Replication 2 Report

Paper

OSNABRÜGGE M, HOBOLT SB, RODON T. Playing to the Gallery: Emotive Rhetoric in Parliaments. American Political Science Review. 2021;115(3):885-899. doi:10.1017/S0003055421000356

Introduction

This paper investigates how legislators strategically deploy emotional rhetoric within the UK Parliament. While emotional appeals are well-documented in campaigns and public speeches, their role in formal legislative settings remains underexplored. The central research question asks: when and why do legislators use emotive language in parliament?

The authors hypothesize that legislators use more emotive rhetoric in high-profile settings—where public visibility and incentives to perform are strongest—compared to the routine, low-profile debates. Drawing on a dataset of nearly one million UK parliamentary speeches (2001–2019), they construct a dictionary-based measure of emotive rhetoric. The approach combines the ANEW lexicon with word embeddings trained on parliamentary corpora, allowing the authors to expand the vocabulary using semantic similarity and score speeches based on the proportion of emotive versus neutral language.

This pattern supports the idea that legislators "play to the gallery"—crafting their language not only for policy discussion, but to appeal to external audiences. Topic-level analysis further shows that emotionally charged language is most common in debates over values and society, and less so in technical domains like the economy.

Differences and Similarities

The replication successfully reproduces the major findings and visualizations presented from the article. Across Figures 2 and 4, as well as regression models in Tables 3 and 4[1], we observe results that align closely with the original in both direction and statistical significance. Specifically, we confirm that emotive rhetoric was more prevalent in high-profile parliamentary settings than in general debates. Likewise, topic-level patterns reveal that emotive language tends to cluster around socially and morally charged policy domains.

We extended the analysis by constructing a new emotive dictionary using the NRC Emotion Lexicon as a foundation. Following a similar embedding-based methodology, we generated an

[1] Figure and table references correspond to those in the original article: Osnabrügge, Hobolt, & Rodon (2021). Replicated versions of Figure 2 and 4 are presented in Appendix Figure 1 and 2, respectively,

[2]alternative lexicon tailored to UK parliamentary language. This allowed us to replicate the scoring procedure and evaluate the robustness of the original approaches.

Our results diverge in several key areas. “Urgent Questions” displayed greater emotive variability and higher peaks than reported in the original paper, suggesting this unscheduled and event-driven debate format may contain more spontaneous or emotionally charged exchanges. Topic rankings also shifted: themes like *Freedom and Democracy* appeared more emotionally intense under the new dictionary, while domains such as social welfare scored lower.

These discrepancies highlight how lexicon construction influences measurement and interpretation. The NRC-based dictionary may emphasize affective tone in abstract or symbolic language while underweighting institutional or technical expressions. This underscores the importance of dictionary choice in computational text analysis and invites further investigation of how emotional rhetoric is operationalized.

Autopsy

The replication process is largely successful in reproducing the core empirical findings of the original study, and several components of the workflow proved to be both robust and adaptable. The availability of a well-structured dataset, clearly labeled variables, and documentation on scoring procedures facilitate efficient reproduction and transparent comparisons.

Several elements work particularly well. The alignment of speech texts with speaker metadata enables smooth integration of contextual variables. The implementation of the dictionary-based scoring method using word embeddings, particularly cosine similarity between seed words and candidate terms, translated effectively from the original methodology. Visualization and modeling steps—such as replicating Figures 2 and 4[2]—are easily transferable and flexible enough to incorporate our own trained dictionary extension.

However, several challenges also emerge. Some R packages used in the original workflow were deprecated or version-sensitive, requiring additional troubleshooting for compatibility. Memory limitations also arise when working with when processing large-scale word embeddings or training supervised models, especially with different dictionary thresholds and classification models. This necessitates the use of subsampling and simplified model configurations.

Finally, while the original word embedding approach worked well for lexicon expansion, its reliance on a small set of seed terms and static embeddings constrained linguistic complexity. Our replication, with NRC-based seeds, confirmed that even small adjustments in seed lexicons can lead to notable shifts in outcomes—highlighting the sensitivity of results to dictionary construction choices.

---

[2] Extended versions produced using the NRC-based dictionary are included in Appendix Figure 2 and Appendix Figure 4, respectively.

Extension

To deepen the study, several methodological improvements could be proposed. First, incorporating contextual language models—such as BERT or other transformer-based embeddings—would allow for a more detailed detection of emotive tone by considering how meaning shifts based on syntactic and semantic context. This would overcome limitations of static word-level scoring, particularly in cases involving irony, negation, or rhetorical contrast.

Second, applying the framework to additional political environments—such as the U.S. Congress, European Parliament, or devolved legislatures like the Scottish Parliament—as well as to key historical periods (e.g., post-Brexit, COVID-19) would test the generalizability of observed patterns and clarify whether emotive rhetoric is context-dependent or structurally embedded across institutions.

Finally, linking emotive rhetoric to downstream outcomes, such as media coverage, public engagement, or legislative effectiveness, would move the research beyond descriptive measurement,a more causal understanding of the role emotional appeals play in shaping political communication and influence.

Appendix

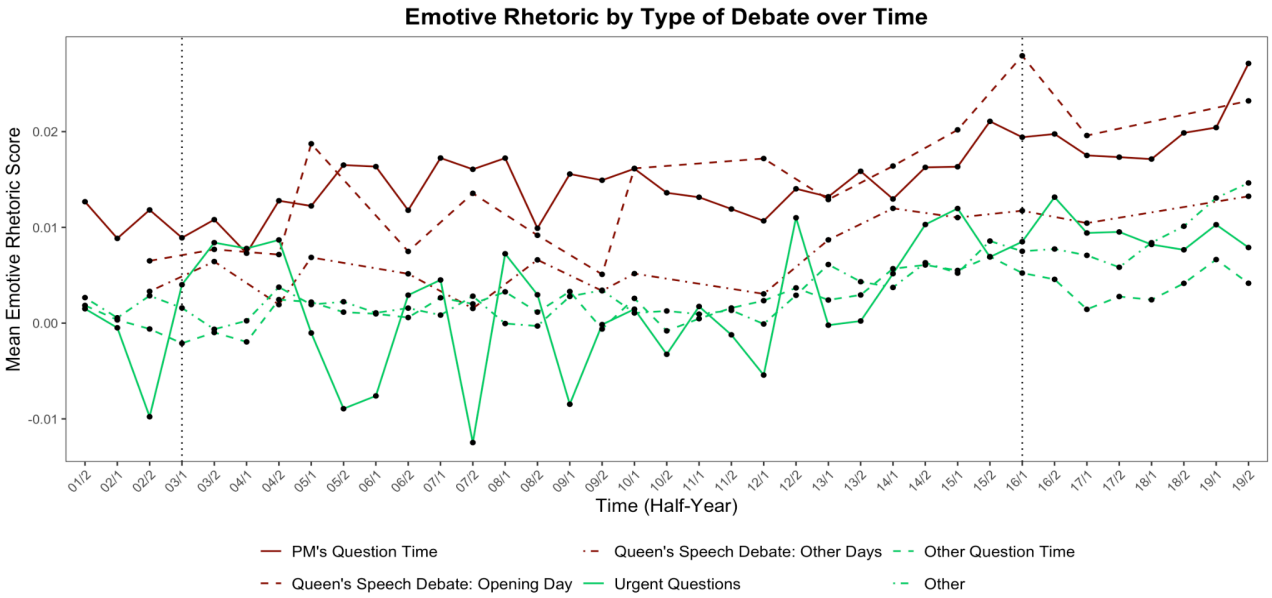

Figure 1. Emotive Rhetoric by Type of Debate (Original Dictionary)

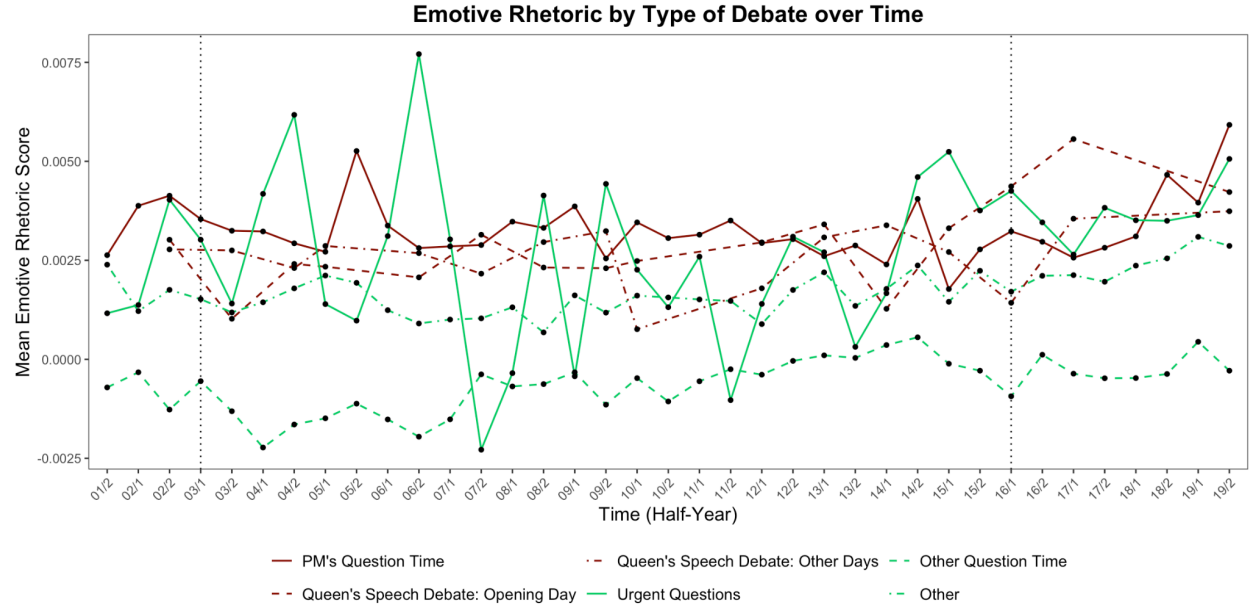

Figure 2. Emotive Rhetoric by Type of Debate (Extended Dictionary)

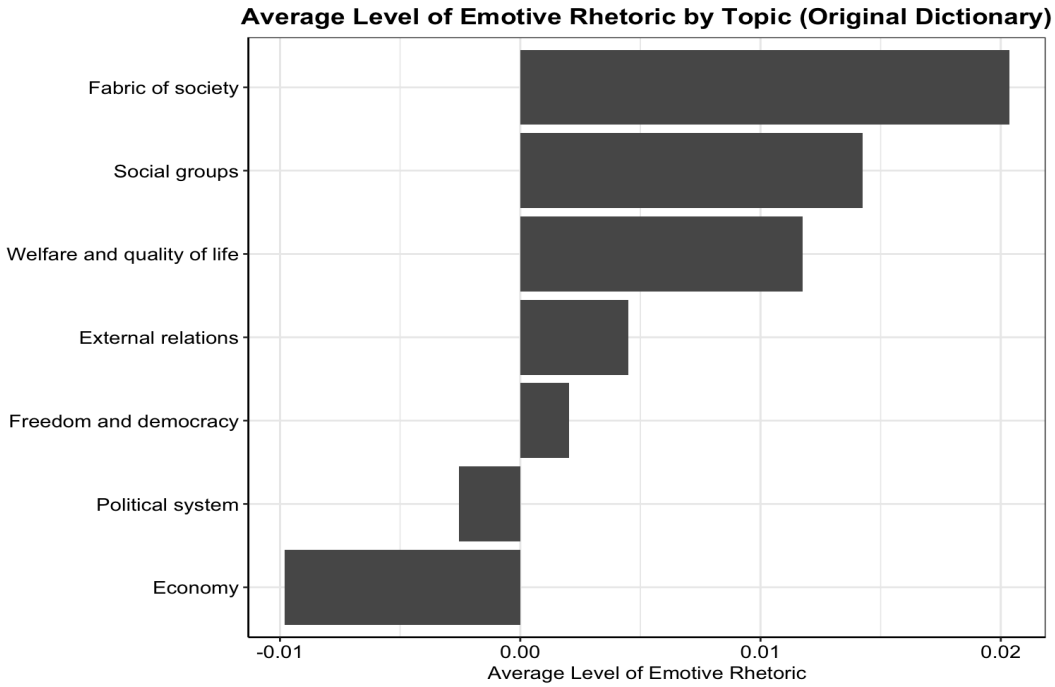

Figure 3. Average Level of Emotive Rhetoric by Topic (Original Dictionary)

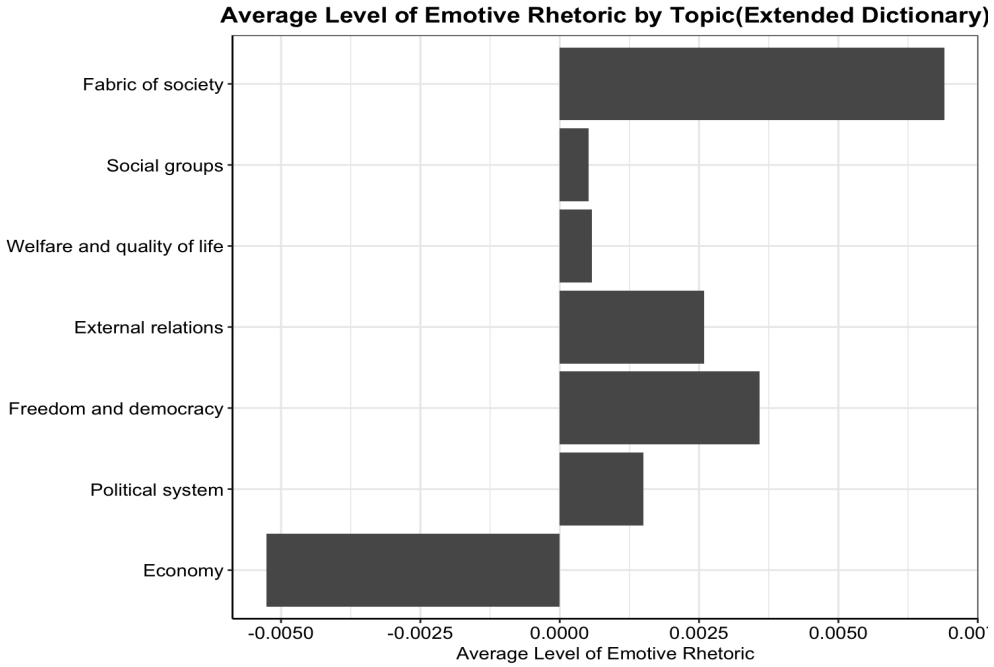

Figure 4. Average Level of Emotive Rhetoric by Topic (Extended Dictionary)