Persuasiveness of News Editorials depending on Ideology and Personality

Roxanne El Baff [1,2] **Khalid Al-Khatib** [2] **Benno Stein** [2] **Henning Wachsmuth** [3]
[1] German Aerospace Center (DLR), Germany, roxanne.elbaff@dlr.de
[2] Bauhaus-Universität Weimar, Weimar, Germany, <first>.<last>@uni-weimar.de
[3] Paderborn University, Paderborn, Germany, henningw@upb.de

Abstract

News editorials aim to shape the opinions of their readership and the general public on timely controversial issues. The impact of an editorial on the reader's opinion does not only depend on its content and style, but also on the reader's profile. Previous work has studied the effect of editorial style depending on general political ideologies (liberals vs. conservatives). In our work, we dig deeper into the persuasiveness of both content and style, exploring the role of the intensity of an ideology (lean vs. extreme) and the reader's personality traits (agreeableness, conscientiousness, extraversion, neuroticism, and openness). Concretely, we train content- and style-based models on New York Times editorials for different ideology- and personality-specific groups. Our results suggest that particularly readers with extreme ideology and non "role model" personalities are impacted by style. We further analyze the importance of various text features with respect to the editorials' impact, the readers' profile, and the editorials' geographical scope.

1 Introduction

News editorials are considered the backbone of a community in which they tackle timely controversial issues, aiming to sway readers towards certain opinions. Nowadays, editorials do not only focus on issues affecting their close entourage (e.g., within a state or a country), but rather tackle issues relevant across continents to shape the views of those living there and worldwide. For example, *The New York Times* and *Der Spiegel*[1] lately invested resources to write news editorials about the August 4th Beirut blast. As such, news editorials represent an important source for research on computational social science.

To be persuasive, a news editorial should comply with the communication-persuasion paradigm defined by O'Keefe (2015) consisting of five factors: source, message, target, impact, and channel: (1) The *source* represents an editorial's author who tries to persuade the readers. Usually, authors reflect the ideology of their newspaper. For example, The New York Times is considered a liberal news portal and its editorials reflect this ideology. (2) The *message* represents an editorial's content and the linguistic choices it made, e.g., in terms of style. (3) The *target* represents the readers, their prior beliefs (e.g., their political ideology or its intensity) and their characteristics (e.g., personality traits or gender). (4) The *impact* represents the actual effect of an editorial on a reader. Halmari and Virtanen (2005) states that persuasive text aims at changing or affecting the behavior of others or at strengthening the existing beliefs of those who already agree. And (5) the *channel* represents the mean used to read the editorial, e.g., an online news portal. We leave an analysis of the impact of the medium used on the editorial's effectiveness to future research.

Previous research tackled how people are affected by arguments depending on their personality traits, interests, and beliefs. However, most studies conducted their analysis on dialogical text from debate portals and similar (Lukin et al., 2017; Durmus and Cardie, 2018; Al Khatib et al., 2020). For news editorials, we recently revealed that liberal readers, unlike conservatives, are affected by the linguistic style (El Baff et al., 2020). Still, it remains unexplored to what extent also the intensity of a political ideology plays a role, let alone a reader's personality traits. In our work here, we fill this gap, and we consider both the content and the style of an editorial.

[1] A German newspaper: https://www.spiegel.de/international/

In particular, this paper analyzes the persuasive effect (the impact) of linguistic content and style choices of news editorials (the message) on readers (the target) with two profile varieties, the intensity of their political ideology and their personality traits. We distinguish *lean* and *extreme* intensity of ideology, and we resort to the "Big Five" personality traits (Goldberg, 1990): *agreeableness*, *conscientiousness*, *extraversion*, *neuroticism*, and *openness*.

For our analysis, we employ a corpus with 1000 New York Times news editorials (El Baff et al., 2018). Each editorial is annotated for a notion of impact that defines the editorial to either *challenge* its readers' stance, by making them rethink their current opinion towards a topic, to *reinforce* their stance, by helping them argue better about a certain issue, or neither. The annotations were added by 24 readers with different political ideologies (liberals and conservatives). For each reader, also the ideology intensity and the Big Five personality traits are reported, but this information has not been used so far to our knowledge.

For each intensity and personality group in the corpus, we train one model to predict the persuasive effectiveness of an editorial, using various content and style features. Our results show that people with extreme ideology are somewhat impacted by style, and the same holds for readers whose personality is relatively high in neuroticism and low in extraversion. We further investigate the role of editorials' geographical scope; whether it tackles a *global*, *national*, or *state* topics.

2 Related Work

News editorials reflect argumentation related to political issues and, therefore, comprise hidden rhetorical means (van Dijk, 1995), which makes them a challenging genre to study. Some works dealt with news editorials for information retrieval purposes (Yu and Hatzivassiloglou, 2003; Bal, 2009) or for analyzing arguments (Bal and Dizier, 2010; Kiesel et al., 2015; Scheffler and Stede, 2016). Al Khatib et al. (2016) represent editorial argumentation explicitly by annotating 300 news editorials with argumentative discourse units on the sub-sentence level. We employ their model to extract features from news editorials for predicting the persuasive effectiveness of news editorials, as detailed in Section 4.

Aristotle (2007) argued that a persuasive effect is best achieved by providing showing a good character (ethos), evoking the right emotions (pathos), and providing logically reasoned arguments (logos) in a well-arranged and well-phrased way. This view was modeled by Wachsmuth et al. (2018) for argumentation synthesis. Instead, we here follow the communication-persuasion paradigm of O'Keefe (2015), stating that an argumentative text, and hence a news editorial, should comply with five factors to be persuasive, as already indicated in Section 1. Each of them is tackled in some way in related work:

(1) *Source* refers to the prior beliefs and behaviors of the writer. Each news portal reflects its beliefs (van Dijk, 1995). (2) *Message* deals with the linguistic choices in the content. In this regard, Hidey et al. (2017) study the semantic types of argument components in an online forum, and El Baff et al. (2020) analyze the persuasive effect of linguistic style on readers. Also, Hidey and McKeown (2018) and Durmus et al. (2020), respectively, exploit the role of argument sequencing in detecting persuasive influence, and the role of pragmatics and discourse context in determining argument impact. (3) *Target* includes the prior beliefs of readers. Lukin et al. (2017) find that emotional and rational arguments are effective depending on the Big Five personality traits (John et al., 1991). Also, Durmus and Cardie (2018) provide a debate portal dataset with a controlled task setting that takes into consideration the reader's religious and political ideology, and Al Khatib et al. (2020) exploit the personal characteristics of debaters to improve persuasiveness prediction. (4) *Impact* reflects the effect of a text, which has been assessed for essays (Persing and Ng, 2015; Wachsmuth et al., 2016) and debate portal arguments (Habernal and Gurevych, 2016; Persing and Ng, 2017). (5) *Channel*, finally, means the communication medium. Joiner and Jones (2003) study the effect of the medium on argumentation. They found that the quality of argumentation in face-to-face discussions is higher than in online discussions.

In previous work, we annotated news editorials at the document level, covering an editorial's persuasive effectiveness by reflecting to what extent a writer persuades a reader (El Baff et al., 2018); the effectiveness concept is based on the argumentation quality taxonomy defined by Wachsmuth et al. (2017). In our annotation setup, we considered the beliefs of readers by profiling annotators based on political ideology (liberals or conservatives). We also provides additional information about the annotators' ideology

(a) Intensity	Lean		Extreme	
Effect	**Train**	**Test**	**Train**	**Test**
Challenging	100	21	156	43
Ineffective	274	70	133	30
Reinforcing	409	105	494	123
Overall	783	196	783	196

(b) Personality	Role Models		Other	
Effect	**Train**	**Test**	**Train**	**Test**
Challenging	74	15	106	26
Ineffective	121	31	412	108
Reinforcing	588	150	265	62
Overall	783	196	783	196

Table 1: Distribution of the majority persuasive effect of the news editorials in the given training and test set for (a) ideology intensity, i.e., *lean* or *extreme*, and (b) the personality group, i.e., *role model* or *other*.

Figure 1: The distribution of the 24 selected annotators over the eight considered political ideologies, which we grouped by ideology intensity into *lean* and *extreme*.

intensity and personality traits. Later, we analyzed linguistic choices in news editorials with respect to the readers' ideology and reported effect (El Baff et al., 2020). However, we did not conduct our analysis on how content and style affect readers with different ideology intensity or with different personality traits. In the paper at hand, we fill this gap by using a similar methodology to detect the effect of style and content on readers with different ideology intensity and personality traits.

Conceptually, the studies of Lukin et al. (2017), Durmus and Cardie (2018), and Al Khatib et al. (2020) are closest to ours, but they tackle single arguments and dialogical argumentation respectively. To our knowledge, there is no computational analysis of linguistic choices related to ideology intensity and personality of readers and writers so far.

3 Data

The analysis is conducted using the Webis-Editorial-Quality-18 corpus (El Baff et al., 2018). 1000 New York Times editorials were annotated regarding their persuasive effects by three liberals and three conservatives each. The persuasive effect of each editorial was determined based on whether the editorial *challenged* their stance by making them rethink it, *reinforced* their stance by helping them argue better, or was *ineffective*. We previously utilized a corpus to investigate the role of editorial's style on readers with different political ideologies (El Baff et al., 2020). In order to ease the comparison to El Baff et al. (2020), we use the same dataset (with similar training/test split) in all our experiments. In particular, we chronologically split the dataset into 80% for training and 20% for testing (Table 1), based on editorials issue date, to imitate real-life prediction.[2] The majority effect votes vary depending on the annotators' profile (e.g., ideology intensity such as *lean*, personality trait such as *low agreeableness*).[3]

The corpus, in addition to the effect labels, includes information about the annotator's ideology and personality traits. In the following, we describe how we leverage this information here.

3.1 Ideology Intensity

The annotators of the Webis-Editorial-Quality-18 corpus took the PEW political typology quiz in order to determine their political ideology.[4] The ideology classes in this test ranges from *Solid Liberals* to *Core Conservatives*, as shown in Figure 1.

[2]In our previous work (El Baff et al., 2020), we found 21 duplicate editorials with the same content but different IDs — for these, they use the majority vote across all duplicates).

[3]In case of a tie between *effective* (challenging or reinforcing) and ineffective, we consider the majority effect as *effective*.

[4]PEW research quiz: `https://www.pewresearch.org/politics/quiz/political-typology/`

Trait	Effect	Low Train	Low Test	Average Train	Average Test	High Train	High Test
Agree-ableness	Challenging	123	30	n/a	n/a	86	17
	Ineffective	291	76	n/a	n/a	162	43
	Reinforcing	369	90	n/a	n/a	535	136
Conscien-tiousness	Challenging	157	48	77	12	64	9
	Ineffective	115	18	263	78	118	24
	Reinforcing	511	130	443	106	408	114
Extra-version	Challenging	106	22	83	18	171	53
	Ineffective	281	80	126	21	119	25
	Reinforcing	396	94	574	157	493	118

Trait	Effect	Low Train	Low Test	Average Train	Average Test	High Train	High Test
Neuro-ticism	Challenging	181	39	80	23	95	15
	Ineffective	104	36	114	28	311	70
	Reinforcing	498	121	396	96	377	111
Open-ness	Challenging	133	29	125	31	148	35
	Ineffective	316	93	220	42	106	28
	Reinforcing	334	74	438	123	529	133

Table 2: Distribution of the majority persuasive effect of the news editorials in the given training and test sets for the three values of the Big Five personality traits: *low*, *average*, and *high*. For agreeableness, *average* was combined with *low* due to the low number of annotators (two only) with *average* values.

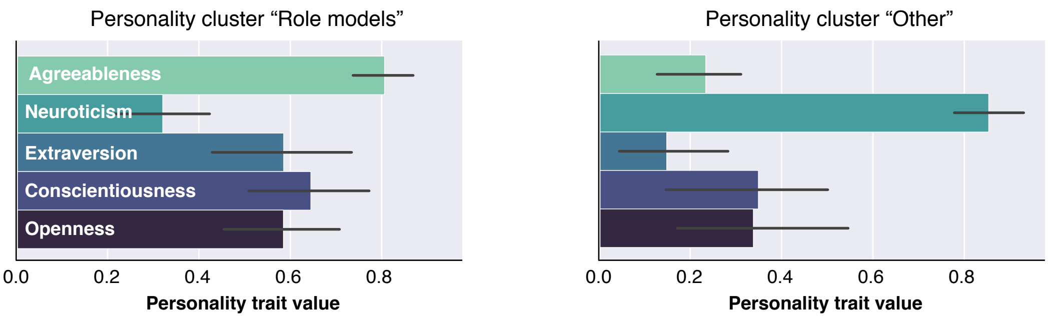

Figure 2: The two personality clusters, *role models* and *other*, based on the Big Five personality traits: agreeableness, conscientiousness, extraversion, neuroticism, and openness.

To analyze the effect of editorials on readers with different ideologies, we abstracted the annotators' ideologies into *Liberals* and *Conservatives* before (El Baff et al., 2018; El Baff et al., 2020). In contrast, we here decide to focus on the ideology intensities. Hence, we group the annotators into *lean* ("Market Skeptic Republicans", "New Era Enterprisers" and "Disaffected Democrats") and *extreme* ("Country First Conservatives", "Opportunity Democrats", "Solid Liberals"), illustrated in Figure 1. Table 1(a) shows the distribution of the persuasive effect (aggregated by majority vote) of the news editorials in the training and test sets for extreme and lean intensities.

3.2 Personality

Personality Traits Besides the ideology test, the annotators took the personality test based on the "Big Five" (Goldberg, 1990) traits. Each annotator was assigned a numerical score (between 0 and 100) for each of five traits: "Agreeableness", "Conscientiousness", "Extraversion", "Neuroticism", and "Openness". Using this information, we investigate the impact of personality traits as follows. We use the same training and test sets mentioned before for each personality trait value of *Low* (≤ 32), *Average* (≥ 33 and ≤ 67), and *High*.[5] Table 2 shows the training and test distribution for each trait value (e.g. Openness *low*) across all effects (challenging, ineffective, and reinforcing).

Personality Groups We categorize the annotators into two personality clusters. To do that, we apply cosine k-means, with $k = 2$, on the annotators' five personality trait values, as shown in Figure 2. The first group contains annotators with relatively high agreeableness, conscientiousness, extraversion, and openness, whereas the second group contains annotators with high neuroticism. Due to the small size of the dataset, we use $k = 2$ only.

[5]The *Low*, *Average* and *High* ranges were defined already in previous work (El Baff et al., 2018). There is one exception: for conscientiousness, the average range is ≥ 33 and < 67.

Feature Base	Overview	Reference	Label
Linguistic inquiry and word count	Psychological meaningfulness in percentile	Pennebaker et al. (2015)	liwc
NRC emotional and sentiment lexicon	Count of emotions (e,g. *fear, etc.) and polarity words*	Mohammad and Turney (2013)	emotion
Webis Argumentative Discourse Units	Count of each evidence type (*anecdote* and *testimony*)	Al Khatib et al. (2017)	evidence
MPQA Arguing Lexicon	Count of 17 types of arguing (*assessments*, *doubt*, etc.)	Somasundaran et al. (2007)	arguing
MPQA Subjectivity Classifier	Count of subjective and objective sentences	Riloff and Wiebe (2003)	subjectivity
Lemma 1–3-grams	TfIdf of lemma for 1–3-grams	Miller (1998)	lemma

Table 3: Summary of the six feature types used. Each feature is quantified at both the level of the editorial. The labels (rightmost column) are used to refer to the respective feature.

Gerlach et al. (2018) developed an approach to identify personality types, which they applied to more than 1.5 million participants. They found robust evidence for at least four distinct personality types and one of them is labeled as the "role model", who is low in *neuroticism* and high in all the other traits. Figure 2 shows that the upper cluster fits the description of the "role model". For simplicity, we refer to the two personality groups by *role models* and *other* reflecting the most discriminating personality trait between the two groups. Table 1(b) shows the distribution of the persuasive effect (majority vote) of the news editorials in the training and test sets for the two personality groups.

4 Features

In this section, we describe the set of features that we select to explore the linguistic choices in editorials. These features encode semantic and pragmatic properties that may manifest the author's means of persuasion, implicitly or explicitly, and follow those in previous work (El Baff et al., 2020): psychological meaningfulness, eight basic emotions, editorial evidence types, argumentativeness, and subjectivity. As those features essentially target the modeling of text style, we also consider standard text features to model text content. In the following, we describe all features in detail (an overview is given in Table 3):

Linguistic Inquiry and Word Count (liwc) LIWC (Pennebaker et al., 2015) is a lexicon-based text analysis that counts words in psychologically meaningful categories (Tausczik and Pennebaker, 2010). It captures the narrative tone, the emotional tone, and the confidence tone among several other categories.

NRC Emotional and Sentiment Lexicons (emotion) The NRC lexicon, compiled with crowdsourcing by Mohammad and Turney (2013), contains a set of English words and their associations with (1) *sentiment*, i.e., negative and positive polarities, and (2) *emotions*, i.e., the eight basic emotions as defined by Plutchik (1980): *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise* and *trust*. We use this lexicon to generate features, where each category is represented as the count of words in an editorial (e.g., *sad* words).

Webis Argumentative Discourse Units (evidence) Al Khatib et al. (2017) developed a computational model to classify the evidence types in news editorials. The model was trained and evaluated using the corpus of Al Khatib et al. (2016), which contains 300 editorials from The Guardian, Al Jazeera, and Fox News. Each segment in the editorial is labeled with six types, including three evidence types: (1) *anecdote*, giving a personal experience of the author, (2) *statistics* citing a quantitative study, and (3) *testimony* quoting an expert's argument. The classifier sees all remaining types (common ground, assumption, and other) as (4) *other*. We apply the evidence classifier at the sentence-level of each editorial and count the occurrence of each type (e.g., number of *testimony* sentences in an editorial).

MPQA Arguing Lexicon (arguing) The MPQA Arguing lexicon, built by Somasundaran et al. (2007), includes various arguing patterns of different types such as *causation*, *conditionals*, *structure*, and *contrast*. Using the lexicon, we generate different features represented as the count of each arguing type in a text (e.g., number of *assessments* patterns in an editorial).

Features	A. Intensity		B. Personality	
	Extreme	Lean	Role Model	Other
liwc	0.28	0.29	0.32	0.29
emotion	0.34	0.32	0.31	0.30
evidence	0.34	0.28	0.24	0.36
arguing	0.29	0.34	0.30	0.29
subjectivity	0.23	0.21	0.29	0.37
Top Style	***0.40**	0.35	0.37	0.37
Content (lemma-based)	0.33	0.37	0.33	0.34
Top Content+Style	*0.38	**0.42**	**0.36**	***0.39**
Random baseline	0.27	0.32	0.23	0.34

Table 4: The macro F_1-score of each feature type and the best combinations in classifying the persuasive effect on readers with different profiles: *extreme* and *lean* ideology intensity (left), as well as *role models* and *other* personality group (right). $*$ indicates significant gains over the *Random baseline* at $p < 0.05$.

MPQA Subjectivity (subjectivity) The MPQA subjectivity classifier, provided in OpinionFinder 2.0 (Riloff and Wiebe, 2003; Wiebe and Riloff, 2005), labels a text as *subjective* or *objective*. We apply the classifier to the editorials, and count the number of *subjective* and *objective* sentences.

Content Features (lemma) We use the Tf-Idf score for lemma (Miller, 1998) 1–3-grams as the base of our content features.

5 Analysis of the Persuasive Effect

In this section, we assess the impact of the style and content of news editorials on their persuasive effectiveness for readers with different ideology intensities (extreme or lean), personality traits, and personality groups (role models or others). Similar to El Baff et al. (2020), we perform the analysis by approaching the following task: *Given a news editorial and a reader's profile characteristic (ideology intensity, personality trait, or personality group), predict the effect of the editorial.* This task is tackled by developing a separate effect prediction model for each ideology intensity (extreme or lean), each personality trait (e.g., low agreeableness), and each personality group (role models or others).

The prediction models use SVM classifiers (with a linear kernel), in which each classifier is trained using its corresponding profile training set and evaluated on the test set (See Section 3). The classifiers employ the features described in Section 4, considering both the style and the content of the editorials. The SVM cost was tuned using grid search with 5-fold cross-validation on the training set. We set the class weight to "balance" because of the skewed distribution of the data, as shown in Tables 1 and 2. The prediction results are reported using the macro-F_1 scores for each style feature alone, for the best combination of style features (*top style*), for the best combination of style and content (*top content+style*), and the *random baseline*. We measure significance using a t-test (Wilcoxon's test if normal distribution is missing) to quantify the differences between each two feature-based models among random baseline', content, top style, and top style+content.

In the following presentation of the results, we see readers as impacted by style and/or content if at least one model based on the respective feature manages to outperform the random baseline significantly.

5.1 Ideology Intensity

As shown in Table 4.A, for *extreme* intensity ideologies, the only two models that significantly beat the random baseline are *top style* (liwc, emotion, arguing) with macro-F_1 = 0.40 and *top content+style* (lemma, arguing, evidence) with macro-F_1 = 0.38. The *content* model alone did not significantly outperform the baseline. For the *lean* ideology, we did not observe any model that yield significant improvements.

5.2 Personality

Traits Table 5 shows the macro-F_1 scores for each personality trait value. In general, the best combination of style and content, *top content+style*, performed best. In detail, we observe the following:

Features	Agreeable. Low	Agreeable. High	Conscientiousness Low	Conscientiousness Avg	Conscientiousness High	Extraversion Low	Extraversion Avg	Extraversion High	Neuroticism Low	Neuroticism Avg	Neuroticism High	Openness Low	Openness Avg	Openness High
liwc	0.35	0.31	0.39	0.35	0.33	0.35	0.27	0.33	0.35	0.28	0.30	0.26	0.30	0.28
emotion	0.30	0.31	0.29	0.34	0.34	0.35	0.26	0.30	0.28	0.23	0.35	0.29	0.35	0.27
evidence	0.37	0.27	0.25	0.28	0.29	0.31	0.21	0.29	0.28	0.27	0.28	0.32	0.31	0.28
arguing	0.29	0.28	0.28	0.28	0.29	0.31	0.26	0.28	0.26	0.18	0.23	0.32	0.22	0.25
subjectivity	0.29	0.28	0.27	0.26	0.30	0.24	0.16	0.23	0.27	0.35	0.25	0.19	0.24	0.31
Top Style	0.39	0.32	0.39	0.36	*0.42	0.40	*0.33	*0.35	0.38	0.35	0.35	0.34	0.37	0.35
Content (lemma-based)	0.34	0.32	*0.40	0.39	0.38	0.35	*0.38	*0.35	0.35	**0.42**	0.33	‡0.39	**0.40**	0.39
Top Content+Style	†**0.41**	**0.37**	***0.44**	$^{\ddagger *}$**0.43**	***0.43**	**0.41**	***0.42**	**0.37**	**0.38**	0.41	**0.42**	‡**0.41**	**0.40**	†**0.43**
Random baseline	0.31	0.26	0.25	0.34	0.21	0.29	0.25	0.30	0.29	0.35	0.23	0.33	0.29	0.26

Table 5: The macro F_1-scores of each feature type and their best combinations in classifying the persuasive effect on readers with different profiles. * and † and ‡ indicate significant differences at $p < 0.05$ against the *Random baseline*, *content* and *style* respectively.

- *Agreeableness.* For the readers with *low* agreeableness, the top content+style model (liwc, evidence and lemma) model was significantly better than the content model. In contrast, for the readers with *high* agreeableness, no model significantly outperformed the baseline.
- *Conscientiousness.* Readers with *low* and *average* values seem to be impacted by content. However, readers with *high* conscientiousness are more impacted by style, i.e., both top style (liwc, emotion, arguing) and top content+style (lemma, liwc, emotion, arguing, subjectivity) models significantly outperformed the baseline.
- *Extraversion.* For the *average* and *highly* extraverted readers, style and content have a similar impact. For *average*, the content, top style (liwc, subjectivity, evidence), and top content+style ({lemma, liwc, subjectivity, evidence} and {lemma, liwc}) models significantly outperformed the baseline. The analog holds for *high* extraversion with content and top style (emotion, arguing) models.
- *Neuroticism.* Here we find that the top content+style model performed best for *low* and *high* neuroticism, while the content model performed better for *average* neuroticism, however, without observed significance.
- *Openness.* Readers with *low* openness are impacted by content (content and top content+style were significantly better than style). However, those with *high* openness are impacted by style since we observe that top content+style is significantly better than the content model.[6]

Groups As shown in Table 4.B, for *other* personalities, the only models that significantly outperformed the random baseline are the two top content+style models ({lemma, liwc, arguing, evidence} and {lemma, liwc, emotion, arguing, evidence}) with macro-F_1 = 0.39. Content alone did not significantly outperform the baseline. On the contrary, *role model* readers seem not to be impacted by style, i.e., we did not observe any significant differences between the style models and the baseline.

6 Analysis of the Impact of Geographical Scopes

Given the importance of the *topic* and its role in persuasive text (Al Khatib et al., 2017), we conduct an analysis study considering both the readers' profiles and the topic of the editorials. In particular, we cluster the topics of the editorials and group them into three geographical scopes: (1) *Global* discusses global issues, such as the Iraq war. (2) *National* discusses national issues such as election, and (3) *State* discusses state (e.g. New York) related issues such as New York governor.

We conduct our analysis on the training sets as in section 5, following two settings: (i) using the whole training sets, and (ii) using the editorials that belong to each geographical scope in the training sets

[6]For *high* openness, two sets of top content+style outperform the content model: {lemma, liwc, emotion, evidence} and {lemma, emotion, subjectivity, evidence}.

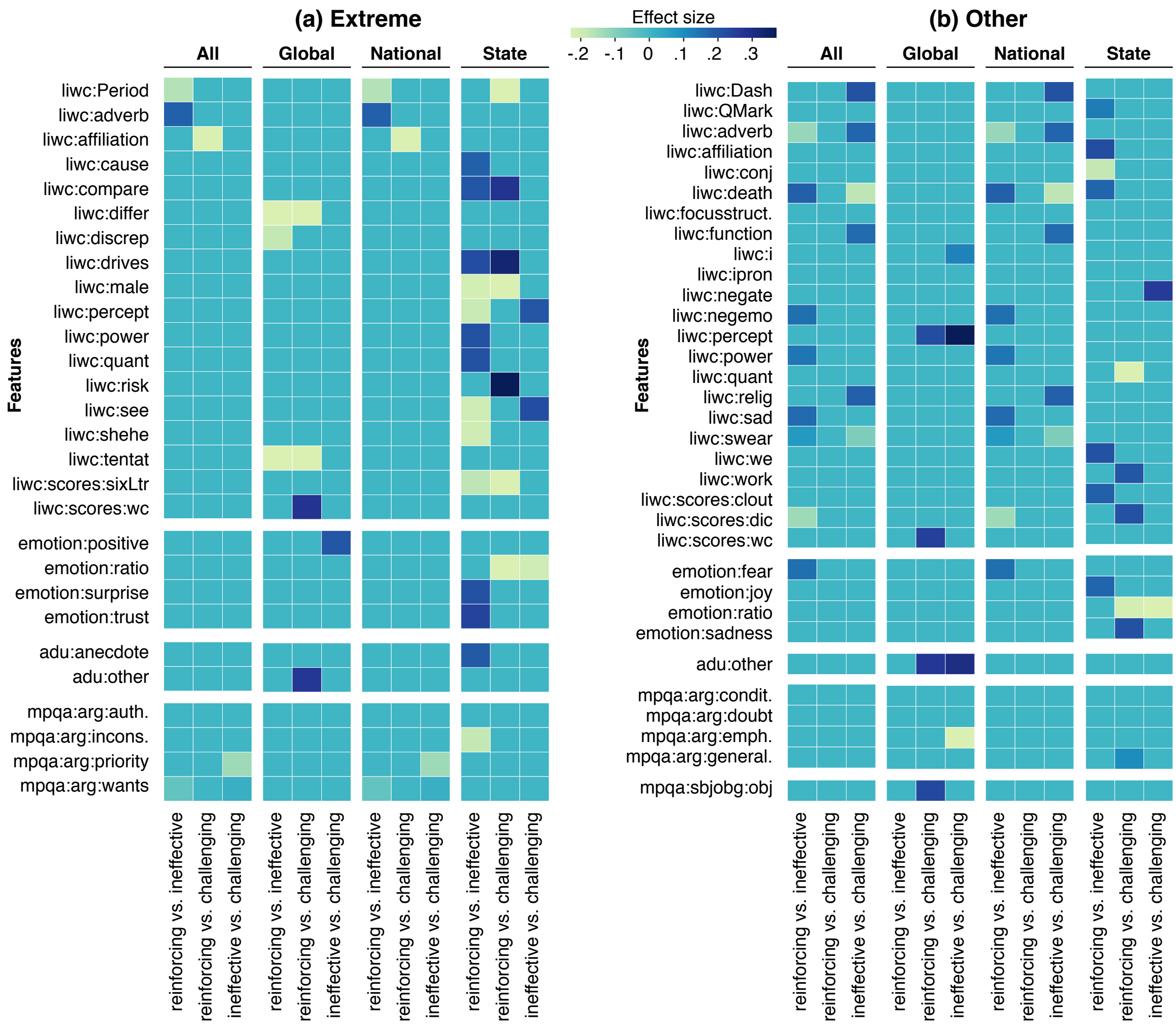

Figure 3: Heatmaps for each feature style for the two reader's profiles: extreme ideology ((a) Extreme) and and "Other" personality group ((b) Other). Each profile has four heatmaps, for each editorials' geographical scopes: *All*, *Global*, *National* and *State*. The y-axis represents the style features and the x-axis represents each effect-pair (a *vs.* b). Each effect size r value is indicated by a cube color: dark (light) color indicates that effect a (b) has significantly higher numbers of a style feature than effect b (a).

separately. For the reader profiles, we only consider the ones that were impacted by style (see Section 5), *extreme* ideology readers and non-*role models* ones.

Overall, our approach is divided into two steps: (1) Cluster the editorials into their three geographical scopes: *Global*, *National* and *State*. And (2) extract feature importance for each setting (i, ii), and profile (ideology intensity, *extreme* and personality group, *Other*).

6.1 Editorial Scope

For editorials topic clustering, we use Mallet latent Dirichlet allocation (Mallet-LDA) (Blei et al., 2003; McCallum, 2002). We employ it for several k (number of topics) and we calculate the coherence value for each k ranging from 2 to 30. The highest coherence value (0.52) is achieved with $k = 18$. The 18 topics cover issues related to the Bush administration, supreme court, tax, Iraqi war, Palestinian/Israeli conflict, immigration, nuclear weapon, energy, election, and more. We, then, hire an American annotator to map the 18 topics into meaningful groups. After inspecting each topic's keywords, he divides these topics into three geographical scopes. In total, we end up with 225 Global editorials, 475 for National editorials and 277 for State editorials.

New York City was put on notice a full decade ago that its black and Hispanic students were on the verge of being shut out of the elite public high schools that serve as a gateway to first-tier colleges and universities. The most damning analysis came from the community group ACORN. It called for sweeping curriculum changes at minority neighborhood middle schools, which typically lack the math, science and critical reading instruction necessary to prepare students for the entry test at the city's flagship high schools – Stuyvesant, Bronx Science and Brooklyn Tech.
The city should have seized on these findings as an opportunity to build a new middle school infrastructure in underserved neighborhoods. Instead, it opted for a poorly conceived and poorly run tutoring program that has now been exposed as a failure. [...]
The idea that students with decent preparation in the lower grades will automatically thrive is faulty. The city needs to attack the weaknesses of middle schools with the same urgency it has directed toward the elementary schools. New York needs a kind of Marshall Plan for its middle schools, especially in minority areas, with the specific aim of producing more high-performing minority students.
Getting there won't be easy. But the city needs to move with urgency and with all the resources at its disposal.

Figure 4: An excerpt of the news editorial with a *State* geographical scope, "Shutting Out Minorities". This editorial challenged the stance of annotators with extreme ideology.

The Supreme Court has been struggling to address the thorny question of when, if ever, punitive damages become so large that they violate the Constitution. The court made a good start when it laid down guidelines on when punitive damages are excessive. But eventually, it went too far. Today, it hears arguments in a case that offers a perfect opportunity to pull back to a more reasonable position.
The case involves Philip Morris's challenge to damages awarded to the widow of a smoker who died of lung cancer. An Oregon jury awarded Jesse Williams's widow, Mayola Williams, more than $821,000 in actual damages, and $79.5 million in punitive damages. Mrs. Williams said Philip Morris had engaged in 40 years of publicity to undercut concerns about cigarettes, even though it knew for most or all of that time that smoking was deadly. [...]
A final problem with the Supreme Court's rule of thumb on punitive damages is that it has been far less restrictive when it comes to punishing people. In 2003, the court held that California did not violate the ban on cruel and unusual punishment when it sentenced a man under its three-strikes law to 50 years for a theft of $153.53 worth of videotapes. That is a far more disproportionate punishment than Philip Morris got, for far less offensive conduct.

Figure 5: An excerpt of the news editorial with a *National* geographical scope, "Assessing the Damages". This editorial reinforced the stance of annotators with *Other* personality.

6.2 Style Impact within Geographical Scopes

Here, we study the impact of style (using style features) on readers with respect to reader's profile (*extreme* and *Other*) and geographical scope (e.g., national, all). To this end, we calculate, for each profile-scope, the significant differences between the persuasive effects (*challenging* vs. *reinforcing* vs. *ineffective*), for each of the style features (e.g. nrc:sad).

More precisely, for each feature (e.g., adu:anecdote), we measure significance using Anova (in case of homogeneity and normality) or Kruskal (otherwise). In the case of $p < 0.05$, we conduct post-hoc analysis (independent t-test in case of normality, Mann-Whitney otherwise) with Bonferroni correction for each effect-pair, and we calculated the effect-size r. Each heatmap, in Figure 3, shows the effect size [-0.23, +0.37] between each persuasive effect pair (e.g. challenging vs. ineffective) for all features with entailing significant differences within a pair.

For each profile-scope, we show, in Figure 3, only the style features if at least one effect-pair (e.g. *challenging* vs. *reinforcing*) has a significant difference for this style feature.

Extreme Ideology As shown in Figure 3.a, *All* and *National* editorials have similar pattern[7]. Whereas, *State* editorials differ from the other scopes. We observe that *reinforcing* editorials have significantly higher *adverbs* (liwc:adverbs) than *ineffective* editorials in both scopes (National/All). Also, within *State* editorials, the *emotional* (emotion:ratio) words are higher in *challenging* than *reinforcing/ineffective* (an excerpt is shown in Figure 4). Whereas, within the same scope, the same can be observed for *non-evidence* sentences (adu:other) for *reinforcing vs. challenging*. Within the *Global* scope, *ineffective* editorials have higher *positive words* (emotion:positive) than challenging ones.

[7]This can be due to the high number of National editorials in the dataset.

Other Personality We observe from Figure 3.b that *fear* (emotion:fear) words are higher in *reinforcing* editorials than *ineffective* ones within *All and National* scopes (an excerpt is shown in Figure 5). However, for *State* editorials, in general, emotional words (emotion:ratio) are higher for *challenging* editorials. And, the *liwc:clout*, which refers to the relative social status, confidence, and leadership displaced in a text, is significantly higher for *reinforcing* editorials compared to *ineffective*.

Figure 3 shows the difference of style features across the different geographical scopes and within different editorial's effect, revealing the importance of the topic when studying persuasiveness.

7 Conclusion

In this paper, we analyzed how linguistic choices, in news editorials, affect readers with different ideology intensities, personality traits and groups, filling the gap for El Baff et al. (2020) analysis. Argumentative text, especially editorials tend to be very challenging to study due to the strategic maneuver used by the authors who are considered (usually) expert writers. Therefore, the performance of predicting effectiveness is limited. In our work, we used one news editorial portal (The New York Times) with an obvious ideology (Liberal). The picture will be more complete if this analysis is conducted on news editorials with different ideologies. However, the purpose of this paper was to shed light on which linguistic choices affect which profile and on the importance of topical information when studying persuasiveness. Our findings can be employed in augmented writing tools, to help editorials writer improve their *message*, based on their *target*'s profile, to have a higher *impact*.

References

Khalid Al Khatib, Henning Wachsmuth, Johannes Kiesel, Matthias Hagen, and Benno Stein. 2016. A News Editorial Corpus for Mining Argumentation Strategies. In *26th International Conference on Computational Linguistics (COLING 2016)*, pages 3433–3443. Association for Computational Linguistics, dec.

Khalid Al Khatib, Henning Wachsmuth, Matthias Hagen, and Benno Stein. 2017. Patterns of Argumentation Strategies across Topics. In *2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017*, pages 1362–1368. Association for Computational Linguistics, sep.

Khalid Al Khatib, Michael Völske, Shahbaz Syed, Nikolay Kolyada, and Benno Stein. 2020. Exploiting Personal Characteristics of Debaters for Predicting Persuasiveness. In *58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, pages 7067–7072. Association for Computational Linguistics, July.

Aristotle. 2007. *On Rhetoric: A Theory of Civic Discourse* (George A. Kennedy, Translator). Clarendon Aristotle series. Oxford University Press.

Bal Krishna Bal and Patrick Saint Dizier. 2010. Towards building annotated resources for analyzing opinions and argumentation in news editorials. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. European Languages Resources Association (ELRA).

Bal Krishna Bal. 2009. Towards an analysis of opinions in news editorials: How positive was the year? (project abstract). In *Proceedings of the Eight International Conference on Computational Semantics*, pages 260–263. Association for Computational Linguistics.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Esin Durmus and Claire Cardie. 2018. Exploring the Role of Prior Beliefs for Argument Persuasion. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, volume 1, pages 1035–1045.

Esin Durmus, Faisal Ladhak, and Claire Cardie. 2020. The role of pragmatic and discourse context in determining argument impact. *arXiv preprint arXiv:2004.03034*.

Roxanne El Baff, Henning Wachsmuth, Khalid Al Khatib, and Benno Stein. 2018. Challenge or empower: Revisiting argumentation quality in a news editorial corpus. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 454–464. Association for Computational Linguistics.

Roxanne El Baff, Henning Wachsmuth, Khalid Al Khatib, and Benno Stein. 2020. Analyzing the persuasive effect of style in news Editorial Argumentation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3154–3160, Online, July. Association for Computational Linguistics.

Martin Gerlach, Beatrice Farb, William Revelle, and Luís A Nunes Amaral. 2018. A robust data-driven approach identifies four personality types across four large data sets. *Nature human behaviour*, 2(10):735–742.

Lewis R. Goldberg. 1990. An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, 59(6):1216–1229.

Ivan Habernal and Iryna Gurevych. 2016. Which argument is more convincing? Analyzing and predicting convincingness of web arguments using bidirectional LSTM. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1589–1599. Association for Computational Linguistics.

Helena Halmari and Tuija Virtanen. 2005. *Persuasion across genres: a linguistic approach*, volume 130. John Benjamins Publishing.

Christopher Hidey and Kathleen R McKeown. 2018. Persuasive influence detection: The role of argument sequencing. In *AAAI*, pages 5173–5180.

Christopher Hidey, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathy McKeown. 2017. Analyzing the semantic types of claims and premises in an online persuasive forum. In *Proceedings of the 4th Workshop on Argument Mining*, pages 11–21.

Oliver P John, Eileen M Donahue, and Robert L Kentle. 1991. The big five inventory – versions 4a and 54.

Richard Joiner and Sarah Jones. 2003. The effects of communication medium on argumentation and the development of critical thinking. *International journal of educational research*, 39(8):861–871.

Johannes Kiesel, Khalid Al Khatib, Matthias Hagen, and Benno Stein. 2015. A Shared Task on Argumentation Mining in Newspaper Editorials. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 35–38. Association for Computational Linguistics.

Stephanie Lukin, Pranav Anand, Marilyn Walker, and Steve Whittaker. 2017. Argument Strength is in the Eye of the Beholder: Audience Effects in Persuasion. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 742–753. Association for Computational Linguistics.

Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu.

George Miller. 1998. *WordNet: An electronic lexical database*. MIT press.

Saif M Mohammad and Peter D Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Daniel J. O'Keefe. 2015. *Persuasion: Theory and research*. Sage Publications.

James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The Development and Psychometric Properties of LIWC2015.

Isaac Persing and Vincent Ng. 2015. Modeling Argument Strength in Student Essays. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 543–552. Association for Computational Linguistics.

Isaac Persing and Vincent Ng. 2017. Lightly-supervised modeling of argument persuasiveness. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 594–604, Taipei, Taiwan, November. Asian Federation of Natural Language Processing.

Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*, pages 3–33. Elsevier.

Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*.

Tatjana Scheffler and Manfred Stede. 2016. Realizing Argumentative Coherence Relations in German: A Contrastive Study of Newspaper Editorials and Twitter Posts. *Patrick Saint-Dizier*, page 73.

Swapna Somasundaran, Josef Ruppenhofer, and Janyce Wiebe. 2007. Detecting arguing and sentiment in meetings. In *Proceedings of the SIGdial Workshop on Discourse and Dialogue*, volume 6.

Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.

Teun A. van Dijk. 1995. Opinions and Ideologies in Editorials. In *Proceedings of the 4th International Symposium of Critical Discourse Analysis, Language, Social Life and Critical Thought*, Athens, June.

Henning Wachsmuth, Khalid Al Khatib, and Benno Stein. 2016. Using Argument Mining to Assess the Argumentation Quality of Essays. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1680–1691. The COLING 2016 Organizing Committee.

Henning Wachsmuth, Nona Naderi, Yufang Hou, Yonatan Bilu, Vinodkumar Prabhakaran, Tim Alberdingk Thijm, Graeme Hirst, and Benno Stein. 2017. Computational argumentation quality assessment in natural language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 176–187. Association for Computational Linguistics.

Henning Wachsmuth, Manfred Stede, Roxanne El Baff, Khalid Al Khatib, Maria Skeppstedt, and Benno Stein. 2018. Argumentation synthesis following rhetorical strategies. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3753–3765. Association for Computational Linguistics.

Janyce Wiebe and Ellen Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. In *International conference on intelligent text processing and computational linguistics*, pages 486–497. Springer.

Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 129–136. Association for Computational Linguistics.