also varies systematically with the institutional variables examined in this paper. In linear probability models with legislator and meeting fixed effects, governing status reduces the probability that an angry script is delivered angrily by 12 to 15 percentage points across the three delivery measures (all $p < 0.01$; Appendix B). Governing-party legislators continue to write angry scripts; they are simply less likely to perform them. A text-only design therefore carries measurement error that is correlated with governing status, the treatment variable in Section 6, and overstates how strongly expressed anger responds to institutional incentives.[12]

Appendix B presents further evidence that the delivery measures accurately capture anger rather than generic arousal or model artifacts. Within each channel, anger correlates strongly with other adversarial emotions (contempt, disgust, and distress) and negatively with calmness and joy, while remaining nearly orthogonal to determination, indicating that the measure distinguishes

[12]Under identical specifications, scandal raises text anger by 23 percentile points but composite delivered anger by only 8 (Appendix Table 7); the full text-based analysis appears in Appendix D.

angry speech from merely forceful speech. Segments independently coded as accusations also exhibit the highest levels of voice anger among all rhetorical modes. Appendix C remeasures anger on the same segments using alternative open-source models and reproduces the paper's main results, whereas composites based on low-level acoustic and facial-muscle features do not. Finally, because all models include legislator and meeting fixed effects and outcomes are standardized within channel, the estimates depend only on the rank ordering of anger within the sample rather than on absolute calibration in the Korean context.

## 4.5 Empirical Specification

Our main estimation framework is a two-way fixed effects model:

$$Y_{it} = \alpha_i + \gamma_{m(t)} + \beta \cdot X_{it} + \varepsilon_{it} \tag{1}$$

where $t$ indexes speech segments, $i$ indexes legislators, and $m(t)$ denotes the meeting to which segment $t$ belongs. The legislator fixed effects $\alpha_i$ absorb all time-invariant individual traits, including party affiliation, personality, and baseline expressiveness; the meeting fixed effects $\gamma_{m(t)}$ absorb all session-level confounds, including the political climate, media attention, and agenda composition. The coefficient $\beta$ is therefore identified from within-legislator, within-meeting variation: the same legislator shifting between contexts, net of individual and session means. The variable of interest $X_{it}$ differs by analysis: topic indicators in Section 5 and the governing-status indicator GOV_TV in Section 6. Standard errors are clustered at the legislator level (Liang and Zeger, 1986) with a finite-sample correction (Angrist and Pischke, 2009).

We report all three outcomes throughout—voice anger, face anger, and the composite—with coefficients expressed in standard deviations of each channel's anger distribution. Appendix B shows that every result is substantively identical under percentile-rank standardization.

# 5 Results

We first examine the topic of legislative speech, where the two accounts make their clearest distinction. The representation account points to policy debate, while the partisan-contest account predicts greater anger in scandal debates. Table 1 reports the effect of each topic relative to procedural business within the same meeting and legislator; Table 2 compares topic effects directly.

Table 1: Topic Effects on Anger Expression

| | (1) Voice anger ($z$) | (2) Face anger ($z$) | (3) Composite ($z$) |
|---|---|---|---|
| SCANDAL | $+0.351^{***}$ | $+0.067^{*}$ | $+0.209^{***}$ |
| | (0.055) | (0.036) | (0.038) |
| EVENT | $+0.222^{***}$ | $+0.043$ | $+0.132^{***}$ |
| | (0.057) | (0.037) | (0.040) |
| POLICY | $+0.139^{***}$ | $+0.096^{***}$ | $+0.118^{***}$ |
| | (0.051) | (0.033) | (0.034) |
| REFORM | $+0.175^{**}$ | $+0.086^{*}$ | $+0.131^{***}$ |
| | (0.069) | (0.051) | (0.048) |
| Legislator FE | Yes | Yes | Yes |
| Meeting FE | Yes | Yes | Yes |
| $N$ | 11,060 | 11,060 | 11,060 |

*Notes:* Each column is a separate regression. Outcomes are raw Hume segment-mean anger scores $z$-standardized within channel on the full sample (columns 1–2) and the average of the two channel $z$-scores (column 3). Procedural segments are the omitted category; stars test each topic against the procedural baseline. Standard errors clustered at legislator level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The composite measure in column 3 shows clear differences across topics. All four substantive topics are associated with more delivered anger than procedural business, but the magnitude varies considerably. Scandal produces the largest increase, at 0.21 standard deviations, compared with 0.12 for policy, with event and reform falling between the two. The ordering is therefore not what

we would expect if anger primarily reflected constituency representation. Policy debates, where district interests are most directly implicated, do not generate the highest levels of anger. Scandal does.

Table 2: Pairwise Topic Contrasts: Scandal versus Each Topic

| | (1) Voice anger ($z$) | (2) Face anger ($z$) | (3) Composite ($z$) |
|---|---|---|---|
| SCANDAL − EVENT | $+0.129^{***}$ | $+0.024$ | $+0.076^{***}$ |
| | (0.038) | (0.028) | (0.026) |
| SCANDAL − POLICY | $+0.212^{***}$ | $-0.030$ | $+0.091^{***}$ |
| | (0.035) | (0.028) | (0.026) |
| SCANDAL − REFORM | $+0.175^{***}$ | $-0.020$ | $+0.078^{*}$ |
| | (0.057) | (0.049) | (0.040) |

*Notes:* Wald tests of the difference between the scandal coefficient and each other topic coefficient from the regressions in Table 1. The stars in Table 1 test each topic against the procedural baseline; the contrasts here test topics against one another, which is the comparison underlying the gradient claim. Standard errors of the differences, clustered at legislator level, in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The pairwise comparisons in Table 2 confirm that the topic gradient is concentrated in the voice measure. Vocal anger is significantly higher during scandal than during event, policy, or reform debates. The scandal-policy difference is 0.21 standard deviations. The corresponding face differences are small and statistically indistinguishable from zero. The composite follows the voice measure, with scandal significantly above event and policy and marginally above reform. Thus, the overall topic gradient is not driven by a broad increase in facial expressiveness across substantive topics; it is primarily a difference in vocal anger. This channel difference also helps

distinguish the result from a general increase in arousal during contentious debate. If all forms of expressive intensity rose on more contentious topics, we would expect a similar pattern across the two delivery channels. Instead, the topic ordering is much sharper in the voice measure. The timing evidence in Appendix O is consistent with this distinction: facial anger displays are briefer than vocal ones.

**Constituency-Based Heterogeneity.** One competing interpretation is that the topic pattern reflects constituency-oriented expression rather than partisan conflict. This account implies that the contrast between scandal and policy should be stronger among district legislators, who represent individual geographic constituencies, than among party-list legislators. We assess this implication by interacting the topic indicators with party-list status.

Table 3: Topic Effects for District and Proportional-Representation Members

| | (1) Voice anger ($z$) | (2) Face anger ($z$) | (3) Composite ($z$) |
|---|---|---|---|
| SCANDAL × PR | −0.050 | −0.028 | −0.039 |
| | (0.124) | (0.075) | (0.076) |
| EVENT × PR | +0.085 | +0.006 | +0.046 |
| | (0.153) | (0.069) | (0.088) |
| POLICY × PR | −0.007 | −0.063 | −0.035 |
| | (0.116) | (0.062) | (0.064) |
| REFORM × PR | +0.083 | −0.091 | −0.004 |
| | (0.174) | (0.143) | (0.121) |
| Topic main effects | Yes | Yes | Yes |
| Legislator FE | Yes | Yes | Yes |
| Meeting FE | Yes | Yes | Yes |
| $N$ | 11,060 | 11,060 | 11,060 |

*Notes:* Each column adds topic × PR interactions to the specification in Table 1. Each coefficient is the difference in that topic's effect for proportional-representation members, who have no district, relative to district members. The PR main effect is absorbed by the legislator fixed effects. Full estimates, including the topic effects for district members, appear in Appendix I. Standard errors clustered at legislator level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The interactions are small and statistically insignificant across all three outcomes and all four topics, with all $p > 0.31$. The estimated topic effects are therefore similar for legislators with and without districts (Appendix I). The absence of a systematic difference by constituency status is

difficult to reconcile with an account in which legislative anger primarily reflects district representation.

**Robustness.** Several additional tests assess the robustness of these results. Results are unchanged when anger is expressed as percentile ranks rather than standardized scores (Appendix B). The voice gradient survives controls for speech duration (Appendix G) and remains when the analysis is conducted within legislator-meeting cells using interacted fixed effects (Appendix N). The scandal effect is also similar for conservative and democratic legislators (Appendix H). The same contrast between adversarial and positive emotions appears when the analysis is extended to the full emotion space (Appendix F). Finally, the topic pattern is reproduced using independent open-source emotion models for both channels, whereas composites based on low-level acoustic and facial-muscle features do not reproduce it (Appendix C).

# 6 The Impeachment Quasi-Experiment

The topic results show that anger is higher when legislators discuss issues that provide identifiable targets for partisan attack. We next examine whether anger also changes with a legislator's position in the partisan contest. The prediction is straightforward. If anger is part of partisan competition, the same legislator should express more anger in opposition and less in government. If anger instead reflects constituency representation, changing governing status should have little effect because the legislator's constituency remains the same. The impeachment of President Park Geun-hye and the subsequent inauguration of President Moon Jae-in provide this change in governing status without an intervening legislative election. The two events moved the conservative bloc from government to opposition and the progressive bloc from opposition to government. The same legislators therefore remained in office and continued to represent the same constituencies while their partisan roles changed. We estimate the following specification using all segments from the

two major-party blocs ($N = 8{,}293$; 190 legislators; 135 meetings):

$$Y_{it} = \alpha_i + \gamma_{m(t)} + \delta \cdot \text{GOV\_TV}_{i,m(t)} + \varepsilon_{it} \tag{2}$$

where $\text{GOV\_TV}_{i,m(t)}$ equals one when legislator $i$'s party holds the presidency at the time of meeting $m$. The indicator flips at Moon Jae-in's inauguration on May 10, 2017.[13] Because legislator fixed effects absorb party affiliation, $\delta$ is identified from within-legislator variation: the same legislators observed under both institutional roles.

Equation 2 compares changes in anger within legislators as their parties move between government and opposition, while meeting fixed effects absorb shocks common to all legislators in a given meeting. The specification therefore removes stable differences across legislators and common changes in the legislative environment. The reversal occurs for both major blocs, so the coefficient combines the transition from opposition to government among progressive legislators with the transition in the opposite direction among conservatives.[14]

[13] During the caretaker period following Park's formal removal (March 10, 2017), conservatives remain coded as governing because acting president Hwang Kyo-ahn was a conservative appointee. In practice no plenary meetings occurred between March 30 and May 29, 2017, so no meeting falls on the transition date. The descriptive crossover below excludes the October 2016–May 2017 window—the scandal's outbreak, the impeachment vote, and the caretaker presidency—to give a clean comparison of means; the regressions in Table 4 use the full sample, with legislator and meeting fixed effects absorbing period-specific shocks. Results are substantively identical either way.

[14] Emotional reactions to the impeachment itself could affect anger independently of governing status. The event study below examines whether the change occurs at the formal role reversal rather than at the scandal's outbreak or the impeachment vote, and whether it persists after the immediate events have passed.

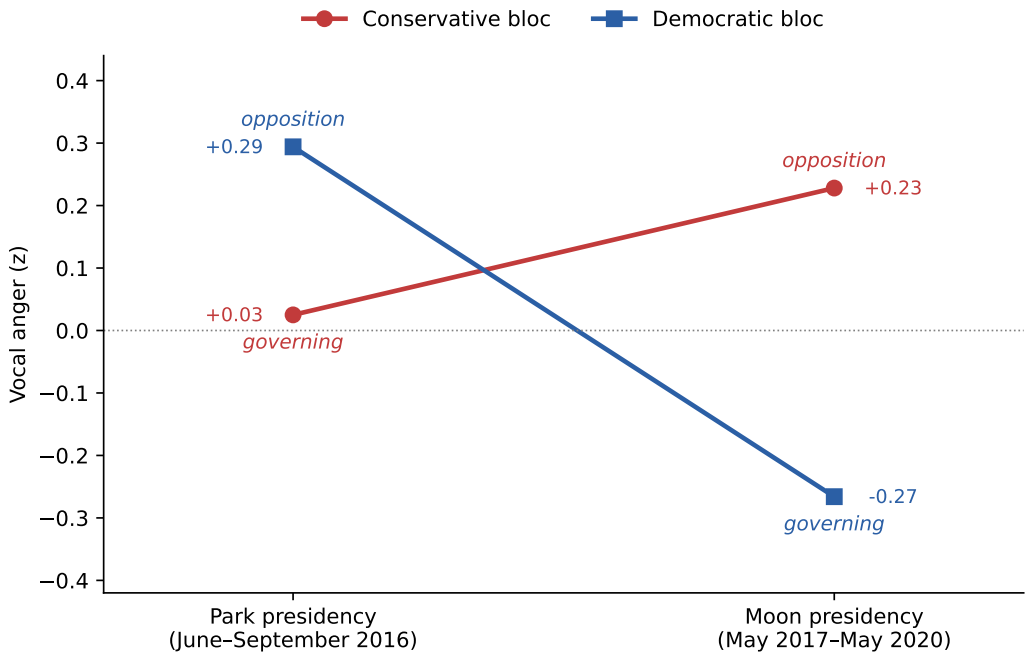


Figure 2: Mean vocal anger by bloc under the Park and Moon presidencies. Means exclude the October 2016–May 2017 transition window. Whichever bloc sits in opposition expresses more anger, and the two blocs converge to nearly the same level when occupying the same seat.

Figure 2 shows the raw means underlying the regression. Democratic legislators' voice anger falls by 0.56 standard deviations when they move from opposition to government, from $+0.29$ to $-0.27$ relative to the sample mean. Conservative legislators' voice anger rises by 0.20 standard deviations when they move from government to opposition, from $+0.03$ to $+0.23$. The two blocs have similar levels of voice anger when they are both in opposition, at $+0.23$ and $+0.29$. The regression results are reported in Table 4.

Table 4: Governing Party Status and Anger Expression

| | (1) Voice anger ($z$) | (2) Face anger ($z$) | (3) Composite ($z$) |
|---|---|---|---|
| GOV_TV | $-0.345^{***}$ | $-0.125^{**}$ | $-0.235^{***}$ |
| | (0.052) | (0.058) | (0.045) |
| Legislator FE | Yes | Yes | Yes |
| Meeting FE | Yes | Yes | Yes |
| $N$ | 8,293 | 8,293 | 8,293 |

*Notes:* Sample restricted to legislators in the two major-party blocs (190 legislators; 135 meetings). GOV_TV $= 1$ if the legislator's party holds the presidency. Outcomes as in Table 1. Standard errors clustered at legislator level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 4 shows lower anger among governing-party legislators on all three measures. The coefficient is $-0.35$ standard deviations for voice, $-0.13$ for face, and $-0.24$ for the composite. The difference is larger for voice than for face, as in the topic analysis. The result also does not vary systematically with constituency status. Interactions between GOV_TV and the proportional-representation indicator are small and statistically insignificant for all three outcomes (composite $-0.04$, $p = 0.64$; Appendix I). Legislators without districts therefore show a similar response to the change in governing status as district legislators.

**Event study.** We then examine the timing of the change and whether it can be attributed to the role reversal rather than to the events surrounding the impeachment. We interact period indicators

with a conservative-bloc indicator:

$$Y_{it} = \alpha_i + \gamma_{m(t)} + \sum_p \delta_p (\text{Period}_p \times \text{Conservative}_i) + \varepsilon_{it} \tag{3}$$

where $\text{Period}_p$ are indicators for each time period (transition, H2 2017, 2018, H1 2019, H2 2019–2020), the pre-scandal period (June–September 2016) is the reference category, and the coefficients $\delta_p$ trace how the conservative–democratic gap in voice anger evolved relative to the pre-scandal baseline, net of legislator and meeting fixed effects. Figure 3 plots the coefficients; Appendix K reports the estimates in table form.

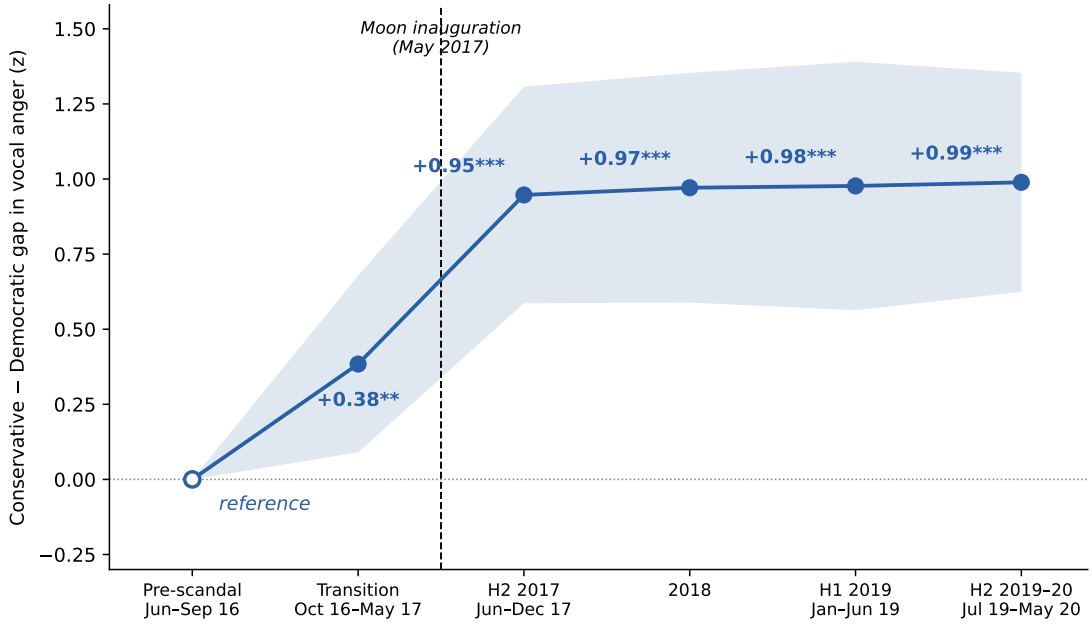


Figure 3: Event Study: Conservative–Democratic Gap in Voice Anger Around the Impeachment. Each point shows the Conservative × Period interaction coefficient on voice anger ($z$-score scale), with 95% confidence intervals; point estimates are printed with significance stars ($^{***}p < 0.01$, $^{**}p < 0.05$). The reference period is pre-scandal (June–September 2016), when democrats sat in opposition and were the angrier bloc. The dashed line marks Moon Jae-in's inauguration (May 2017). Sample: two major-party blocs ($N = 8{,}293$; 190 legislators; 135 meetings). Standard errors clustered at legislator level. Full estimates appear in Appendix K.

Figure 3 provides the timing evidence. In the pre-scandal period, when democrats were in opposition, the conservative–democratic difference was $-0.27$ SD. The difference moves toward zero during the transition period and then changes sharply after the formal change in governing

status. It reaches +0.95 SD in the first full session after Moon's inauguration and remains close to that level thereafter, at +0.97 in 2018, +0.98 in the first half of 2019, and +0.99 thereafter (all $p < 0.01$). The transition period shows a smaller difference of +0.38 SD, when scandal accounted for 27% of all speech segments, more than twice its share in any subsequent period.[15] The larger and persistent difference after May 2017 is therefore difficult to attribute to the initial scandal shock alone.

**Emotion specificity.** The role reversal should affect the adversarial emotions most directly associated with partisan conflict rather than simply changing overall expressiveness. We estimate the governing-status specification separately for each vocal emotion, using other adversarial emotions and positive emotions as comparison outcomes. Table 5 reports the estimates.

[15] During the transition period (October 2016–May 2017), scandal accounted for 27% of all speech segments, more than double its share in any other period.

Table 5: Governing-Status Effects Across Vocal Emotions

| Vocal emotion ($z$) | GOV_TV | |
|---|---|---|
| *Adversarial* | | |
| Anger | $-0.345^{***}$ | (0.052) |
| Contempt | $-0.210^{***}$ | (0.038) |
| Disgust | $-0.185^{***}$ | (0.042) |
| Distress | $-0.160^{***}$ | (0.039) |
| *Positive* | | |
| Joy | $+0.130^{***}$ | (0.042) |
| Contentment | $+0.144^{***}$ | (0.037) |
| Calmness | $+0.227^{***}$ | (0.033) |

*Notes:* Each row is a separate regression of the vocal $z$-score for that emotion on GOV_TV with legislator and meeting fixed effects, on the two-party sample ($N = 8{,}293$; 190 legislators; 135 meetings). Estimates for all 48 vocal emotion dimensions appear in Appendix M. Standard errors clustered at legislator level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The estimates extend beyond anger to the broader emotional register. Governing status is associated with lower levels of all four adversarial emotions in the table, including anger ($-0.35$ SD), contempt ($-0.21$), disgust ($-0.19$), and distress ($-0.16$). At the same time, joy ($+0.13$), contentment ($+0.14$), and calmness ($+0.23$) increase. All seven estimates are statistically significant at $p < 0.01$. Across the full set of 48 vocal dimensions, anger has the largest negative coefficient and calmness the largest positive coefficient; determination does not change (Appendix M). The results

therefore do not indicate a general decline in vocal expressiveness among governing-party legislators. Rather, the changes are concentrated in the emotional dimensions most closely associated with adversarial and positive expression.

We also conducted several additional tests to assess the robustness of these results. The voice and face estimates are unchanged or larger after controlling for topic and speech duration, with the voice estimate increasing when duration is held fixed (Appendix G). The results also hold under percentile-rank standardization (Appendix B) and under alternative inference procedures for the single-shock design, including two-way clustering, the wild cluster bootstrap, and randomization inference based on party-bloc permutations (Appendix L). Face anger is less stable under the most demanding corrections and should therefore be interpreted more cautiously.

# 7 Conclusion

Scholars of legislative behavior have read anger as a performance for voters (Osnabrügge et al., 2021; Park, 2021a; Stout et al., 2025). Because an angry outburst risks humiliation and damage to reputation, its display is treated as a costly signal of passion on the district's behalf. This article argues that anger is aimed at fellow legislators instead and moves with the contest between parties. Measuring anger directly in the voice and face of every speaker across four years of South Korea's National Assembly, we find that legislators are angriest when discussing the misconduct of their colleagues, roughly twice as angry as when debating the policy questions that matter most to their districts. Misconduct hands the accusing side a target that cannot easily strike back, so these are the moments when attack is cheapest. The opposition seat works the same way. That opposing the government is the opposition's institutionalized role is well established (Lijphart, 2012; Tuttnauer, 2018). We show that anger serves this institutionalized role. The impeachment quasi-experiment demonstrates it directly. When the roles of the two parties reversed without an election, the same legislators became angrier upon entering opposition and calmer upon entering government. Anger is an instrument of the power contest, deployed where attack is cheap.

These findings carry implications for representation and polarization in two-party democracies. A growing literature treats elite emotional expression as representation, with legislators voicing what their constituents feel (Stout et al., 2025; Osnabrügge et al., 2021). We do not claim that emotion never serves representation. Our evidence shows that the contest between parties is the dominant driver of anger in the chamber. Members with no constituents at all display the same anger as members facing district voters, so party competition by itself is sufficient to produce what looks like passionate representation.

The results also draw a bleaker picture of what a polarized legislature spends its passion on. The emotional peak of the chamber we study is not the debate over what government should do but the assignment of blame for what the other side has done. This is the emotional counterpart of what Lee (2009) documents in the United States Senate, where parties manufacture conflict even on issues without ideological stakes because weakening the other side is itself the goal. Our evidence shows that this war over power rather than policy is waged not only in votes and words but in the emotional register of the legislature.

More broadly, the emotional register of politics matters more than ever because citizens increasingly encounter politics as video clips on social media rather than as written text. Emotional delivery is precisely what these clips carry and what transcripts strip away. Emotion is central to how political messages capture attention and move behavior (Brader, 2006; Valentino et al., 2011), and anger in particular mobilizes citizens and spreads from elites to the public (Stapleton and Dawkins, 2022; Webster, 2020). Our results imply that the anger citizens see on their screens reflects the competitive position of the speaker's party, not the mood of any district.

This article also makes a measurement contribution to the study of political emotion. Emotion in politics has mostly been measured indirectly, through surveys or through the text of what speakers wrote. We measured delivery itself, in the voice and the face of every speaker in a legislature, and the two layers disagree in ways that matter. Governing legislators keep writing angry scripts while delivering them calmly, so a text-only design would miss the suppression this article documents. The framework is portable to any legislature with video archives.

Democratic theory has long worried about passion overwhelming deliberation, and political science has shown how much of mass politics runs on emotion (Iyengar et al., 2012; Mason, 2018). We show that elite anger is neither an eruption nor a reflection of the constituencies legislators serve. It is driven by the contest for power between parties, rising where misconduct licenses attack and falling with the responsibilities of government. Rather than reading the anger of politicians as a display on the voters' behalf, future research should treat it as a measure of how fiercely power is being contested.

# References


Abramowitz, A. I. and Webster, S. (2016). The rise of negative partisanship and the nationalization of U.S. elections in the 21st century. *Electoral Studies*, 41:12–22.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ.

Arceneaux, K., Bakker, B. N., and Schumacher, G. (2024). Being of one mind: Does alignment in physiological responses and subjective experiences shape political ideology? *Political Psychology*, 45(4):741–759.

Boussalis, C., Coan, T. G., Holman, M. R., and Müller, S. (2021). Gender, candidate emotional expression, and voter reactions during televised debates. *American Political Science Review*, 115(4):1242–1257.

Brader, T. (2006). *Campaigning for Hearts and Minds: How Emotional Appeals in Political Ads Work*. University of Chicago Press, Chicago, IL.

Brooks, J. A., Tran, N., and Cowen, A. S. (2023). Deep learning and the measurement of expressive behavior. Hume AI technical report. See https://www.hume.ai/research.

Cheong, Y.-J. and Haggard, S. (2023). Political polarization in Korea. *Democratization*, 30(7):1215–1239.

Cowen, A. S., Fang, X., Sauter, D., and Keltner, D. (2020). What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures. *Proceedings of the National Academy of Sciences*, 117(4):1924–1934.

Cowen, A. S. and Keltner, D. (2020). What the face displays: Mapping 28 emotions conveyed by naturalistic expression. *American Psychologist*, 75(3):349–364.

Cowen, A. S. and Keltner, D. (2021). Semantic space theory: A computational approach to emotion. *Trends in Cognitive Sciences*, 25(2):124–136.

Cowen, A. S., Keltner, D., Schroff, F., Jou, B., Adam, H., and Prasad, G. (2021). Sixteen facial expressions occur in similar contexts worldwide. *Nature*, 589:251–257.

Dietrich, B. J., Hayes, M., and O'Brien, D. Z. (2019). Pitch perfect: Vocal pitch and the emotional intensity of congressional speech. *American Political Science Review*, 113(4):941–962.

Fenno, R. F. (1978). *Home Style: House Members in Their Districts*. Little, Brown, Boston, MA.

Gennaro, G. and Ash, E. (2022). Emotion and reason in political language. *The Economic Journal*, 132(643):1037–1059.

Grimmer, J. (2013). *Representational Style in Congress: What Legislators Say and Why It Matters*. Cambridge University Press, New York.

Grimmer, J. and King, G. (2011). General purpose computer-assisted clustering and conceptualization. *Proceedings of the National Academy of Sciences*, 108(7):2643–2650.

Grimmer, J. and Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3):267–297.

Horiuchi, Y. and Lee, S. (2008). The presidency, regionalism, and distributive politics in South Korea. *Comparative Political Studies*, 41(6):861–882.

Iyengar, S., Sood, G., and Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, 76(3):405–431.

Jeon, D., Kim, J., Lee, S., Cho, W. I., and Seo, J. (2022). KOTE: Korean online comments Tone and Emotion dataset. *arXiv preprint arXiv:2205.05300*.

Keltner, D., Brooks, J. A., and Cowen, A. S. (2023). Semantic space theory: Data-driven insights into basic emotions. *Current Directions in Psychological Science*, 32(3):225–233.

Lazarus, R. S. (1991). *Emotion and Adaptation*. Oxford University Press, New York.

Lee, F. E. (2009). *Beyond Ideology: Politics, Principles, and Partisanship in the U.S. Senate*. University of Chicago Press, Chicago.

Lee, F. E. (2016). *Insecure Majorities: Congress and the Perpetual Campaign*. University of Chicago Press, Chicago.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Lijphart, A. (2012). *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*. Yale University Press, New Haven, CT, 2nd edition.

Mason, L. (2018). *Uncivil Agreement: How Politics Became Our Identity*. University of Chicago Press, Chicago.

Mayhew, D. R. (1974). *Congress: The Electoral Connection*. Yale University Press, New Haven, CT.

Miller, W. E. and Stokes, D. E. (1963). Constituency influence in congress. *American Political Science Review*, 57(1):45–56.

Osnabrügge, M., Hobolt, S. B., and Rodon, T. (2021). Playing to the gallery: Emotive rhetoric in parliaments. *American Political Science Review*, 115(3):885–899.

Park, C. W. (1988). Legislators and their constituents in South Korea: The patterns of district representation. *Asian Survey*, 28(10):1049–1065.

Park, J. Y. (2021a). When do politicians grandstand? measuring message politics in committee hearings. *Journal of Politics*, 83(1):214–228.

Park, S. (2021b). Elite polarization in South Korea: Evidence from a natural language processing model. *Journal of East Asian Studies*, 21(2):245–267.

Pitkin, H. F. (1967). *The Concept of Representation*. University of California Press, Berkeley, CA.

Proksch, S.-O. and Slapin, J. B. (2012). Institutional foundations of legislative speech. *American Journal of Political Science*, 56(3):520–537.

Rask, M. and Hjorth, F. (2025). Partisan conflict in nonverbal communication. *Political Science Research and Methods*.

Rheault, L., Beelen, K., Cochrane, C., and Hirst, G. (2016). Measuring emotion in parliamentary debates with automated textual analysis. *PLoS ONE*, 11(12):e0168843.

Smith, C. A. and Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48(4):813–838.

Stapleton, C. E. and Dawkins, R. (2022). Catching my anger: How political elites create angrier citizens. *Political Research Quarterly*, 75(3):730–743.

Stout, C., Leslie, G. J., Phoenix, D., and Schroeder, E. (2025). Emotional representation: Identifying the characteristics and consequences of elected officials mirroring the emotions of their constituents. *Political Psychology*.

Theriault, S. M. (2013). *The Gingrich Senators: The Roots of Partisan Warfare in Congress*. Oxford University Press, New York.

Tuttnauer, O. (2018). If you can beat them, confront them: Party-level analysis of opposition behavior in European national parliaments. *European Union Politics*, 19(2):278–298.

Valentino, N. A., Brader, T., Groenendyk, E. W., Gregorowicz, K., and Hutchings, V. L. (2011). Election night's alright for fighting: The role of emotions in political participation. *Journal of Politics*, 73(1):156–170.

Webster, S. W. (2020). *American Rage: How Anger Shapes Our Politics*. Cambridge University Press, New York.

Yildirim, T. M., Traber, D., and Rauh, C. (2025). Emotions in the aisles: Unpacking the use of emotive language in the UK House of Commons. *European Journal of Political Research*, 64(1):1–22.

# Supporting Information

## What Drives Legislative Anger? Multimodal Evidence from South Korea's National Assembly

*Intended for online publication*

# Contents of the Supporting Information

# Appendix A: Sample Descriptives

Table 6: Sample Composition

| | Legislators | Segments |
|---|---|---|
| *Panel A: Full sample* | | |
| Total | 250 | 11,462 |
| Meetings | | 156 |
| Segments per legislator (median) | | 31.5 |
| *Panel B: Gender* | | |
| Male | 202 (80.8%) | 9,584 (83.6%) |
| Female | 48 (19.2%) | 1,878 (16.4%) |
| *Panel C: Party bloc* | | |
| Conservative | 92 (36.8%) | 4,224 (36.8%) |
| Democratic | 101 (40.4%) | 4,408 (38.5%) |
| Other / independent | 57 (22.8%) | 2,830 (24.7%) |
| *Panel D: Election type* | | |
| Direct (district) | 206 (82.4%) | 9,819 (85.7%) |
| Proportional | 44 (17.6%) | 1,643 (14.3%) |
| *Panel E: Topic distribution* | | |
| Policy | | 4,051 (35.3%) |
| Procedural | | 2,804 (24.5%) |
| Event | | 2,114 (18.4%) |
| Scandal | | 1,727 (15.1%) |
| Reform | | 364 (3.2%) |
| Unknown | | 402 (3.5%) |
| *Panel F: Period* | | |
| Pre-inauguration (Jun 2016 – May 2017) | | 3,227 (28.2%) |
| Post-inauguration (May 2017 – May 2020) | | 8,235 (71.8%) |

*Notes:* The raw corpus contains 11,587 segments from 250 legislators. Of these, 125 segments with missing audio data are excluded, yielding the 11,462-segment sample reported here. A further 402 segments whose topic label did not match one of the five predefined categories are excluded from all analyses, yielding an analytic sample of 11,060 segments from 247 unique legislators.

# Appendix B: Measurement: Descriptive Statistics, Text versus Delivery, and Validity

This appendix collects supporting material for the measurement strategy described in Section 4: descriptive statistics for the raw anger measures, the analysis of the gap between scripted and delivered anger, estimation results using each measure as the outcome, and validity evidence for the multimodal measures.

## B.1 Descriptive statistics for the raw anger measures

Before standardization, the two delivery channels operate on similar absolute means but differ in spread. Voice anger (mean Hume score across the segment's duration) has mean 0.069, standard deviation 0.050, and range 0.001–0.507; face anger has mean 0.075, standard deviation 0.020, and range 0.020–0.286; text anger (the sum of KOTE's anger-family label probabilities) has mean 1.69 and standard deviation 1.32 ($N = 11{,}060$ segments; 11,057 for text). Voice anger thus has more than twice the standard deviation of face anger and a much wider range, indicating greater variability in vocal anger expression. Because the raw scores are produced by separate models with different architectures and training data, differences in their distributional properties may reflect model calibration as much as behavior; this is one motivation for the within-channel standardization used throughout the paper, and all comparisons across measures use within-measure standardization.

## B.2 The script–delivery gap and institutional position

Section 4 reports that, conditional on writing an angry script (top-quartile text anger), governing-party legislators are substantially less likely than opposition legislators to deliver it angrily. Figure 4 presents the underlying comparison. In the two-party sample, 2,269 segments carry top-quartile text anger. Opposition legislators deliver these scripts with top-quartile voice anger 45% of the time, against 34% for governing-party legislators; the corresponding rates are 25% versus 23% for the face and 24% versus 17% for the composite. In linear probability models with legislator and meeting fixed effects, estimated on the angry-script subsample with standard errors clustered on legislator, governing status reduces the probability of angry delivery by 12.2 percentage points in the voice (SE 4.3), 15.4 in the face (SE 5.7), and 13.5 in the composite (SE 3.7).

The pattern is not sensitive to the choice of threshold. At the 90th percentile, 76% of text-flagged segments lack an angry voice and 89% lack an angry face; at a median split, the same ordering across institutional positions holds.

## B.3 Estimation with each measure as the outcome

Table 7 runs the paper's two main specifications with each of the four measures as the outcome, all on the common percentile-rank scale so that coefficients are directly comparable across columns. The text measure detects both patterns, and detects them more strongly than any delivery channel: scandal raises text anger by 23.3 percentile points, against 13.8 for voice, 2.0 for face, and 7.9 for the composite, and governing status lowers text anger by 11.0 points, against 7.8, 4.0, and

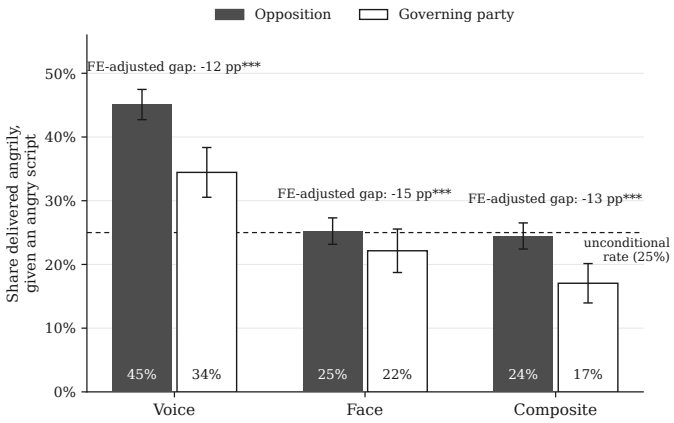


Figure 4: The script–delivery gap by institutional position. Among segments with top-quartile text anger ($N = 2{,}269$ in the two-party sample), bars show the share also delivered with top-quartile anger in each channel, by institutional position; whiskers are 95% confidence intervals. By construction the unconditional rate is 25% (dashed line). FE-adjusted gaps are from linear probability models with legislator and meeting fixed effects, standard errors clustered on legislator. $^{***}p < 0.01$.

5.9. This ordering is what the scripted-channel account predicts. Text is the most deliberate layer of expression, so when institutional incentives make anger useful, word choice moves most. The comparison implies that a text-only design would overstate how strongly expressed anger responds to institutional incentives, in addition to carrying the position-dependent measurement error documented in Section B.2.

## B.4 Validity of the multimodal measures

**Face validity.** The mean level of every emotion dimension paints a recognizably parliamentary portrait. The voice channel is dominated by determination, interest, and concentration, with anger ranking 6th of 48 dimensions; the face channel is dominated by concentration and calmness, with anger ranking 37th. The models thus describe a professionalized institution in which anger is a real but regulated presence, more audible than visible. This asymmetry is consistent with the controllability hierarchy discussed in the main text, since the face is the channel speakers manage most deliberately.

**Convergent and discriminant validity.** If the anger scores measure anger rather than generic activation, they should co-move with the other adversarial, high-arousal negative emotions and move against positive and low-arousal states. Figure 5 confirms both patterns in both channels. Within the voice, anger correlates with contempt at $r = 0.67$, disgust at 0.66, and distress at 0.64, and negatively with calmness ($-0.48$), interest ($-0.27$), and joy ($-0.18$). The facial measure shows the same structure (disgust 0.83, negative surprise 0.62; calmness $-0.61$, joy $-0.51$). Anger is nearly orthogonal to determination in both channels ($r = 0.12$ voice, 0.11 face), so the measure separates angry speech from merely forceful, emphatic speech. This distinction matters: Appendix C shows that a composite of low-level acoustic arousal features fails exactly this test.

Table 7: Topic and Governing-Status Effects Across the Four Outcome Measures

| | Text (KOTE) | Voice (Hume) | Face (Hume) | Composite (Voice + Face) |
|---|---|---|---|---|
| *Panel A: Topic effects (procedural omitted), full sample* | | | | |
| SCANDAL | $+0.233^{***}$ | $+0.138^{***}$ | $+0.020^{*}$ | $+0.079^{***}$ |
| | (0.012) | (0.014) | (0.011) | (0.010) |
| EVENT | $+0.114^{***}$ | $+0.107^{***}$ | $+0.012$ | $+0.059^{***}$ |
| | (0.014) | (0.014) | (0.012) | (0.011) |
| POLICY | $+0.085^{***}$ | $+0.072^{***}$ | $+0.028^{***}$ | $+0.050^{***}$ |
| | (0.011) | (0.013) | (0.011) | (0.010) |
| REFORM | $+0.072^{***}$ | $+0.103^{***}$ | $+0.026$ | $+0.065^{***}$ |
| | (0.019) | (0.020) | (0.016) | (0.014) |
| *Panel B: Governing-party status, two-party sample* | | | | |
| GOV_TV | $-0.110^{***}$ | $-0.078^{***}$ | $-0.040^{**}$ | $-0.059^{***}$ |
| | (0.019) | (0.014) | (0.018) | (0.013) |
| Legislator FE | Yes | Yes | Yes | Yes |
| Meeting FE | Yes | Yes | Yes | Yes |
| *N* (Panel A / B) | 11,057/8,290 | 11,060/8,293 | 11,060/8,293 | 11,060/8,293 |

*Notes:* All outcomes are within-measure percentile ranks (0–1). Two-way legislator and meeting fixed effects; standard errors clustered on legislator. The text column loses three segments with no scoreable transcript. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

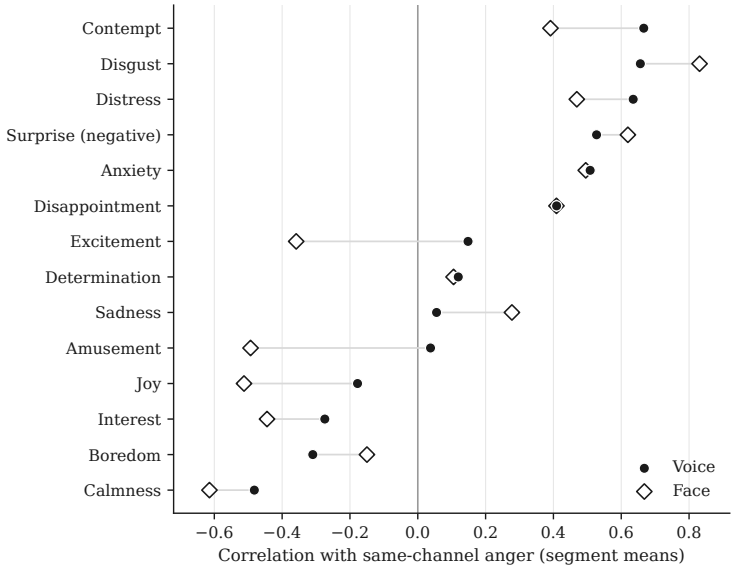


Figure 5: Correlation of each emotion's segment-mean score with same-channel anger, for the voice (filled circles) and face (open diamonds), $N = 11{,}060$. Adversarial negative emotions co-move with anger; positive and low-arousal states move against it; determination is nearly orthogonal.

**Behavioral validity.** The measures also track independently coded speech behavior. As part of topic classification, the LLM separately labeled each segment's rhetorical mode and hostility without access to any emotion score. Segments coded as ACCUSATION show the highest mean voice anger of any major mode (percentile rank 0.67, against 0.48 for QUESTIONING and 0.23 for ANSWERING), and the 209 segments coded as openly hostile average 0.68 on voice anger and 0.59 on the composite, against baselines of 0.50.

**Cross-cultural calibration.** Hume's models are grounded in cross-cultural evidence on the recognition of emotional expression (Cowen et al., 2021). Our design does not, however, depend on the models being perfectly calibrated for Korean speakers. All regressions include legislator fixed effects, which absorb any speaker-specific measurement bias, and outcomes are standardized within channel, so only the rank-ordering of anger within the sample matters for the estimates. For measurement error to bias our results, a model would need to misread the same legislator differently depending on the topic under discussion within the same session. The open-source replications in Appendix C, including an English-language speech model applied to Korean audio, provide a direct check on this possibility.

# Appendix C: Robustness to Alternative, Open-Source Emotion Measurement

A natural concern is that our findings are an artifact of Hume AI's proprietary, black-box models. We re-measure anger on the same 11,060 segments with independent, open-source models and rerun the paper's regressions, keeping legislator and meeting fixed effects and clustering on legislator throughout. Two choices shape the design. We score each nonverbal channel with a model built for that channel: a wav2vec2 speech-emotion model fine-tuned on IEMOCAP for voice (Baevski et al., 2020; Yang et al., 2021), and py-feat, an AffectNet-trained facial classifier, for the face (Cheong et al., 2023). Each check then tests the construct the main text measures in that same channel. We also use trained emotion classifiers rather than composites of low-level features from OpenFace (Baltrušaitis et al., 2018) or openSMILE (Eyben et al., 2016); Section C.3 shows that those composites track arousal and muscle movement, not anger. We also score the transcripts with KOTE, a Korean text-emotion model (Jeon et al., 2022). Because text measures the deliberate content of speech

rather than its delivery, we analyze it separately in Appendix D, where it serves as a comparison with the text-based measurement common in the literature.

## C.1 Same-modality replication with independent emotion models

Table 8 re-estimates the scandal effect (Section 5) and the impeachment governing-status effect (Section 6) with each open-source model as the outcome. The outcomes are percentile-ranked and the specification is the main-text model (topic dummies with procedural omitted; the two-party sample for GOV_TV); the scandal column reports the scandal coefficient from that model.

Table 8: Same-modality replication: independent emotion models vs. Hume

| Channel | Measure | Corr. w/ Hume | Scandal $\beta$ | GOV_TV $\beta$ |
|---|---|---|---|---|
| Voice | Hume (original) | — | $+0.138^{***}$ | $-0.078^{***}$ |
| Voice | wav2vec2 (IEMOCAP) | $r$=0.37 | $+0.086^{***}$ | $-0.058^{***}$ |
| Face | Hume (original) | — | $+0.020^{*}$ | $-0.040^{**}$ |
| Face | py-feat (AffectNet) | $r$=0.20 | +0.006 | +0.011 |

*Notes:* Percentile-rank outcomes; legislator + meeting FE; SE clustered on legislator ($N =$ 11,060 for scandal; $N =$ 8,293 for GOV_TV). Procedural is the omitted topic. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

**Voice replicates across models and languages.** An English-language model applied to Korean speech without adaptation recovers both headline patterns with the same signs and significance as Hume. The agreement is specific to anger: wav2vec2's anger probability tracks Hume voice anger at $r$=0.37, while its neutral, happy, and sad probabilities all correlate negatively with it. A model that has never encountered Korean still reproduces the substantive results, which is hard to square with a proprietary-model artifact.

**The face carries a weaker scandal signal.** Hume's facial anger rises modestly during scandal (+0.020, $p < 0.10$), far below the voice (+0.138), and py-feat shows no scandal effect at all (+0.006, n.s.). Both models agree that the face carries much less scandal-specific anger than the voice, consistent with facial expression being the more controllable channel; the two facial measures also agree less with each other than the two voice measures do ($r$=0.20 against 0.37), as one expects given the difficulty of facial-expression recognition on parliamentary video.

**Joint activation survives independent measurement.** The joint elevation of voice and face anger during scandal requires both channels to be measured well at once, so it is the most demanding test. Under Hume the scandal effect on joint activation (both channels in their top quartile) is $+0.040^{***}$; swapping both channels for the open-source pair leaves it at $+0.024^{**}$, smaller, as expected with a noisier facial measure, but still significant. The cross-channel pattern is therefore not an artifact of the proprietary model.

## C.2 Discriminant validity: a double dissociation

Emotion classifiers and raw-feature composites do not just differ in noise; they measure different things, and the difference is plain when we compare scandal with national-crisis events. Both topics are emotionally charged. Only one is angry (Table 9).

All three emotion models (Hume, the English wav2vec2, and the Korean text model) place anger highest on scandal. Raw acoustic arousal does the opposite, peaking during national-crisis events such as the Sewol ferry disaster and the North Korea standoffs, which are intense but not angry. The scandal effect therefore reflects emotion-specific anger, not how loudly a legislator happens to be speaking.

Table 9: Discriminant validity: scandal vs. national-crisis event

| Measure | Type | Scandal vs. Event |
|---|---|---|
| Hume voice | emotion model | scandal $>$ event ($t = 6.0$) |
| wav2vec2 voice | emotion model | scandal $>$ event ($t = 5.9$) |
| KOTE text | emotion model | scandal $>$ event ($t = 20.3$) |
| openSMILE arousal | acoustic feature | **event** $>$ **scandal** ($t = -5.3$) |
| OpenFace AU | muscle feature | scandal $>$ event ($t = 4.5$) |

*Notes:* Standardized topic means, two-sample $t$. Raw-feature comparators (openSMILE, OpenFace) from the local pipeline on the same segments.

## C.3 Emotion-specific anger versus raw acoustic and facial features

The headline regressions on the raw-feature composites separate the two channels. For voice the dissociation is clean. The acoustic-arousal composite from openSMILE (Eyben et al., 2016) (F0, loudness, jitter, shimmer, $-$HNR, $-$spectral slope) shows no scandal effect ($-0.004$, n.s.) and no governing-status effect ($+0.003$, n.s.), even though both voice emotion models reproduce both. Loudness and pitch are not what carry the scandal signal.

The face behaves differently, and the difference is itself evidence. The FACS muscle composite from OpenFace (Baltrušaitis et al., 2018) (AU4, AU5, AU7, AU23) rises sharply during scandal ($+0.057$, $p < 0.01$): the brow lowers, the lids tighten, the lips press. Yet the facial emotion models barely register it as anger (Hume $+0.020$; py-feat $+0.006$, n.s.). Scandal moves the face, but the movement does not resolve into a display the emotion models call angry. This is what facial management looks like in the data, and it matches the controllability account of Section 4.3: the face is the most suppressible channel, so what survives is raw muscle activation rather than a legible anger expression. The underlying signal is present before fixed effects in both channels — AU4 rises during scandal ($t = 6.5$), so does the acoustic composite ($t = 3.0$, $r = 0.25$ with Hume voice anger). What an emotion model adds is the anger-specific, within-legislator, within-session component; for the voice that component survives, and for the face it stays faint.

## C.4 Interpretation and caveats

Two emotion models built on different data and architectures reproduce the voice results, one of them an English model applied to Korean, and the joint voice–face activation survives swapping both channels for open-source models. A raw acoustic-arousal composite does not reproduce the scandal effect, and a Korean text model recovers the same substantive pattern. Together these point to emotion-specific content behind the findings, not a Hume-specific artifact or a loudness effect.

**Caveats.** wav2vec2 is trained on English, so applying it to Korean is a conservative test in which any cross-lingual mismatch works against finding agreement. For a handful of very long speeches we score voice emotion on the first 120 seconds; these speeches are rare and the restriction does not move the estimates. Open-source facial-emotion recognition is the lowest-fidelity component under parliamentary video; even so, the joint-activation result survives the substitution.

# Appendix D: Comparison with Text-Based Emotion Measurement

Most existing measurement of legislative emotion works from text (Grimmer and Stewart, 2013). A fair question is whether the multimodal apparatus earns its cost: the transcripts are already in hand, so why not measure anger in what legislators say, as the literature does? We answer by scoring every transcript with KOTE, a Korean text-emotion model (Jeon et al., 2022), summing its anger-family labels, and running the paper's regressions on the result. Table 7 in Appendix B reports the estimates in full alongside the delivery channels.[16]

Text-based anger is not silent. It tracks the same two patterns as the delivery channels, and tracks them more sharply. Scandal raises text anger by 23.3 percentile points, against 13.8 for voice; governing-party status lowers it by 11.0 points, against 7.8 for voice. Pairwise, text scandal anger outranks every other topic by a wide margin (scandal $-$ event $+0.119$, $t = 9.8$; scandal $-$ policy $+0.148$, $t = 15.7$; scandal $-$ reform $+0.161$, $t = 9.4$; all $p < 0.01$), and text separates scandal from national-crisis events more cleanly than any other measure ($t = 20.3$). A scholar working only from the transcripts would already see that legislators write angrier speeches about scandal, and that opposition legislators write angrier speeches than the governing party.

Two things follow. First, the ordering is what the controllability hierarchy of Section 4.3 predicts. Text is the most deliberate channel, scripted and filed in writing before the session, so when institutional incentives make anger useful, word choice is the easiest lever to pull and it is the lever that moves most. Second, text records that anger is present but cannot show how it is delivered. A written line of moral condemnation looks the same whether the legislator reads it in a level voice or with a raised voice and a visibly angry face. What our framework adds is the pattern of activation across delivery channels, and that cross-channel pattern, not the presence of angry words, is what reveals the institutional structure of legislative anger.

# Appendix E: Topic Classification Prompt

Each of the 11,462 speech segments was classified using Claude Haiku (model `claude-haiku-4-5-20251001`, Anthropic) via the Anthropic Messages API. The classification script is included in the replication materials (`classify_segments.py`). Each segment was classified into one of five topic categories (plus a residual UNKNOWN label) according to what the speaker discusses. For segments under 20 characters (e.g., "yes," "thank you"), the segment was automatically coded as PROCEDURAL without an API call. For segments exceeding 3,000 characters, the middle portion was truncated to keep the first and last 1,500 characters.

Each API call received a system prompt specifying the topic categories and a user message containing the speaker name, meeting name, and segment transcript text. The model returned a JSON object with the topic label. Invalid responses were parsed with fallback regex matching. A total of 402 segments (3.5%) received a topic label that did not match one of the five predefined categories and are coded as unknown topic in Table 6; these segments are excluded from all

[16] Three segments whose ASR transcripts contain only a two- or three-character fragment carry no scoreable text and are dropped from the text models, leaving $N = 11{,}057$ for the topic regressions. They are retained in the voice and face channels, which score the audio and video stream directly.

analyses.

The full system prompt is reproduced below.

```
You classify speech segments from South Korea's 20th National Assembly (2016-2020).
Classify each segment by TOPIC. === TOPIC (what the speaker discusses) ===

SCANDAL -- Political scandal, corruption, or misconduct.  Includes:  Choi Soon-sil,
impeachment, Cho Guk, corruption (bribery, abuse of power, political funds, family
scandals, election violations), any accusation of personal/party misconduct or ethical
violation.

EVENT -- Politicized national events:  Sewol ferry, THAAD, North Korea/inter-Korean
relations, national security crises, economic crises, COVID. REFORM -- Major institutional/legisla
reforms:  electoral reform, fast-track legislation, prosecution reform, constitutional
amendment.

POLICY -- Substantive policy discussion not tied to scandal/event/reform:  economy,
labor, housing, healthcare, education, environment, technology, agriculture, defense/security
policy, welfare, trade, etc.

PROCEDURAL -- Session management, agenda setting, vote counting, scheduling.  No
substantive content.  === RESPONSE FORMAT === Respond ONLY with valid JSON: {"topic":

"TOPIC"} If a segment covers multiple topics, choose the DOMINANT one.
```

The user message for each segment followed this template:

```
Speaker:  [legislator name]
Meeting:  [meeting name]

[segment transcript text]
```

# Appendix F: Multi-Emotion Validation

If the scandal anger finding reflects a pattern specific to adversarial emotions rather than a general measurement artifact, the same analysis applied to other emotions should produce a clear contrast: adversarial emotions should show elevated joint voice–face activation during scandal, while positive emotions should not. Table 10 reports the scandal coefficient from the same two-way fixed effects specification (legislator + meeting FE, cluster-robust SEs) for eight emotions, each measured independently across voice and face.

The results support the specificity of the anger finding across several dimensions.

Adversarial emotions show joint voice–face elevation during scandal; positive emotions show the opposite pattern. The continuous Voice $\times$ Face measure captures this: anger ($+0.079^{***}$), disgust ($+0.078^{***}$), fear ($+0.053^{***}$), and contempt ($+0.039^{***}$) all show significant positive joint activation, while positive emotions show significant negative effects: joy ($-0.024^{**}$) and calmness ($-0.062^{***}$). If scandal merely increased general physiological arousal, all emotions would increase uniformly. The sharp contrast between adversarial and positive emotions indicates that scandal activates a specific emotional register.

Two individual emotions merit comment. Contempt shows a channel dissociation: scandal strongly increases vocal contempt ($+0.093$***) but reduces facial contempt ($-0.030$**). Facial contempt involves a distinctive unilateral lip raise that is particularly visible and socially costly,

Table 10: Scandal Effect Across Voice and Face Emotions

| Emotion | Type | Voice (pct rank) | Face (pct rank) | Voice × Face (continuous) |
|---|---|---|---|---|
| Anger | adversarial | $+0.138^{***}$ | $+0.020^{*}$ | $+0.079^{***}$ |
| | | (0.014) | (0.011) | (0.012) |
| Fear | adversarial | $+0.067^{***}$ | $+0.040^{***}$ | $+0.053^{***}$ |
| | | (0.013) | (0.010) | (0.009) |
| Disgust | adversarial | $+0.111^{***}$ | $+0.044^{***}$ | $+0.078^{***}$ |
| | | (0.013) | (0.010) | (0.010) |
| Contempt | adversarial | $+0.093^{***}$ | $-0.030^{**}$ | $+0.039^{***}$ |
| | | (0.013) | (0.012) | (0.012) |
| Sadness | negative | $+0.001$ | $+0.044^{***}$ | $+0.017^{**}$ |
| | | (0.014) | (0.011) | (0.009) |
| Joy | positive | $-0.019$ | $-0.023^{**}$ | $-0.024^{**}$ |
| | | (0.014) | (0.010) | (0.010) |
| Triumph | positive | $+0.022$ | $-0.022^{**}$ | $-0.001$ |
| | | (0.014) | (0.010) | (0.009) |
| Calmness | positive | $-0.076^{***}$ | $-0.040^{***}$ | $-0.062^{***}$ |
| | | (0.018) | (0.009) | (0.013) |

*Notes:* Each cell reports the scandal coefficient from a separate regression with legislator and meeting fixed effects. Procedural segments are the omitted topic category. Voice × Face = product of voice and face percentile ranks. Standard errors clustered at legislator level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

which may explain why legislators suppress it even during scandal. Sadness shows the reverse pattern: vocal sadness is unaffected (+0.001, n.s.) while facial sadness is elevated (+0.044***), suggesting that scandal evokes visible facial sadness without the vocal counterpart.

# Appendix G: Robustness to Duration Controls

One might worry that the topic effects in Table 1 partly reflect speech duration: scandal discussions tend to be longer than procedural business, and longer segments mechanically produce more vocal variation. We address this by adding log speech duration as a control to both the GOV_TV and the topic specifications.

**GOV_TV.** Table 11 re-estimates the GOV_TV effect under three specifications: the baseline with no additional controls, adding topic indicators, and adding both topic indicators and log speech duration.

Table 11: GOV_TV Effect Under Alternative Specifications

| | (1) Baseline | (2) + Topic | (3) + Topic + Duration |
|---|---|---|---|
| *Panel A: Voice anger (pct rank)* | | | |
| GOV_TV | $-0.078^{***}$ | $-0.075^{***}$ | $-0.096^{***}$ |
| | (0.013) | (0.013) | (0.015) |
| *Panel B: Face anger (pct rank)* | | | |
| GOV_TV | $-0.040^{**}$ | $-0.040^{**}$ | $-0.037^{**}$ |
| | (0.017) | (0.017) | (0.017) |
| Legislator FE | Yes | Yes | Yes |
| Meeting FE | Yes | Yes | Yes |
| Topic controls | No | Yes | Yes |
| Duration controls | No | No | Yes |
| *N* | 8,293 | 8,293 | 8,293 |

*Notes:* Each cell reports the GOV_TV coefficient from a separate regression. Sample restricted to two major-party blocs ($N = 8{,}293$; 190 legislators; 135 meetings). Topic controls include indicators for scandal, event, policy, and procedural segments. Duration controls include log speech duration. Standard errors clustered at legislator level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The GOV_TV estimate is robust across specifications. Adding topic controls reduces it modestly ($-0.078$ to $-0.075$). Adding duration controls actually *strengthens* the voice estimate to $-0.096$, because governing legislators tend to give longer speeches (ministerial responses, policy explanations), and longer speeches carry more vocal anger; controlling for duration removes this positive confound on the governing side. Face anger is stable across all three specifications

($-0.037$ to $-0.040$, $p < 0.05$). The institutional role effect on voice anger is not an artifact of topic composition or speech duration.

**Topic effects.** Table 12 adds log speech duration as a control to the topic regressions, reporting all four outcomes (voice and face, mean and peak).

Table 12: Topic Effects Controlling for Speech Duration

| | (1) Voice anger (mean) | (2) Voice anger (peak) | (3) Face anger (mean) | (4) Face anger (peak) |
|---|---|---|---|---|
| SCANDAL | $+0.082^{***}$ | $+0.083^{***}$ | $+0.027^{**}$ | $+0.044^{***}$ |
| | (0.015) | (0.012) | (0.012) | (0.011) |
| EVENT | $+0.043^{**}$ | $+0.057^{***}$ | $+0.020$ | $+0.045^{***}$ |
| | (0.017) | (0.014) | (0.014) | (0.014) |
| POLICY | $+0.026^{*}$ | $+0.037^{***}$ | $+0.033^{***}$ | $+0.050^{***}$ |
| | (0.015) | (0.012) | (0.012) | (0.012) |
| REFORM | $+0.043^{**}$ | $+0.042^{**}$ | $+0.034^{**}$ | $+0.045^{***}$ |
| | (0.020) | (0.017) | (0.017) | (0.016) |
| log(duration) | $+0.069^{***}$ | $+0.150^{***}$ | $-0.009^{*}$ | $+0.091^{***}$ |
| | (0.007) | (0.006) | (0.005) | (0.006) |
| Legislator FE | Yes | Yes | Yes | Yes |
| Meeting FE | Yes | Yes | Yes | Yes |
| *N* | 11,060 | 11,060 | 11,060 | 11,060 |

*Notes:* Each column is a separate regression. Outcomes are percentile-ranked anger scores. Procedural segments are the omitted category. Standard errors clustered at legislator level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Compared with the baseline, voice mean anger attenuates by roughly 40% (scandal drops from $+0.138$ to $+0.082$) but the topic gradient is preserved: scandal remains the highest voice anger topic at $p < 0.01$, and the full ordering scandal $>$ event $>$ policy holds. Face mean anger is essentially unchanged by duration controls: coefficients are stable or slightly larger (scandal $+0.020$ to $+0.027$), consistent with face anger being unrelated to speech length. Face peak anger attenuates from $\sim$0.12 to $\sim$0.04 but remains significant and uniform across topics ($+0.044$ to $+0.050$), preserving the pattern observed in the baseline. This attenuation is itself informative: because the segment peak is a maximum over frames, it is mechanically increasing in speech length, so much of the raw face-peak elevation reflects segment duration rather than topic-specific anger. What survives the duration control is the cross-topic uniformity, not a scandal-specific facial signal—consistent with reading the face peaks as evidence that the face does not encode the topic gradient the voice does, rather than as a legible anger display that is actively suppressed.

The duration coefficient itself differs across channels. For voice, log duration is positive and large ($+0.069^{***}$ for mean, $+0.150^{***}$ for peak): longer speeches carry more vocal anger. For face mean anger, duration is negligible ($-0.009$), confirming that facial anger is not driven by the structural features of speech delivery that govern vocal expression.

# Appendix H: Scandal Effect by Party Bloc

The main text argues that scandal lowers the cost of anger expression for all legislators, not just the opposition. Table 13 tests this claim by adding a SCANDAL × Conservative interaction to the topic regression in Table 1. If the scandal effect on anger expression were concentrated in one party, the interaction term would be large and statistically significant.

The interaction is small and insignificant for both outcomes: voice anger ($\delta = -0.015$, $p = 0.40$) and face anger ($\delta = -0.014$, $p = 0.45$). The scandal effect on anger expression does not differ significantly between conservative and democratic legislators, in line with the theoretical expectation that attributable blame lowers the cost of anger expression symmetrically across parties.

Table 13: Scandal × Party Bloc Interaction

| | (1) Voice anger (pct rank) | (2) Face anger (pct rank) |
|---|---|---|
| SCANDAL | $+0.139^{***}$ | $+0.021$ |
| | (0.017) | (0.017) |
| EVENT | $+0.102^{***}$ | $+0.004$ |
| | (0.015) | (0.011) |
| POLICY | $+0.068^{***}$ | $+0.025^{***}$ |
| | (0.014) | (0.009) |
| REFORM | $+0.083^{***}$ | $+0.008$ |
| | (0.022) | (0.017) |
| SCANDAL × Conservative | $-0.015$ | $-0.014$ |
| | (0.018) | (0.018) |
| Legislator FE | Yes | Yes |
| Meeting FE | Yes | Yes |
| $N$ | 8,293 | 8,293 |

*Notes:* Same specification as Table 1, with an additional SCANDAL × Conservative interaction term. The interaction tests whether the scandal effect on anger expression differs between conservative and democratic legislators. Procedural segments are the omitted category. Percentile ranks are computed within the full sample before subsetting to the two-party sample ($N = 8{,}293$). Standard errors clustered at legislator level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

# Appendix I: Constituency Tests: District versus Proportional-Representation Members

This appendix reports the full estimates behind the constituency tests in the main text. Table 14 presents the complete topic × PR interaction model summarized in Table 3. Panel A reports the topic effects for district members, the reference group; Panel B reports the differences for proportional-representation members. The district-member effects in Panel A are nearly identical to the full-sample estimates in Table 1, and no difference in Panel B approaches significance. Table 15 reports the corresponding test for the impeachment quasi-experiment. The governing-status effect for district members matches the full-sample estimate, and the PR interaction is small and insignificant on every outcome, so members with and without districts respond to the role reversal alike.

# Appendix J: Inter-Coder Reliability

We evaluate the reliability of the LLM topic classification by comparing it against independent human coding. Two coders—one of the authors (Coder A) and a trained research assistant (Coder B)—independently classified a stratified random sample of 200 segments drawn from the analytic corpus. The sample was stratified by LLM-assigned topic (40 segments each from SCANDAL, EVENT, POLICY, and PROCEDURAL; 20 each from REFORM and UNKNOWN) to ensure adequate coverage of rarer categories. Neither coder had access to the LLM labels or to each other's codes. Coders read the Korean transcript of each segment and assigned one of six labels: the five substantive categories used in the analysis plus UNKNOWN for segments that did not clearly fit any category.

**Agreement metrics.** Table 16 reports pairwise and three-way agreement statistics. On the five substantive categories (excluding segments labeled UNKNOWN by either member of the pair), all three pairwise Cohen's $\kappa$ values fall between 0.76 and 0.78, indicating substantial agreement (Landis and Koch, 1977). Krippendorff's $\alpha$ across all three raters is 0.80 on the five-way scheme, meeting the conventional reliability threshold (Krippendorff, 2011). For the binary scandal-vs.-other distinction—the classification most critical to the paper's inferences—pairwise raw agreement ranges from 89% to 92%, with $\kappa$ between 0.72 and 0.78.

**Confusion patterns.** Table 17 reports the confusion matrix between the LLM and each human coder. Three patterns emerge. First, the primary source of disagreement is the residual UNKNOWN category, which all three raters apply inconsistently; only 14 segments receive UNKNOWN from both human coders, validating our decision to exclude these segments from all analyses. Second, the most common substantive confusion is between EVENT and POLICY, a genuine boundary ambiguity that does not threaten our inferences because both serve as comparison categories rather than key predictors. Third, the scandal classification is reliable but slightly conservative on the human side. Both coders agree with the LLM on 30 of 40 LLM-labeled SCANDAL segments (75% recall), and most disagreements are reclassifications to UNKNOWN rather than to another substantive category. The LLM is thus, if anything, more liberal with the scandal label than human

Table 14: Topic Effects by Electoral Type: Full Estimates

| | (1) Voice anger ($z$) | (2) Face anger ($z$) | (3) Composite ($z$) |
|---|---|---|---|
| Panel A. Topic effects, district members | | | |
| SCANDAL | $+0.359^{***}$ | $+0.071^{*}$ | $+0.215^{***}$ |
| | (0.062) | (0.040) | (0.043) |
| EVENT | $+0.212^{***}$ | $+0.043$ | $+0.127^{***}$ |
| | (0.064) | (0.042) | (0.045) |
| POLICY | $+0.141^{**}$ | $+0.107^{***}$ | $+0.124^{***}$ |
| | (0.058) | (0.038) | (0.040) |
| REFORM | $+0.163^{**}$ | $+0.101^{*}$ | $+0.132^{***}$ |
| | (0.076) | (0.052) | (0.051) |
| Panel B. Differences for proportional-representation members | | | |
| SCANDAL $\times$ PR | $-0.050$ | $-0.028$ | $-0.039$ |
| | (0.124) | (0.075) | (0.076) |
| EVENT $\times$ PR | $+0.085$ | $+0.006$ | $+0.046$ |
| | (0.153) | (0.069) | (0.088) |
| POLICY $\times$ PR | $-0.007$ | $-0.063$ | $-0.035$ |
| | (0.116) | (0.062) | (0.064) |
| REFORM $\times$ PR | $+0.083$ | $-0.091$ | $-0.004$ |
| | (0.174) | (0.143) | (0.121) |
| Legislator FE | Yes | Yes | Yes |
| Meeting FE | Yes | Yes | Yes |
| $N$ | 11,060 | 11,060 | 11,060 |

*Notes:* Single regression per column. Panel A coefficients are topic effects for district members (reference group); Panel B coefficients are the PR differences. The PR main effect is absorbed by legislator fixed effects. Procedural segments omitted. Standard errors clustered at legislator level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 15: Governing-Status Effect by Electoral Type

| | (1) Voice anger ($z$) | (2) Face anger ($z$) | (3) Composite ($z$) |
|---|---|---|---|
| GOV_TV (district members) | $-0.350^{***}$ | $-0.111^{*}$ | $-0.231^{***}$ |
| | (0.060) | (0.062) | (0.050) |
| GOV_TV × PR | $+0.046$ | $-0.129$ | $-0.041$ |
| | (0.130) | (0.099) | (0.089) |
| Legislator FE | Yes | Yes | Yes |
| Meeting FE | Yes | Yes | Yes |
| $N$ | 8,293 | 8,293 | 8,293 |

*Notes:* Two-party subsample (190 legislators; 135 meetings). The first row is the governing-status effect for district members; the second is the difference for proportional-representation members. Standard errors clustered at legislator level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 16: Inter-Coder Reliability: Summary Statistics

| | Coder A vs. LLM | Coder B vs. LLM | Coder A vs. Coder B |
|---|---|---|---|
| *Six categories (including* UNKNOWN*),* $N = 200$ | | | |
| Raw agreement (%) | 71.0 | 68.0 | 72.5 |
| Cohen's $\kappa$ | 0.64 | 0.61 | 0.66 |
| *Five categories (excluding* UNKNOWN*)*[a] | | | |
| Raw agreement (%) | 81.3 | 82.5 | 82.9 |
| Cohen's $\kappa$ | 0.76 | 0.78 | 0.78 |
| *Binary: scandal vs. other,* $N = 200$ | | | |
| Raw agreement (%) | 89.0 | 92.0 | 91.0 |
| Cohen's $\kappa$ | 0.72 | 0.78 | 0.77 |
| *Three-way (all raters)* | | | |
| Krippendorff's $\alpha$ (6-way) | | 0.64 | |
| Krippendorff's $\alpha$ (5-way) | | 0.80 | |
| Krippendorff's $\alpha$ (scandal binary) | | 0.71 | |

[a] Excludes segments labeled UNKNOWN by either member of each pair ($N$ ranges from 149 to 166 depending on the pair).

coders, which would attenuate the estimated scandal effects by including ambiguous segments in the scandal category.

Table 17: Confusion Matrix: LLM (rows) vs. Human Majority (columns)

| | SCAN. | EVENT | POLICY | REFORM | PROC. | UNKN. | Total |
|---|---|---|---|---|---|---|---|
| SCANDAL | **30/30** | 1/1 | 3/2 | 0/0 | 0/0 | 6/7 | 40 |
| EVENT | 3/1 | **27/27** | 8/7 | 0/1 | 1/0 | 1/4 | 40 |
| POLICY | 2/1 | 1/3 | **35/33** | 0/1 | 0/0 | 2/2 | 40 |
| REFORM | 3/0 | 0/1 | 4/2 | **12/11** | 0/1 | 1/5 | 20 |
| PROCEDURAL | 0/1 | 0/0 | 5/5 | 0/0 | **31/26** | 4/8 | 40 |
| UNKNOWN | 4/3 | 1/1 | 7/6 | 0/1 | 1/0 | **7/9** | 20 |
| Total | 42/36 | 30/33 | 62/55 | 12/14 | 33/27 | 21/35 | 200 |

*Note.* Each cell reports Coder A / Coder B counts. LLM–Coder A raw agreement = 142/200 = 71.0%; LLM–Coder B = 136/200 = 68.0%.

**Category-specific precision and recall.** Table 18 reports precision, recall, and $F_1$ for each category, treating the LLM label as the reference. Both coders achieve highest $F_1$ for PROCEDURAL (0.78–0.85) and lowest for UNKNOWN (0.33–0.34), reinforcing the decision to drop the latter. SCANDAL $F_1$ is 0.73 for Coder A and 0.79 for Coder B. The slightly lower precision for Coder A (0.71 vs. 0.83) reflects a modest tendency to assign scandal more liberally (42 vs. 36 segments), but both coders identify the same 30 core scandal segments. Of the 200 segments, 122 (61%) receive

Table 18: Category-Specific Agreement: Human Coders vs. LLM

| | Coder A vs. LLM | | | Coder B vs. LLM | | |
|---|---|---|---|---|---|---|
| | Prec. | Recall | $F_1$ | Prec. | Recall | $F_1$ |
| SCANDAL | 0.71 | 0.75 | 0.73 | 0.83 | 0.75 | 0.79 |
| EVENT | 0.90 | 0.68 | 0.77 | 0.82 | 0.68 | 0.74 |
| POLICY | 0.56 | 0.88 | 0.69 | 0.60 | 0.82 | 0.69 |
| REFORM | 1.00 | 0.60 | 0.75 | 0.79 | 0.55 | 0.65 |
| PROCEDURAL | 0.94 | 0.78 | 0.85 | 0.96 | 0.65 | 0.78 |
| UNKNOWN | 0.33 | 0.35 | 0.34 | 0.26 | 0.45 | 0.33 |

*Note.* LLM label is the reference. Precision = share of human-assigned labels that match the LLM; Recall = share of LLM-assigned labels recovered by the human coder. $N = 200$.

the same label from all three raters, and when a majority exists the LLM agrees with it 87% of the time, against 92% for Coder A and 89% for Coder B. In sum, the LLM topic classification achieves reliability comparable to human coding. The five substantive categories are coded with substantial inter-rater agreement ($\kappa \geq 0.76$; $\alpha = 0.80$), and the scandal classification, the distinction most consequential for the paper's findings, is reliable across all three rater pairs ($\kappa = 0.72$–$0.78$). The main

source of disagreement is the residual UNKNOWN category, which we exclude from the analytic sample.

# Appendix K: Event Study Estimates

Table 19 reports the estimates plotted in Figure 3.

Table 19: Event Study: Conservative–Democratic Gap in Voice Anger by Period

| | Voice anger ($z$) |
|---|---|
| Pre-scandal (Jun–Sep 2016) | [reference] |
| Transition (Oct 16–May 17) | $+0.384^{**}$ |
| | (0.150) |
| H2 2017 (Jun–Dec 17) | $+0.947^{***}$ |
| | (0.184) |
| 2018 (Jan–Dec 18) | $+0.971^{***}$ |
| | (0.195) |
| H1 2019 (Jan–Jun 19) | $+0.977^{***}$ |
| | (0.211) |
| H2 2019–2020 (Jul 19–May 20) | $+0.989^{***}$ |
| | (0.186) |

*Notes:* Coefficients show how the conservative–democratic gap in voice anger ($z$-score scale) changed relative to the pre-scandal baseline. The baseline gap is negative; in the reference period conservatives held the presidency and democrats, in opposition, were the angrier bloc by 0.27 SD in raw means. The conservative main effect is absorbed by the legislator fixed effects. Sample restricted to two major-party blocs ($N = 8{,}293$; 190 legislators; 135 meetings). Key events: Choi Soon-sil scandal breaks October 2016; Moon Jae-in inaugurated May 2017; Cho Guk controversy August 2019. All models include legislator and meeting FE with cluster-robust SEs. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

# Appendix L: Inference Corrections for the Single-Shock Design

The quasi-experiment rests on a single institutional shock that moved the two party blocs in opposite directions, so conventional legislator-clustered standard errors may understate uncertainty. We subject the governing-status estimates in Table 4 to three corrections. Two-way clustering allows

errors to be correlated both within legislators over time and within meetings across legislators, so the result cannot be driven by a handful of heated sessions. The wild cluster bootstrap (1,999 Rademacher draws at the legislator level, null imposed) replaces the asymptotic approximation behind clustered standard errors with a distribution rebuilt from the data, guarding against unbalanced cluster sizes. Randomization inference sidesteps standard errors entirely; it reassigns the 190 legislators into placebo blocs of the observed sizes 1,000 times, recomputes the governing-status coefficient for each placebo assignment, and asks how often a placebo produces an effect as large as the observed one. Table 20 reports the results. The voice and composite estimates are unaffected by every correction, with all $p < 0.001$; no placebo assignment of legislators to blocs reproduces an effect of the observed size. The face estimate weakens, with $p$ between 0.03 and 0.09 across corrections, and should be read as suggestive, consistent with the face being the managed channel.

Table 20: Governing-Status Estimates under Inference Corrections

| | | Standard errors | | $p$-values | | |
|---|---|---|---|---|---|---|
| Outcome | $\hat{\delta}$ | Legislator | Two-way | Two-way | Wild bootstrap | Randomization |
| Voice anger ($z$) | $-0.345$ | (0.052) | (0.056) | $< 0.001$ | $< 0.001$ | $< 0.001$ |
| Face anger ($z$) | $-0.125$ | (0.058) | (0.073) | 0.084 | 0.030 | 0.086 |
| Composite ($z$) | $-0.235$ | (0.045) | (0.052) | $< 0.001$ | $< 0.001$ | $< 0.001$ |

*Notes:* Each row corresponds to a column of Table 4 (two-party sample; $N = 8{,}293$; 190 legislators; 135 meetings; legislator and meeting fixed effects). Two-way standard errors cluster on legislator and meeting. Wild cluster bootstrap $p$-values use 1,999 Rademacher draws at the legislator level with the null imposed. Randomization inference $p$-values reassign legislators to party blocs of the observed sizes 1,000 times and report the share of placebo assignments yielding $|\hat{\delta}_{\text{placebo}}| \geq |\hat{\delta}|$.

# Appendix M: The Role Reversal across All 48 Vocal Emotions

Table 5 reports the governing-status effect for seven theoretically chosen vocal emotions. Figure 6 widens the lens to every dimension the vocal model measures, estimating the same specification separately for each of the 48 vocal emotion $z$-scores. The estimates order themselves by adversarialness. The four largest negative effects are anger, contempt, disgust, and distress, the adversarial cluster identified in Appendix B, with anger the largest of all 48. The largest positive effect is calmness, followed by composed and positive registers such as love, satisfaction, and contentment. No clearly negative emotion rises under governing status. The high-arousal hostile ones fall, while fear, sadness, and guilt do not move, and neither does determination, so entering government does not make legislators less forceful or less engaged; it specifically retires the emotions whose expression constitutes an attack.

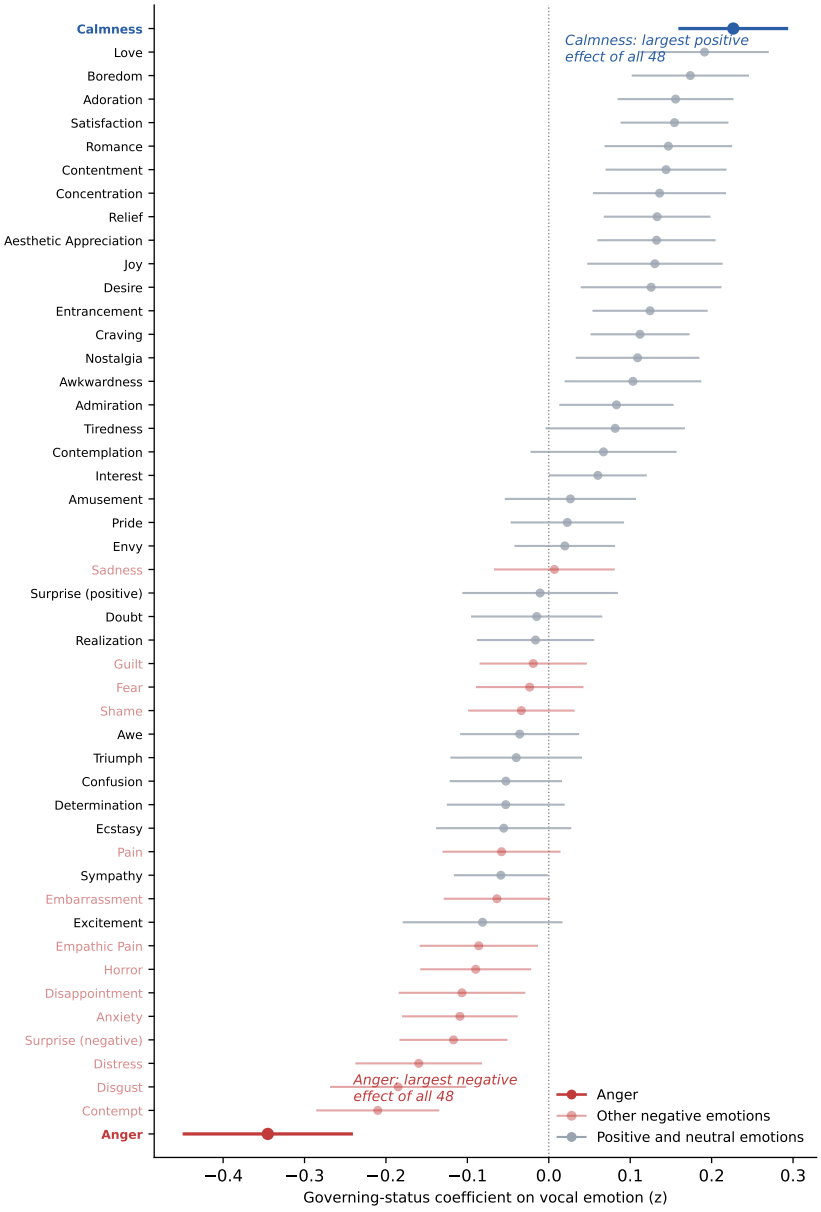


Figure 6: Governing-Status Effect on All 48 Vocal Emotion Dimensions. Each row is a separate regression of that emotion's vocal $z$-score on GOV_TV with legislator and meeting fixed effects, on the two-party sample ($N = 8{,}293$). Horizontal bars are 95% confidence intervals from legislator-clustered standard errors. Anger is shown in bold red and other clearly negative emotions in light red. No clearly negative emotion rises under governing status; every significant increase is a positive or neutral state.

# Appendix N: Robustness to Legislator × Meeting Interacted Fixed Effects

A natural concern is that unobserved legislator × meeting heterogeneity could confound the topic effects: if certain legislators are assigned to certain topics within meetings for reasons correlated with their baseline anger, the additive fixed effects specification may not fully absorb this variation. We address this by replacing the separate legislator and meeting fixed effects with their full interaction, legislator × meeting fixed effects, which identify topic effects strictly from within-cell variation—the same legislator discussing different topics in the same meeting. This is a more demanding specification: identification comes only from legislators who discuss multiple topic types within a single session, and the effective sample shrinks as singleton cells are absorbed.

Table 21 reports the results alongside the main additive specification for comparison.

Table 21: Topic Effects Under Additive vs. Interacted Fixed Effects

| | Voice anger (pct rank) | | Face anger (pct rank) | |
|---|---|---|---|---|
| | Additive FE | Interacted FE | Additive FE | Interacted FE |
| SCANDAL | $+0.138^{***}$ | $+0.112^{***}$ | $+0.020^{*}$ | $-0.004$ |
| | (0.013) | (0.013) | (0.011) | (0.009) |
| EVENT | $+0.107^{***}$ | $+0.090^{***}$ | $+0.012$ | $+0.000$ |
| | (0.014) | (0.013) | (0.012) | (0.009) |
| POLICY | $+0.072^{***}$ | $+0.056^{***}$ | $+0.028^{***}$ | $+0.016^{*}$ |
| | (0.013) | (0.013) | (0.010) | (0.009) |
| REFORM | $+0.103^{***}$ | $+0.080^{***}$ | $+0.026$ | $+0.025^{**}$ |
| | (0.019) | (0.020) | (0.016) | (0.013) |
| Fixed effects | Legislator + Meeting | Legislator × Meeting | Legislator + Meeting | Legislator × Meeting |
| *N* | 11,060 | 11,060 | 11,060 | 11,060 |

*Notes:* Each column is a separate regression. Additive FE columns re-estimate the specification of Table 1 on the percentile-rank scale (matching Table 7). Interacted FE columns replace separate legislator and meeting effects with their full interaction (904 legislator × meeting cells). Identification in the interacted specification comes from the same legislator discussing different topics within the same meeting. Procedural segments are the omitted category. Standard errors clustered at legislator level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The results are consistent with the controllability framework. Voice anger survives the interacted specification with modest attenuation: scandal drops from $+0.138$ to $+0.112$ (19% reduction), and the topic gradient is fully preserved (scandal > event > reform > policy). All voice anger coefficients remain highly significant ($p < 0.001$). Face anger, by contrast, is essentially eliminated for the high-anger topics under interacted FE: scandal drops from $+0.020^{*}$ to $-0.004$ (n.s.) and event drops from $+0.012$ to $+0.000$. This pattern is what one would expect if the face anger observed under additive FE partly reflects unobserved legislator-meeting heterogeneity that

the more demanding specification absorbs, while the voice signal—harder to control and therefore more driven by within-session topic variation—survives.

# Appendix O: Post-Peak Dynamics of the Two Channels

This appendix provides timing evidence for the peak–mean dissociation documented in Appendix G, where facial anger is high at the peak but flat in the segment mean. We ask how quickly each channel returns toward its resting level after an anger peak.

**Aligning two sampling rates.** The face model returns about three frames per second; the voice prosody model returns one score per spoken window (median 1.2 s). A naive comparison would make the denser face series look faster as an artifact. We resample both channels onto a common one-second grid within each segment and smooth both with the *same* kernel, matched to the voice channel's own window (2 s), so the two series carry identical effective resolution. Within each segment we detect local anger peaks (`find_peaks`; prominence at least half the within-segment standard deviation; height above the segment's 75th percentile), drop peaks within ten seconds of an edge, and measure each peak's *half-life*: the time for the normalized trajectory $(x_t - b)/(\text{peak} - b)$ to fall below 0.5, where the resting baseline $b$ is the segment's median anger after excluding all points within five seconds of any peak. The analysis uses the 2,729 segments at least forty seconds long with peaks in both channels (242 of 247 legislators).

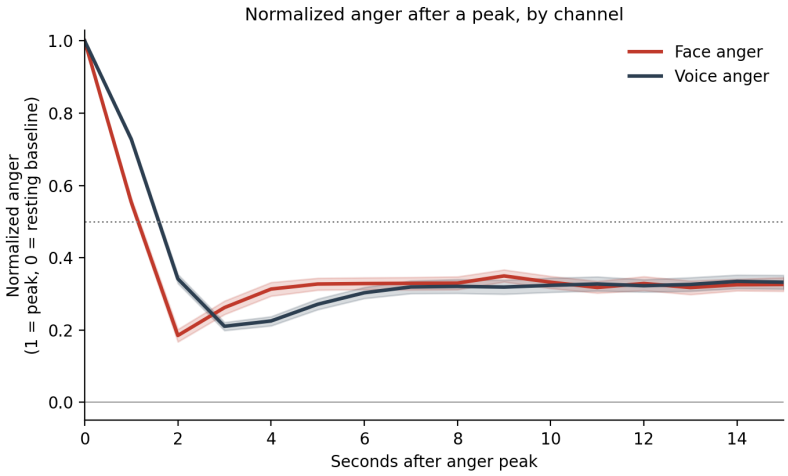


Figure 7: Normalized anger in the fifteen seconds after a peak, by channel (shaded bands: legislator-clustered 95% bootstrap intervals). The two channels settle at the same residual level; they differ in speed. Face anger completes its drop within about two seconds, roughly half a second before voice anger.

**The channels converge to the same level; they differ in speed.** Both channels settle at an indistinguishable long-run residual (Figure 7), so the difference is not where they end up but how fast they get there. Facial anger reaches half-peak in 2.1 s against 2.6 s for the voice, a gap of

$-0.43$ s $[-0.52, -0.33]$. Table 22 shows the gap is stable across four definitions of the resting baseline, across peak-height thresholds, and within each party bloc.

**The gap lives at the sub-two-second timescale.** The smoothing-kernel sweep in Panel C localizes the effect in time. The gap is largest without smoothing ($-0.88$ s), holds at the resolution matched to the voice window ($-0.43$ s at 2 s), attenuates as the kernel coarsens ($-0.17$ s at 3 s), and is gone by 5 s ($+0.15$, n.s.), the expected signature of a process that completes within about two seconds. A test that involves no kernel choice agrees: sampling the face series only at the voice model's own window midpoints gives $-0.69$ s $[-0.76, -0.60]$, so sampling density does not produce the difference.

Table 22: Post-peak half-life gap between channels (face − voice, seconds)

| | Gap (s) [95% CI] |
|---|---|
| *Panel A: resting-baseline definition (2 s kernel)* | |
| Segment median, peaks excluded $\pm 5$ s (primary) | $-0.43$ $[-0.52, -0.33]$ |
| Pre-peak window mean | $-0.45$ $[-0.54, -0.36]$ |
| Legislator's procedural resting level | $-0.29$ $[-0.44, -0.15]$ |
| Segment median (peaks included) | $-0.54$ $[-0.60, -0.49]$ |
| *Panel B: threshold, party, and voice-native grid* | |
| Peak height $\geq$ segment p90 | $-0.47$ $[-0.57, -0.38]$ |
| Conservative bloc only | $-0.47$ $[-0.61, -0.31]$ |
| Democratic bloc only | $-0.43$ $[-0.54, -0.32]$ |
| Face sampled on voice's window grid | $-0.69$ $[-0.76, -0.60]$ |
| *Panel C: smoothing kernel (primary baseline)* | |
| 1 s (unsmoothed) | $-0.88$ $[-0.94, -0.81]$ |
| 2 s (matched to voice window; primary) | $-0.43$ $[-0.52, -0.33]$ |
| 3 s | $-0.17$ $[-0.28, -0.05]$ |
| 5 s (coarser than the process) | $+0.15$ $[-0.04, +0.34]$ |

*Notes:* Half-life is the time for a peak's normalized anger to fall below half its peak-to-baseline height; the baseline is peak-free unless stated. Gaps are differences in legislator-mean half-lives; all 95% intervals come from a single legislator-clustered bootstrap (1,500 draws, common seed), so identical specifications give identical intervals. "Voice-native grid" samples the face series at the voice model's window midpoints before gridding. Panel C localizes the effect: present at resolutions at or finer than the voice window, attenuating monotonically under coarser smoothing.

**The peaks are signal, and their timing does not depend on topic.** Facial *peak* anger rises on every substantive topic relative to procedural business (Appendix G); pure jitter would carry no such topic structure, so the facial peaks analyzed here are expressions, not noise. Their return speed is topic-indifferent: the scandal-minus-other-substantive difference in half-life is small and not significant in either channel (face $-0.10$ s $[-0.29, +0.07]$; voice $-0.08$ s $[-0.27, +0.11]$). Peak

*frequency* is likewise flat across topics: peaks per minute on scandal / other substantive / procedural business are 5.5 / 5.6 / 5.5 for the face and 3.9 / 3.8 / 4.2 for the voice, with no significant contrast in either channel. With both the tempo and the rate of excursions flat across topics, topic structure can enter either channel only through *amplitude*, which is what the peak regressions of Appendix G show: vocal peak amplitude preserves the topic gradient while facial peak amplitude does not, and it is the voice's higher, slower-fading excursions that accumulate into the segment mean.

**Interpretation.** The face's excursions are briefer than the voice's. A facial anger display completes its round trip within about two seconds, with symmetric onset and offset (2.11 vs 2.12 s, $p = 0.51$), while the voice is slower on both ends and fades more slowly than it builds (2.59 vs 2.43 s, $p < 0.001$). Both channels return to the same residual level, so these are differences in tempo, not in where the channels settle, and they are consistent with the peak–mean dissociation: facial anger confined to brief flashes contributes little to a segment-long average even when its peaks are pronounced.[17]

[17]The timing gap alone does not account for the full size of the peak–mean dissociation; the mean-level result is established directly in Appendix G. Because the analysis requires long segments, it rests on a subset skewed toward longer, more procedural speeches (mean 154 s vs 50 s in the full sample; procedural 38% vs 25%); the topic- and party-indifference of the gap indicates the pattern is not specific to that subset.

# Appendix References


Baevski, A., Zhou, Y., Mohamed, A., and Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 12449–12460.

Baltrušaitis, T., Zadeh, A., Lim, Y. C., and Morency, L.-P. (2018). OpenFace 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 59–66. IEEE.

Cheong, J. H., Jolly, E., Xie, T., Byrne, S., Kenney, M., and Chang, L. J. (2023). Py-feat: Python facial expression analysis toolbox. *Affective Science*, 4(4):781–796.

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., and Truong, K. P. (2016). The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing*, 7(2):190–202.

Jeon, D., Lee, J., and Kim, C. (2022). KOTE: Korean online that-gul emotions dataset. Please verify author list and details against the source.

Krippendorff, K. (2011). Computing Krippendorff's alpha-reliability. Departmental paper, University of Pennsylvania, Annenberg School for Communication. Available at `https://repository.upenn.edu/asc_papers/43`.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Yang, S.-w., Chi, P.-H., Chuang, Y.-S., et al. (2021). SUPERB: Speech processing universal performance benchmark. In *Proceedings of Interspeech*, pages 1194–1198.