

Affective valence in posts on X (formerly Twitter) and evaluation of a politician's image

Abstract

The aim of this study was to examine whether the use of affective language by politicians impacts the social evaluation of their image, as measured by the semantic differential method developed by Cwalina et al. (2000). In the study, Polish participants ($N = 958$) evaluated the profiles of three well-known Polish politicians from different parties: Bosak (right-wing), Trzaskowski (center), and Biedroń (left-wing). Participants made evaluations before and after reading hateful, kind or neutral tweets about refugees from Ukraine. We controlled for participants' political views and demographic variables. The results confirmed changes in evaluations of the image of politicians depending on the language they used. The "positivity bias" identified in the study provides encouragement to use kindness speech in public discourse, showing that this type of communication can improve a politician's ratings. Detailed analyses also showed that: a) the participants' congruence of political views was correlated with ratings for all dimensions, b) there were statistical differences in evaluations between women and men, and c) there was a relationship between age and evaluations on several dimensions of the scale.

Key words: hate speech, kindness speech, politician's image

In the view of many researchers (Farkas & Bene, 2021; Grofman, 2005; Wattenberg, 1991) we are living in an era of candidate-centered politics. This means that the electorate's attention shifts from political parties to specific candidates running for various political offices. The shift

is accompanied by the growing importance of a candidate's individual characteristics of which makes up his or her image: a particular type of representation that can impart additional values to an object by evoking emotional associations (Cwalina et al., 2015). The reception of an image is affected both by its core features, the most significant ones for its evaluation, and by peripheral features related to the politician's additional activity, lifestyle, and manner of communicating with the public. Contemporary social discourse promotes a negative image of reality, and the statements of public figures reinforce and legitimize the use of language marked by hateful emotions (Bilewicz & Soral, 2020). Much rarer in public debate is the so-called 'kindness speech', which expresses positive emotions and a friendly attitude towards one's interlocutor (Cavazza, 2017). The language used by politicians influences the assessment of their credibility (Frimer & Skitka, 2018), the affective polarisation of social attitudes (Druckman et al., 2019) and can also build (or lower) trust in the political system and institutions (Van't Riet & Van Stekelenburg, 2022).

The aim of this study was to examine whether the social evaluation of a politician's image is affected by their use of hate speech and kindness speech in political posts on X. We decided to choose "Twitter" as a platform to analyse because, Twitter is often the primary political platform in Poland, and also, unlike other social networks (such as Facebook or Instagram), it facilitates short and direct messages that are comparably impulsive and affective. Twitter's character limit constraints discourse complexity inhibits reflexivity and tends to encourage uncivil behaviour due to the impersonal nature of the platform (Ott, 2017; Salimi & Mortazavi, 2024).

Hate speech and kindness speech in the context of politics

Hate speech, an extreme form of derogatory language (Cervone et al., 2021), is used by journalists, commentators and internet users, but also (and perhaps especially) by politicians, whether in debates in parliament, in the media or in their election campaigns (Cwalina &

Koniak, 2023). Its emotional basis is contempt, which triggers anger and revulsion (Winiewski et al., 2017). Hate speech plays an enormous role in present-day social life and is a tool in political struggle and in public debate, excluding the possibility of dialogue and understanding (Obrębska, 2020). The number of hateful messages is steadily rising, especially on the Internet, which offers a sense of impunity and anonymity (Cervone et al., 2021). In the digital environment, hate speech emphasizes group identity and denigrates people in the ‘‘out-group’’ (Papcunová et al., 2023). It carries a number of negative consequences on many levels: not only leave a trauma in discriminated people (Leets & Giles, 1999; Mullen & Smyth, 2004; Swim et al., 2009), but also changes the way in which minority groups are perceived by the rest of the society (Cervone et al., 2021). Previously tolerant people after contact with verbal aggression may change their attitude towards discriminated groups, assuming that if they are evaluated as such by the majority (or by political leaders), they must be to some extent responsible for the unfriendliness that they encounter. The studies on desensitisation (Allen et al., 2018; Fasoli et al., 2015; Soral et al., 2018) show that the very exposure to speech hate strengthens the inclination for dehumanisation and also causes that after some time it is no longer considered offensive, shocking or violating social standards.

The use of aggressive language in politics can be beneficial: when talking about social categories disliked by listeners, politically incorrect speakers are perceived as more authentic (Rosenblum et al., 2019), negative messages decreased the perceived warmth of the politician while simultaneously increasing the perceived competence (Carraro & Castelli, 2010), the use of vulgar language and a sarcastic tone when referring to political opponents in an election context indicates a politician's strength and determination to win (Ntontis et al., 2024). Furthermore, Sturm-Wikerson and Johnson (2021), in an analysis of comments on immigration posted on Facebook during the 2016 US presidential primaries, found that when candidates used uncivil language, the number of uncivil comments posted by other users on their profiles

also increased. The use of uncivil language can therefore result in an increase in affective polarisation, which may in turn lead to more extreme views.

Much more rarely encountered in public debate is what may be called “kindness speech”, expressing positive emotions and a kindly attitude to one’s interlocutor. Meanwhile, research shows many positive consequences of this type of communication, including improved interpersonal sensitivity (Williams et al., 2018), well-being (Cotney & Banerjee, 2019), and even psychological welfare (Shin et al., 2020; Cregg & Cheavens, 2022). In relation to minority groups who were described in a kind manner, researchers observed increased sympathy and more positive assessments (Cervone et al., 2021). Much research has been devoted to the role of flattery in political campaigns. Cavazza (2017) showed experimentally, that praising a political opponent evokes an audience’s positive emotions, which in turn positively affects source reliability, and finally enhances the likelihood of voting for that source. Thus, kindness speech fosters positive evaluation of the persons using it, but in line with the negativity bias effect (Baumeister et al., 2001; Holleman et al., 2021; Rozin et al., 2010) its impact is weaker than in the case of negative language.

Demographic variables, such as the gender of the politician or the audience (Schultz & Pancer’s, 1997), race and religion (McLaughlin & Thompson, 2016), ethnicity and age (Flannelly, 2002), sexual orientation (Beyerlein & Klocek, 2020), and similarity of political views (Frimer & Skitka, 2018), also play an important role in assessing the language used by politicians. Schultz and Pancer’s (1997) study examined perceptions of political candidates (male and female) making negative statements about the personality and integrity of their political opponents. Participants who read a speech by a candidate of their own gender were more likely to rate the candidate as having greater integrity when the candidate attacked his or her opponent than when he or she did not. When judging a candidate of the opposite sex, participants tended to rate the candidate who attacked his/her opponent as having less integrity than a candidate who did not attack his/her opponent. The results are discussed in terms of the

impact that aggressive campaign tactics may have on voters' perceptions, and how similarity between voters and candidates may influence perceptions of such tactics.

Affective messages may thus play a complex, interactive role in various social behaviors, as is well explained by the classical Affect Infusion Model (AIM) of Joseph Forgas (2001, 2002) and its more recent validations (Fehrenbacher, 2017; Ling et al., 2023; Pérez-Fernández et al., 2022; Winkielman et al., 2007). The phenomenon of affect infusion occurs as a result of affect priming, a process in which affectively marked information has an impact on an individual's cognitive and behavioral processes, becoming a part of them and coloring them with a specific affective valence. Research has shown that affect infusion occurs most frequently during the active processing of new, ambiguous, and complex information, because broader, more complex processing increases the probability of subconscious inclusion of affectively colored information in the process of planning one's behaviors. Social evaluations are an example of complex substantive processing which is often ambiguous and requires personal engagement, and which is susceptible to affect infusion and affect priming effects.

The effect of emotionally marked messages on the social evaluation of the speaker can also be explained in a good way by two other mechanisms: spontaneous trait transference (STT; Carlston & Skowronski, 2005) and transfer of attitudes recursively (TAR; Gawronski & Walther, 2008). Spontaneous trait transference occurs when communicators are perceived as possessing the very traits they describe in others. It reflects simple associative processes that occur even when there are no logical bases for making inferences. The transfer of attitudes recursively, on the other hand, refers to the feedback effect of another person's pronounced evaluation on the formation of a similar attitude towards the speaker. Gawronski and Walther (2008) consider this to be an attributional process. When people hear a person's (e.g. a politician's) insulting and repeated statements about others, they infer that the person who has a negative and arrogant attitude towards others deserves such treatment themselves. And vice versa: people who like others acquire a positive valence.

Based on the psychological mechanisms presented, it can be assumed that a positively marked message (kindness speech) will lead to the formulation of more positive social evaluations, and a negative message (hate speech) to more negative ones. This suggests the following research problems:

1. Will evaluations of the image of individual politicians become more negative following their use of hate speech?
2. Will evaluations of the image of individual politicians become more positive following their use of kindness speech?
3. Are the evaluations of politicians made before and after the use of hate speech or kindness speech dependent on the political views, gender and age of the persons making the evaluations?

The present research

The research was of an experimental nature, conducted in a scheme of 3 (images of right-wing, left-wing, and centrist politicians) \times 3 (messages with negative, positive, and neutral emotional marking). The dependent variable in the analysis was a politician's image. The primary independent variable was the emotional marking of the tweets (negative and positive), while secondary independent variables were the gender, age, and political views of the study participants. The emotionally neutral message had the status of a control variable, providing the ability to examine whether a change in the evaluation of a politician's image was really affected by the emotional marking of the tweets used in the study.

This study was carried out under a research project funded by the Polish National Science Center titled *ComPathos: towards the new model of pathos for computational rhetoric* (2020/39/D/HS1/00488), and received a positive opinion regarding the ethical problems described in the submitted research project, issued by the Ethics Committee for Research

Projects functioning at the Faculty of Psychology and Cognitive Science of Adam Mickiewicz University in Poznań.

Method

Pilot study. The main study was preceded by a pilot study carried out online, on a group of 105 persons, in April 2021. The purpose of the pilot study was to test the research procedure. The results were published in an article xxx. As a consequence of the pilot study, the following changes were made to the project: 1) so-called “kindness speech” was included in the project; 2) messages with emotionally neutral marking were included in the project as controls; 3) the list of politicians was updated; 4) it was planned to present the politicians to subjects in random order rather than in a fixed order; 5) the tweets selected for study were made uniform in terms of their affective valence – positive for kindness speech and negative for hate speech – using methods of computational linguistics.

Participants. A group of 958 participants (53.4% women; $M_{age} = 46$, $SD_{age} = 16$) was recruited for the study. There was an equal number of participants for each of the experimental and control conditions. Participants were recruited via the Ariadna research panel (<https://panelariadna.pl/>), an online platform where users can register and complete questionnaires in exchange for points redeemable for material rewards. Users on this platform may choose to participate in any number of available surveys. Identification is ensured by delivering these rewards to physical addresses provided by users, confirming their identities.

The largest number of participants was from rural areas (36.64%), followed by participants from medium-sized metropolitan areas (20.46%), large metropolitan areas (17.22%), small metropolitan areas (13.05%), and very large metropolitan areas (12.63%). Almost half of the group had completed higher education (45.72%), followed by those with a

high school diploma (including undergraduate students; 39.35%), vocational education (12.42%), and a primary school diploma (2.51%). All subjects were of Polish nationality.

Materials. The Polish politicians selected for study were ascertained by competent judges to be most representative for particular political groupings: Krzysztof Bosak (representing the right), Rafał Trzaskowski (the center), and Robert Biedroń (the left). The selected politicians are also similar to each other in terms of such demographic variables as gender, age, and level of education. None of the politicians selected is known to use or spread direct hate speech. To confirm if the politicians are comparable in terms of the emotional valence of their usual Twitter (X) activity, we collected a sample of approximately 2,000 tweets for each politician and performed an analysis of valence using the method established in Konat et al. (2024), who applied an automated method based on psychologically established lexicons of emotional words (mainly Wierzba et al. 2015; Wierzba et al. 2022). We found that Rafał Trzaskowski's tweets were predominantly positive (68.7%), with a smaller proportion being neutral (13.0%) or negative (18.3%). Similarly, Robert Biedroń's tweets were largely positive (52.8%), followed by neutral (17.1%) and negative (30.1%). In this context, Krzysztof Bosak's tweets were slightly different, with 41.6% positive, 18.3% neutral, and 40.1% negative tweets. These findings suggest that even though Biedroń and Trzaskowski are in general more positive in their online posting than Bosak, still, for all three candidates the dominant category is positive tweets. This empirical evidence supports our claim that the three selected politicians are comparable and none is associated with abnormal negative activity in the online setting.

The study included specific items to assess participants' familiarity and political alignment with the politician. Familiarity was measured with a binary question: "Is the presented politician known to you?", with response options "yes" and "no". To assess political alignment, participants were asked to rate their agreement with the views of a politician on a five-point Likert scale, from 1 (strongly disagree) to 5 (strongly agree). A midpoint option, "I

don't know", was provided to capture ambivalence or unfamiliarity with the politician's stance. Respondents unfamiliar with the politician were instructed to skip this question. These items provided a basic but direct measure of familiarity and political alignment as relevant to our analysis.

The image of the selected politicians was evaluated using the Polish adaptation of the semantic differential scale developed by Cwalina et al. (2000). The scale comprises 14 bipolar seven-point dimensions formed by pairs of opposite adjectives: "qualified–unqualified," "cosmopolitan–provincial," "honest–dishonest," "believable–unbelievable," "successful–unsuccessful," "attractive–unattractive," "sincere–insincere," "calm–excitable," "unaggressive–aggressive," "strong–weak," "active–inactive," "believer–nonbeliever," "sophisticated–unsophisticated," and "friendly–unfriendly."

The material employed in the study consisted of tweets on the topic of Ukrainian refugees in Poland. The topic was selected in view of its current relevance in social discourse and its strong emotional appeal, polarizing Polish society (Wypych & Bilewicz, 2024). We designed three categories of tweets for each politician – one expressing positive sentiment, one negative, and one neutral. The tweets used in the study were inspired by real tweets written by politicians; however, they were fabricated by the researchers for the purpose of the study. We decided to rely on fabricated posts rather than real ones due to several challenges encountered with natural language data. Initially, we collected a sample of over 1,000 tweets for each politician and conducted an automated valence analysis to identify posts with the most emotional content. Based on this analysis, we selected candidate tweets for use in the experiment. However, we found it challenging to control for key variables such as length, concise topic focus, and sentiment consistency. Additionally, it was particularly difficult to identify truly neutral tweets on the topic, as it is uncommon for politicians to produce neutral content on such polarizing issues. To address these challenges, we created fabricated posts inspired by the selected material and informed by our understanding of each politician's typical

style. To ensure intersubjectivity and validate the emotional tone of the fabricated posts, we tested their valence using computational linguistic methods.

The validation methodology included two stages – the first using automatic transformer-based classifiers available in the Transformers Python library (Wolf et al., 2020), and the second following lexicon-based sentiment analysis. The first method allows a text sample to be assigned to one of three categories of sentiment based on the full content of tweets. The second involves the extraction of emotion-laden words based on emotional word lists created by psychologists (Riegel et al., 2015; Imbir, 2016; Wierzba et al., 2021). Here, if the number of positive-valenced words was higher than the number of negative-valenced words, the tweet was regarded as positive, and vice versa. In case of an equal number of positive and negative words, we checked the intensity of sentiment available in emotional word lists and chose the higher value as the final category of expressed sentiment. Neutral tweets were required to have no emotional words. Both methods confirmed the category of sentiment expressed in the tweets; that is, tweets designated as positive were confirmed by both methods to express positive sentiment, negative tweets yielded negative labels from both methods, and neutral tweets were given neutral labels by both methods. The choice of the method for validating valence of tweets is dictated by the nature of the task - judgements of emotional tone of a message are highly subjective, thus the consensus is rarely achieved even with large samples of raters (Gajewska & Konat, 2023). On top of that, the choice of the topic in material is highly polarizing for Poles (Ukrainian refugees in Poland), therefore, it is highly probable that the judgements of competent judges could have been biased. For that reason, we decided to use more objective and intersubjective methods from computational linguistics, which were thoroughly tested and normalized by psychologists to ensure their high reliability. The tweets used in the study are available in the appendix.

In addition, we collected X (Twitter) profiles of the chosen politicians from their X (Twitter) accounts, in order to ask participants about their political alignment with the politician

they were evaluating. As distractor material we employed a short animated video from the YouTube platform. The video material had been tested positively as a distractor in the pilot study.

Procedure. The survey was conducted online in April and May 2023. Participants were first informed about the aims of the study and made familiar with the tasks. They were then asked to provide demographic information relating to their sex, age, place of residence, education, and nationality. Each participant was randomly assigned to one of the three politicians in the study. They were then presented with the politician's profile associated with his Twitter account, and asked about their familiarity and political alignment with the politician. Next, participants evaluated the image of the selected politician on the semantic differential scale (Cwalina et al., 2000). A short video was then displayed as distractor material prior to the second part of the study. Participants answered control questions relating to the content of the video. Analysis of the answers to the control questions also made it possible to determine the level of attentiveness of the participants during the testing procedure.

Each participant was then randomly assigned to one out of three stimulus conditions – a tweet expressing positive, negative, or neutral sentiment. Finally, participants were asked again to evaluate the image of the selected politician on the semantic differential scale. After the study, participants were informed that the tweets presented in the study were fake, not authored by the politicians, and designed only for the purpose of the study.

Statistical analysis. With the use of one-way MANOVA, we test the role of tweet valence (positive, negative, neutral) in the ratings of politicians' image. Then, we perform a two-way MANOVA to test the influence of tweet valence and political orientation of the presented politician (images of right-wing, left-wing, and centrist politicians) on the ratings of politicians' image; and the influence of political orientation and participants' gender on the ratings of

politicians' image. Post hoc Tukey's HSD tests are performed to unveil group differences in such cases. Posttest–pretest comparisons were performed with the paired-samples t -test for each stimulus condition (positive, negative, neutral). Differences between genders and between politicians on the semantic differential scale were tested with the independent-samples t -test and Bonferroni correction for multiple comparisons when applicable. Statistical analysis was run in Python with the use of the Statsmodels library for MANOVA and ANOVA and post hoc Tukey's HSD analysis, as well as the Scipy library implementation of the t -test, Spearman's rho coefficient, and Kendall's tau-b (τ_b) rank correlation coefficient. Effect sizes are calculated with the use of Cohen's d .

Results

Negative stimulus. Posttest–pretest comparisons for the negative stimulus condition showed statistically significant differences in seven dimensions of the scale. Posttest ratings were significantly lower for provincial–cosmopolitan ($t = -5.79, p < 0.01, d = 0.29$), dishonest–honest ($t = -2.76, p < 0.01, d = 0.12$), insincere–sincere ($t = -3.21, p < 0.01, d = 0.15$), excitable–calm ($t = -6.19, p < 0.01, d = 0.30$), aggressive–unaggressive ($t = -6.67, p < 0.01, d = 0.37$), nonbeliever–believer ($t = -2.50, p < 0.01, d = 0.11$), and unfriendly–friendly ($t = -4.99, p < 0.01, d = 0.25$). The described relationships are illustrated in Figure 1.

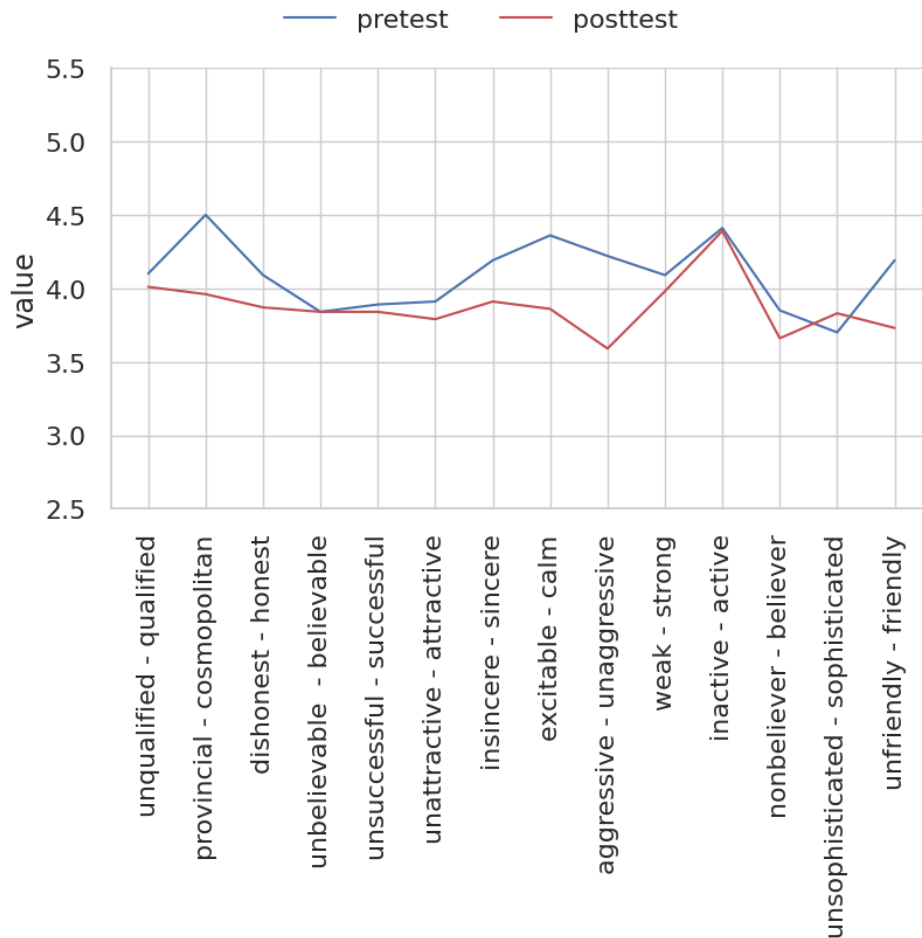


Fig. 1. Pretest–posttest differences in ratings for the negative stimulus condition

Posttest–pretest comparisons for individual politicians under the negative stimulus condition reveal several statistically significant results. First, Krzysztof Bosak (right) was seen as more provincial ($t = -3.86, p < 0.01, d = 0.37$), dishonest ($t = -1.75, p = 0.04, d = 0.16$), insincere ($t = -2.39, p < 0.01, d = 0.22$), excitable ($t = -3.89, p < 0.01, d = 0.41$), aggressive ($t = -4.14, p < 0.01, d = 0.42$), nonbelieving ($t = -4.56, p < 0.01, d = 0.4$), and unfriendly ($t = -3.78, p < 0.01, d = 0.39$) than in the pretest. Second, Robert Biedroń (left) was viewed as more provincial ($t = -3.64, p < 0.01, d = 0.37$), dishonest ($t = -2.72, p < 0.01, d = 0.22$), unattractive ($t = -1.66, p = 0.04, d = 0.22$), insincere ($t = -2.27, p = 0.01, d = 0.2$), excitable ($t = -4.49, p < 0.01, d = 0.39$), aggressive ($t = -4.65, p < 0.01, d = 0.48$), and unfriendly ($t = -3.13, p < 0.01, d = 0.31$). Lastly, Rafał Trzaskowski (center) was seen as more provincial ($t = -2.56, p < 0.01,$

$d = 0.19$), excitable ($t = -2.08, p = 0.02, d = 0.14$), and aggressive ($t = -2.62, p < 0.01, d = 0.22$) by participants after reading a negative tweet than in the pretest.

Positive stimulus. Posttest–pretest comparisons for the positive stimulus condition show statistically significant differences for all dimensions of the scale – participants gave significantly higher ratings after reading a positive tweet. The results are given in Table 1.

Table 1

Pretest–posttest differences in ratings under a positive stimulus condition

Adjective	Pretest	Posttest	Statistic		
	<i>M</i>	<i>M</i>	<i>t</i>	<i>p</i>	<i>d</i>
unqualified–qualified	4.21	4.47	3.43	<0.001	0.14
provincial–cosmopolitan	4.24	4.64	4.57	<0.001	0.21
dishonest–honest	4.12	4.33	3.01	0.001	0.11
unbelievable–believable	4.04	4.25	2.64	0.004	0.11
unsuccessful–successful	3.95	4.22	3.43	<0.001	0.15
unattractive–attractive	4.01	4.21	2.77	0.003	0.11
insincere–sincere	4.17	4.37	2.64	0.004	0.11
excitable–calm	4.34	4.57	3.32	<0.001	0.14

aggressive– unaggressive	4.25	4.47	2.54	0.01	0.13
weak– strong	4.16	4.38	2.88	0.002	0.12
inactive– active	4.58	4.61	0.39	0.35	0.12
nonbeliever– believer	3.96	4.22	3.23	<0.001	0.14
unsophisticated– sophisticated	3.82	3.95	1.43	0.08	0.14
unfriendly– friendly	4.25	4.55	3.71	<0.001	0.16

The results are further illustrated in Figure 2.

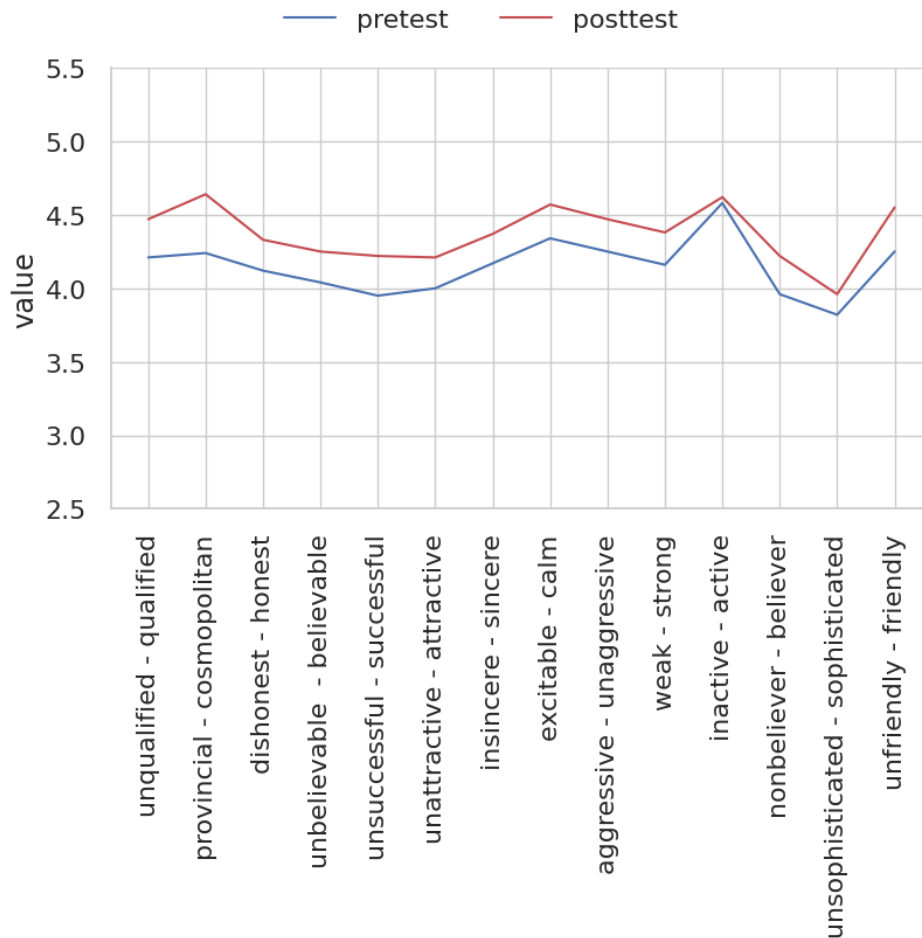


Fig. 2. Pretest–posttest differences in ratings under a positive stimulus condition

Posttest–pretest comparisons performed separately for each politician under the positive stimulus condition show two statistically significant differences for right-wing Krzysztof Bosak (in the provincial–cosmopolitan and unfriendly–friendly dimensions), 11 statistically significant results for Robert Biedroń (left), and eight such differences for Rafał Trzaskowski (center). Ratings on all dimensions of the scale here are significantly higher in the posttest than in the pretest. Detailed results for individual politicians are given in the appendix.

Neutral stimuli. Analysis of differences in posttest–pretest ratings under a neutral stimulus condition reveals two statistically significant results – higher ratings on the unattractive–attractive ($t = 2.75, p < 0.01, d = 0.11$) and aggressive–unaggressive ($t = 3.03, p < 0.01, d = 0.13$) dimensions – although the effect sizes are marginal. The results are illustrated in Figure 3.

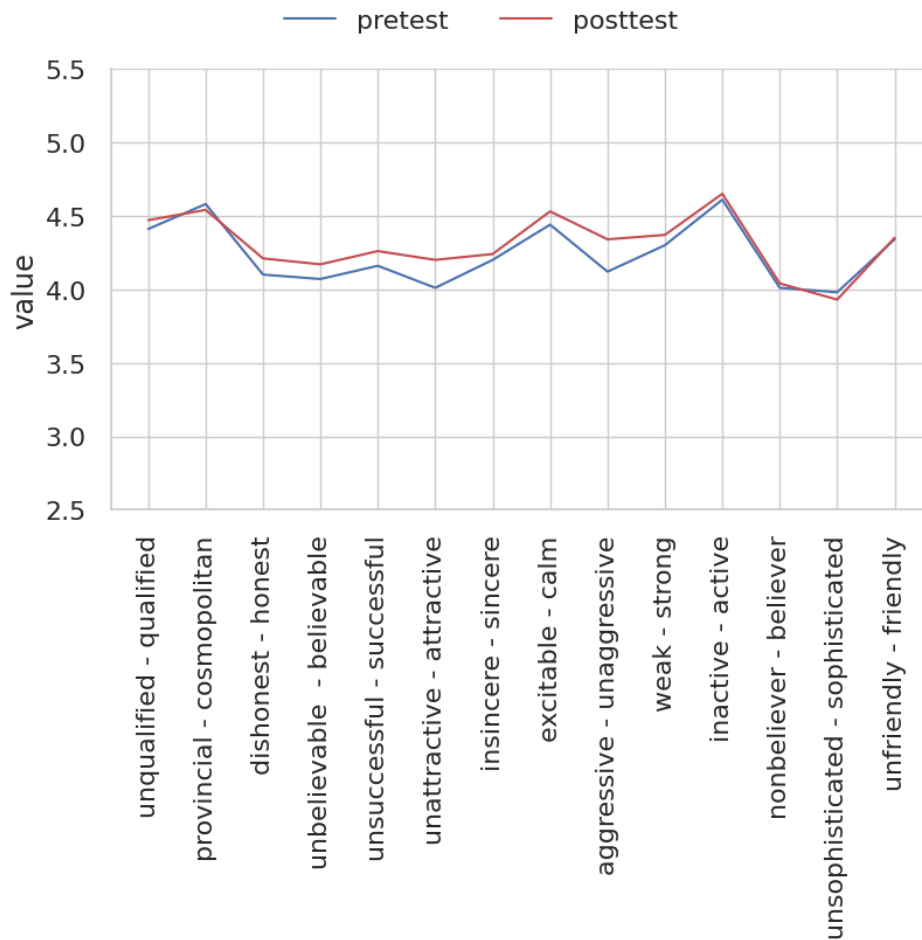


Fig. 3. Pretest–posttest differences in ratings under a neutral stimulus condition

Posttest–pretest comparisons performed separately for each politician under the neutral stimulus condition show four statistically significant differences for Krzysztof Bosak – representing political right, in the unsuccessful–successful ($t = 2.13, p = 0.03, d = 0.17$), unattractive–attractive ($t = 2.42, p = 0.02, d = 0.18$), aggressive–unaggressive ($t = 2.11, p = 0.04, d = 0.16$) and nonbeliever–believer ($t = -1.99, p = 0.048, d = 0.16$) dimensions; one statistically significant result for Robert Biedroń - representing political left (excitable–calm: $t = 2.3, p = 0.02, d = 0.16$); and one such difference for Rafał Trzaskowski – representing political center (unfriendly–friendly: $t = 2.1, p = 0.04, d = 0.15$).

Stimulus. The results indicate a significant difference in the post-test ratings of politicians' image between stimulus conditions, $F(28, 1884) = 3.53, p < 0.001$. Post hoc Tukey test reveals differences between positive and negative conditions for all but two dimensions of the image: active and sophisticated. Differences between negative and neutral conditions are significant for all but four ratings: believable, sincere, active and sophisticated. We find no significant differences between positive and neutral conditions. Results of a one-way MANOVA, performed separately for each politician, are also significant, Bosak (right): $F(28, 614) = 2.56, p < 0.001$; Trzaskowski (center): $F(28, 606) = 1.52, p = 0.04$; Biedroń (left): $F(28, 600) = 2.58, p < 0.001$.

Comparisons between politicians. In the case of the negative stimulus condition, statistically significant differences were recorded for two dimensions. Participants who had read a negative tweet gave significantly lower ratings in the nonbeliever–believer dimension to Krzysztof Bosak (right) than to both Rafał Trzaskowski (center) and Robert Biedroń (left), with medium effect sizes ($t = -4.65, p < 0.01, d = 0.64$; and $t = -2.98, p = 0.01, d = 0.41$). The change in ratings was -0.64 for Bosak (right), -0.06 for Biedroń, and 0.13 for Trzaskowski (center). Thus, Bosak (right) was seen as more of a nonbeliever by participants who had read a negative tweet than Biedroń (left) and Trzaskowski (center). We also observe a statistically significant difference for the excitable–calm dimension, where participants rate Biedroń (left) more negatively after reading a negative tweet than in the case of Trzaskowski ($t = 2.44, p = 0.05, d = 0.33$). The change in this dimension was larger for Biedroń than for Trzaskowski (-0.69 vs. -0.23). Among participants who had read a positive tweet, Bosak (right) was rated significantly lower than Trzaskowski (center) on the inactive–active dimension ($t = -2.68, p = 0.02, d = 0.37$). Here, on average, Trzaskowski was rated as 0.32 more active in the posttest than in the pretest, while Bosak (right) was seen as 0.26 less active in the posttest. Finally, in the neutral stimulus case, the change in semantic differential ratings was significantly higher for Bosak

(right) than for Trzaskowski (center) and Biedroń (left) on the unfriendly–friendly ($t = -2.41, p = 0.05, d = 0.33$) and nonbeliever–believer ($t = -2.74, p = 0.02, d = 0.37$) dimensions, respectively.

Congruence of political views. We found that congruence of participants’ political views was correlated with ratings for all dimensions of the semantic differential scale in pretest and posttest with negative and neutral stimuli, and in all but one dimension in posttest with a positive stimulus. Correlation coefficients are calculated with the use of Kendall’s tau-b (τ_b) rank correlation. Detailed results are given in Table 2.

Table 2

Kendall’s tau-b (τ_b) correlation coefficient between congruence of political views and dimensions of the semantic differential in pretest and posttest

Adjective	Pretest		Posttest positive		Posttest negative		Posttest neutral	
	τ_b	p	τ_b	p	τ_b	p	τ_b	p
unqualified–qualified	0.44	<0.001	0.45	<0.001	0.48	<0.001	0.50	<0.001
provincial–cosmopolitan	0.41	<0.001	0.41	<0.001	0.43	<0.001	0.44	<0.001
dishonest–honest	0.61	<0.001	0.54	<0.001	0.50	<0.001	0.57	<0.001
unbelievable–believable	0.54	<0.001	0.46	<0.001	0.53	<0.001	0.56	<0.001
unsuccessful–successful	0.47	<0.001	0.45	<0.001	0.52	<0.001	0.47	<0.001

unattractive– attractive	0.51	<0.001	0.46	<0.001	0.52	<0.001	0.55	<0.001
insincere– sincere	0.57	<0.001	0.49	<0.001	0.52	<0.001	0.55	<0.001
excitable– calm	0.39	<0.001	0.35	<0.001	0.42	<0.001	0.43	<0.001
aggressive– unaggressive	0.49	<0.001	0.44	<0.001	0.46	<0.001	0.49	<0.001
weak–strong	0.44	<0.001	0.45	<0.001	0.53	<0.001	0.46	<0.001
inactive– active	0.38	<0.001	0.42	<0.001	0.34	<0.001	0.34	<0.001
nonbeliever– believer	0.11	<0.001	0.13	0.01	0.28	<0.001	0.10	0.02
unsophisticated– sophisticated	0.11	<0.001	0.04	0.41	0.17	<0.001	0.11	0.01
unfriendly– friendly	0.48	<0.001	0.49	<0.001	0.51	<0.001	0.54	<0.001

The differences between groups were tested using the independent sample *t*-test with Bonferroni correction for multiple comparisons. The results of this analysis are given in Table 4 in the appendix. Effect sizes for these results vary from small to large. Small effects are observed for the nonbeliever–believer dimension, and medium or large effect sizes for the other 13 dimensions. The results show that participants who agree with a politician’s political views tend to rate the image of that politician more positively in the pretest than do neutral/undecided participants and those with opposite political views. Similarly, neutral/undecided participants rate the image more positively than opponents of the politician. Moreover, the effect of these differences is larger between aligned and neutral participants than between the neutral and

opposing groups. Therefore, the positivity of supporters is larger than the negativity of opponents in evaluating a politician’s image. These relationships are illustrated in Figure 4.

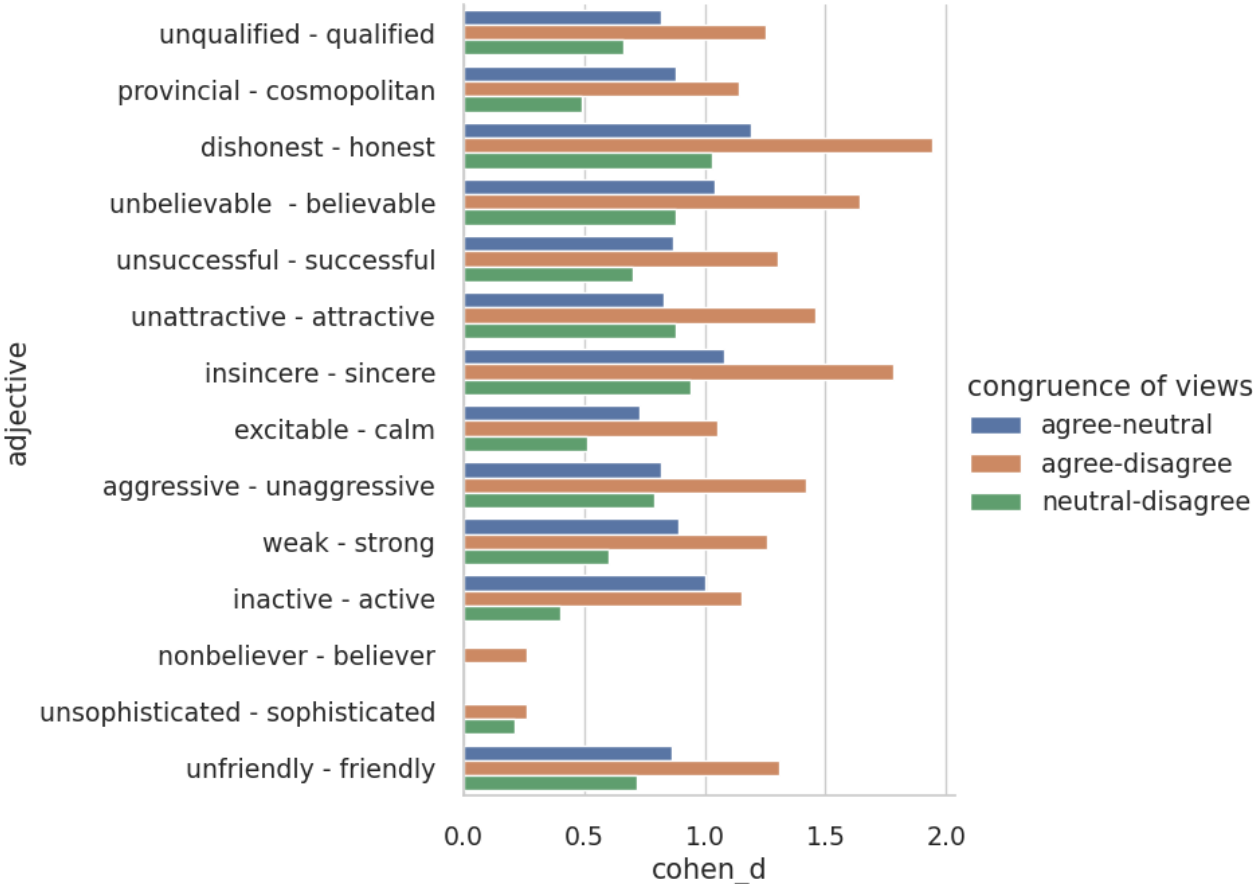


Fig. 4. Differences in semantic differential ratings in the pretest between groups with different congruence of political views with the selected politician

The effect of the interaction between stimulus type and participants’ congruence of political views on the change in ratings (pretest ratings subtracted from posttest ratings) was statistically significant in the case of one dimension (unsophisticated–sophisticated: $F = 2.7$, $p = 0.03$) according to a two-way ANOVA design (stimulus and congruence of political views). The post hoc Tukey’s HSD test, however, showed the results for group comparisons on the unsophisticated–sophisticated dimension to be statistically insignificant.

Gender differences. We observed statistically significant differences in pretest judgements on the semantic differential scale between women and men using a one-way MANOVA, $F(14, 942) = 1.74, p = 0.04$. Group differences are tested with the t -test for independent samples. On average, women gave higher responses on all six of the dimensions in question: unqualified–qualified ($t = 1.97, p = 0.048, d = 0.13$), dishonest–honest ($t = 2.01, p = 0.044, d = 0.13$), unsuccessful–successful ($t = 1.99, p = 0.046, d = 0.13$), unattractive–attractive ($t = 2.41, p = 0.016, d = 0.16$), aggressive–unaggressive ($t = 2.74, p = 0.006, d = 0.18$), and unfriendly–friendly ($t = 3.76, p < 0.01, d = 0.24$).

We also find a statistically significant interaction effect with regard to differences on the semantic differential scale between the gender of participants and evaluated politicians in a pre-test, $F(42, 2789) = 10.89, p < 0.001$. Tests of group differences reveal that statistically significant results are obtained only for Robert Biedroń (left). On average, women evaluated that politician higher on the following dimensions: unqualified–qualified ($t = 3.38, p < 0.01, d = 0.38$), dishonest–honest ($t = 3.77, p < 0.01, d = 0.43$), unbelievable–believable ($t = 2.48, p = 0.013, d = 0.28$), unsuccessful–successful ($t = 3.26, p < 0.01, d = 0.37$), unattractive–attractive ($t = 3.86, p < 0.01, d = 0.44$), insincere–sincere ($t = 2.81, p < 0.01, d = 0.32$), aggressive–unaggressive ($t = 2.05, p = 0.04, d = 0.23$), weak–strong ($t = 2.89, p < 0.01, d = 0.33$), inactive–active ($t = 2.77, p < 0.01, d = 0.31$), and unfriendly–friendly ($t = 3.78, p < 0.01, d = 0.43$). Women, therefore, rated Biedroń more positively in the pretest than men did. No statistically significant results between women and men were obtained for the other two politicians.

Differences between genders in evaluation of politicians on the semantic differential scale were also found in the posttest analyses, $F(42, 2789) = 8.95, p < 0.001$. Following a positive stimulus, women viewed politicians as more qualified ($t = 2.69, p < 0.01, d = 0.31$), cosmopolitan ($t = 3.1, p < 0.01, d = 0.35$), believable ($t = 2.55, p = 0.01, d = 0.29$), successful ($t = 2.0, p = 0.046, d = 0.23$), attractive ($t = 2.3, p = 0.02, d = 0.26$), calm ($t = 2.07, p = 0.039,$

$d = 0.23$), believing ($t = 2.25, p = 0.025, d = 0.25$), and friendly ($t = 2.76, p < 0.01, d = 0.31$). Following a negative stimulus, no differences were found between the genders.

Analyses performed for each politician individually revealed that women viewed Krzysztof Bosak (right) as more friendly ($t = 2.27, p = 0.025, d = 0.44$) after reading a positive tweet compared with men under the same condition. They also rated Robert Biedroń (left) as more qualified ($t = 2.01, p = 0.046, d = 0.4$), attractive ($t = 2.28, p = 0.025, d = 0.46$), sincere ($t = 2.66, p < 0.01, d = 0.53$), believing ($t = 2.33, p = 0.021, d = 0.47$) and friendly ($t = 2.01, p = 0.047, d = 0.4$) compared with men in the posttest. Rafał Trzaskowski was rated as more cosmopolitan ($t = 2.48, p = 0.015, d = 0.48$) by women than men after reading a positive tweet. No statistically significant results between women and men in the posttest with a negative stimulus were obtained for Bosak (right), or Trzaskowski (center). One significant difference was obtained for Biedroń (left): men viewed him as less honest ($t = 2.06, p = 0.042, d = 0.39$) than women did in the posttest.

The influence of stimulus type and participants' gender on the change in ratings (pretest ratings subtracted from posttest ratings) was tested using two-way ANOVA (stimulus and gender as factors). No statistically significant results were found. We may conclude that the influence of the stimulus on the change in ratings given by participants in the posttest compared with the pretest is the same in the case of women and men.

Age. The association between ratings on the semantic differential scale and participants' age were tested using Spearman's correlation coefficient as implemented in the statsmodels Python library. Statistically significant results were found for unqualified–qualified (a weak negative correlation) and inactive–active (a weak positive correlation) in the pretest. In the posttest analyses, several statistically significant coefficients were obtained: seven positive correlations under the positive stimulus condition, and one negative and one positive correlation in the

posttest with a negative stimulus. No statistically significant coefficients were recorded in the posttest with a neutral stimulus. Detailed results are given in the appendix.

Factor analysis. Lastly, we conduct exploratory factor analysis in the search of key dimensions of the politicians' image that change after exposition to experimental stimuli. We tested several variations with a different number of factors ($n = 3, 4, 5, 6, 7$), with and without a rotation (varimax, promax, quartimax, oblimin). **Data entered into the factor analysis model involves all variables from pretest and posttest. We use a dedicated *FactorAnalyzer* python library to conduct factor analysis.** The best suited model groups items into 3 factors using a promax rotation (similar results are also achieved with the varimax rotation) and MINRES fitting method for all politicians. Factor 1 ("leader's abilities") comprises qualified, cosmopolitan, believable, successful, attractive, and strong items, all with loadings between 0.52 and 0.87. Factor 2 ("morality"): items honest, sincere, and active with loadings between 0.65 and 0.71. Factor 3 ("sociability"): items calm, unaggressive, and friendly with loadings around 0.60. Because of high percentage of variability unexplained by the three factors ($>70\%$) items believer and sophisticated were discarded. Similar results from factor analysis were obtained by Cwalina, Falkowski and Kaid (2005) who distinguished a two-factor model ("leader's abilities" and "sociability/morality") on a Polish and American sample. **Running a MANOVA multivariate linear model with stimulus type and politician as independent variables, we observe a statistically significant interaction effect between both variables ($F(12, 2503) = 2.09, p=0.02$).** Regarding Factor 1, significant changes between stimuli conditions only for Robert Biedroń - between negative vs. neutral conditions ($t = -3.47, p=0.002$), and negative vs. positive conditions ($t = -3.01, p=0.008$). The difference in Factor 2 between stimuli conditions is again found only for Robert Biedroń (negative vs. neutral ($t = -2.90, p=0.01$), and negative vs. positive conditions ($t = -3.19, p=0.004$). Differences in Factor 3 are significant for Robert Biedroń and Krzysztof Bosak – for both politicians between negative vs. neutral ($t = -5.71,$

$p < 0.001$, and $t = -3.24$, $p = 0.004$), and negative vs. positive conditions ($t = -5.41$, $p < 0.001$, and $t = -4.59$, $p < 0.001$). Results of factor analysis indicate that the perception of the leader's abilities and morality is rather stable. The perception of a politician's sociability and friendliness may, however, change depending on the language they use in their media appearances. Although the use of factor analysis is a suitable approach, employed by many scientists, it may result in losing much of their application value (Cwalina, Falkowski & Kaid, 2000). Factor analysis can be useful for "preparing an election campaign, because they suggest how to control, to some extent, the image of a candidate. But the hints on what to remove from the image and what to add to it are somewhat general and equivocal." (Cwalina, Falkowski & Kaid, 2000, p. 122). Therefore, we conduct factor analysis as a supplementary analysis rather than the central one, following previous studies on this topic (Cwalina, Falkowski & Kaid, 2000).

Discussion

The results confirmed the significance of the emotional marking of an utterance in terms of social evaluations. For both negative and positive messages, a change in evaluations of a politician's image was observed in a direction consistent with the affective valence of the message. The change applied to evaluations of the politicians' morality, their leader's abilities, and their attitudes to other people (sociability). In the control conditions, using neutral stimuli, changes were recorded on only two dimensions, with small effect sizes. This result is in accordance with the assumptions of the Affect Infusion Model (AIM) proposed by Forgas (2002), which indicate that words with strong emotional valence often activate evaluative reactions in a rapid, automatic way. In our study, the influence of affect priming proved stronger in the case of positive stimuli; this is in contradiction to the theory of negativity bias (Baumeister et al., 2001; Holleman et al., 2021; Rozin et al., 2010), which suggests that automatic emotion activation is stronger in response to negative words than to positive words.

Negative words are generally more marked and show a more extreme deviation from neutrality in their semantic meanings than their positive counterparts. The “positivity bias” obtained in our study may result from the dominance in public political debate in Poland of a negative and aggressive form of communication, which means that positively marked political tweets featuring kindness speech are viewed as atypical, and thus more distinctive and noticeable. This observation is in line with the research of Skytte (2021), who found that the cultural communication of political opponents is something surprising and unique to citizens, and therefore influences the attitudes of respondents more than the style of discourse to which they are accustomed. In other words, people assume that politicians behave aggressively, so when they see something different, they are pleasantly surprised.

The interplay between positivity bias and participants' expectations is complex. Political communication often reflects a “Pollyanna effect” (Boucher & Osgood, 1969), where positive emotional appeals dominate. This is visible in Konat et al. (2024), who found that both right- and left-wing politicians heavily use emotional language, much of it positive, which aligns with trends in computational linguistics and self-reported preference scales. Positivity bias is also present in self-reports, as in our own study (e.g., xxx., submitted) where participants claim to prefer positive messaging. Research using the semantic differential method is also subject to the risk of this effect (Boucher & Osgood, 1969). However, this evidence from language use and self-reported scales contrasts with evidence from online studies on the spread of hate speech, as described in the introduction of this paper. Studies on hate speech online, including our own (xxx, in preparation), show that negative messages receive more engagement. Similar findings can be seen in voting patterns and the amplification of negative content online. This highlights a fundamental limitation of self-reported data: preferences often diverge from behavior, a problem as old as La Pierre's observation of differences between declarations and actions (1934). While we cannot fully resolve this discrepancy within the scope of this paper, it remains a crucial area for further research.

The results also confirmed the significance of the congruence of participants' political views for their evaluations of politicians' images. Participants who agree with the political views of a particular politician tend to rate that politician's image more positively than do neutral/undecided participants and those with opposite political views. Similarly, neutral/undecided participants rate the image more positively than do opponents of the politician. Interestingly, it was found that the positivity of the supporters was greater than the negativity of opponents in evaluating the image of a politician; this may result from the greater degree of engagement of the first group, or may be due to the "Pollyanna effect".

There were also observed weak correlations between participants' age and their evaluations of politicians, and women were found to make more favorable assessments than men, particularly in the case of the left-wing politician Robert Biedroń. A significant factor here may be the fact that Biedroń is a politician who strongly affirms his homosexual identity, and many studies have shown (cf. Kite et al., 2021; Vieira de Figueiredo & Pereira, 2021) that the attitudes of heterosexual men toward gay men are more negative than those of heterosexual women. Other studies (e.g. Peresie, 2004) show that women tend to be softer in their judgements than men, which may be due to their greater agreeableness and different patterns of socialisation. However, this hypothesis would require further study.

In terms of limitations, this study's focus on a single issue limits generalizability across different topics, such as climate change or vaccination. Further research should extend this investigation cross-platform and cross-issue to fully distinguish the impact of hate and kindness speech across diverse settings. We acknowledge that platforms vary in content regulation and user culture, which may influence the nature of interactions and outcomes. For example, Theocharis et al. (2024) found that Twitter, compared to other platforms, is associated with lower conspiracy belief rates, suggesting a different user experience. Facebook, on the other hand, has faced criticism for delayed enforcement against hate speech and harassment, though recent changes aim to address these issues (Dubois & Reepschlager, 2024). While limited in

scope, our study contributes by demonstrating that the impacts of affective language are identifiable even with numerous confounding factors common in social psychology. In future research, it would be worthwhile to include additional variables such as gender of politicians, their ethnicity, age or cultural context, as these may serve as mediators of differences in preferences towards politicians and the evaluation of their statements. Another limitation of this study is the reliance on fabricated posts instead of real tweets. This decision, while addressing certain methodological challenges, may influence the perceived authenticity of the posts and thus the participants' responses.

Despite these limitations, the results of our study may serve as important input to the discussion on the consequences of the language used in public debates. Emotionally marked statements may improve or worsen social evaluations of the person who makes them, a fact which is particularly significant in the case of persons holding high government positions. The “positivity bias” identified in the study provides encouragement to use kindness speech in public discourse, showing that this type of communication can improve a politician's ratings. This result is consistent with the ‘Montagu Principle’: people generally evaluate civil people more favorably than uncivil people. This is confirmed by Frimer and Skitka's (2018) study, which included real-life exchanges between President Trump and his opponents, as well as a speech by a fictional politician. Civility helped or did not affect—but never harmed—the reputation of the speaker, supporting the Montagu Principle. Even self-identified supporters of President Trump evaluated the president more favorably after he responded with civility to a personal attack. Incivility made the speaker seem less warm and did less to affect perceptions of dominance or honesty. It is therefore recommended that efforts be made to promote civility in political debates.

Acknowledgements

This research was funded in part by National Science Centre, grant number 2020/39/D/HS1/00488 awarded to Barbara Konat, and in part by Faculty of Psychology and Cognitive Sciences, Adam Mickiewicz University in Poznań internal research grant awarded to Monika Obrębska.

References

- Allen, J. J., Anderson, C. A., & Bushman, B. J. (2018). The General Aggression Model. *Current Opinion in Psychology*, *19*, 75–80. <https://doi.org/10.1016/j.copsyc.2017.03.034>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. <https://doi:10.1037/1089-2680.5.4.323>
- Beyerlein, K. & Klocek, J. (2020). How Do Religion and Sexual Orientation Affect Support for U.S. Presidential Candidates? Evidence from a Survey Experiment. *Journal for the Scientific Study of Religion*, *59*(4), 551-568. <http://dx.doi.org/10.1111/jssr.12690>
- Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, *41*(S1), 3–33. <https://doi.org/10.1111/pops.12670>
- Boucher, J., Osgood, C.E. (1969). The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior*, *8*(1), 1–8. [https://doi.org/10.1016/S0022-5371\(69\)80002-2](https://doi.org/10.1016/S0022-5371(69)80002-2)
- Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: Evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of Personality and Social Psychology*, *89*(6), 884–898. <https://doi.org/10.1037/0022-3514.89.6.884>

- Carraro, L., & Castelli, L. (2010). The implicit and explicit effects of negative political campaigns: Is the source really blamed? *Political Psychology*, 31(4), 617–645. <https://doi.org/10.1111/j.1467-9221.2010.00771.x>
- Cavazza, N. (2017). The tone dilemma: comparing effects of flattery and verbal aggression in a political speech. *Journal of Language and Social Psychology*, 36(5), 585–598. <https://doi.org/10.1177/0261927X17698186>
- Cervone, C., Augoustinos, M., & Maass, A. (2021). The Language of Derogation and Hate: Functions, Consequences, and Reappropriation. *Journal of Language and Social Psychology*, 40(1), 80–101. <https://doi.org/10.1177/0261927X20967394>
- Cotney, J.L., & Banerjee, R. (2019). Adolescents' Conceptualizations of Kindness and its Links with Well-being: A Focus Group Study. *Journal of Social & Personal Relationships*.36(2), 599–617.
- Cregg, D.R., & Cheavens, J.S. (2022). Healing through helping: An experimental investigation of kindness, social activities, and reappraisal as well-being interventions. *The Journal of Positive Psychology*, Dec 14, 2022, <http://dx.doi.org/10.1080/17439760.2022.2154695>
- Cwalina, W., Falkowski, A., & Kaid, L. L. (2000). Role of advertising in forming the image of politicians: Comparative analysis of Poland, France, and Germany. *Media Psychology*, 2, 119-146. https://doi.org/10.1207/S1532785XMEP0202_2
- Cwalina, W., Falkowski, A., & Kaid, L. L. (2005). Advertising and the image of politicians in evolving and established democracies: Comparative study of the Polish and the US presidential elections in 2000. *Journal of Political Marketing*, 4(2-3), 19-44. https://doi.org/10.1300/J199v04n02_02
- Cwalina, W., Falkowski, A., & Newman, B.I. (2015). *Political marketing: theoretical and strategic foundations*. Routledge.

- Cwalina, W. & Koniak, P. (2023). Obrażliwy język w komunikacji polityków i jego konsekwencje na poziomie indywidualnym, grupowym i społecznym [Offensive language in political communication and its consequences at individual, group and societal levels]. In: E. Eliaż & K. Skarżyńska (eds.). Nieufność: źródła i konsekwencje [Distrust: sources and consequences] (253-284). Instytut Problemów Współczesnej Cywilizacji im. Marka Dietricha.
- Dobrowolska, M., & Obrębska, M. (2023). Use of hate speech and social evaluation of a politician's image. *Psychology of Language and Communication*, 27(1), 256-277. <https://doi.org/10.58734/plc-2023-0013>
- Druckman, J.N., Gubitz, S.R., Levendusky, M.S. & Lloyd, A.M. (2019). How incivility on partisan media (de)polarizes the electorate. *Journal of Politics*, 81(1), 291–295. <https://psycnet.apa.org/doi/10.1086/699912>
- Dubois, E., & Reepschlager, A. (2024). How harassment and hate speech policies have changed over time: Comparing Facebook, Twitter, and Reddit (2005–2020). *Policy & Internet*. <https://doi.org/10.1002/poi3.387>
- Farkas, X., & Bene, M. (2021). Images, Politicians, and Social Media: Patterns and Effects of Politicians' Image-Based Political Communication Strategies on Social Media. *International Journal of Press/Politics*, 26(1), 119-142. <https://doi.org/10.1177/1940161220959553>
- Fasoli, F., Paladino, M. P., Carnaghi, A., Jetten, J., Bastian, B., & Bain, P. G. (2015). Not 'just words': Exposure to homophobic epithets leads to dehumanizing and physical distancing from gay men. *European Journal of Social Psychology*, 46(2), 237–248. <https://doi.org/doi/10.1002/ejsp.2148>
- Fehrenbacher, D.D. (2017). Affect Infusion and Detection through Faces in Computer-mediated Knowledge-sharing Decisions. *Journal of the Association for Information Systems*, 18(10), 703-726.

- Frimer, J. A., & Skitka, L. J. (2018). The Montagu Principle: Incivility decreases politicians' public approval, even with their political base. *Journal of Personality and Social Psychology*, 115(5), 845–866. <https://doi.org/10.1037/pspi0000140>
- Flannelly, K.J. (2002). Voting for Female Candidates: Effects of Voters' Age, Ethnicity, and Gender. *The Journal of Social Psychology*, 142(3), 397-399.
- Forgas, J.P. (2001). Affect and the “Social Mind“: Affective Influences on Strategic Interpersonal Behaviors (46-72). In: J.P. Forgas, K.D. Williams, & L. Wheeler (ed.). *The Social Mind. Cognitive and Motivational Aspects of Interpersonal Behavior*. Cambridge University Press.
- Forgas, J. P. (2002). Toward Understanding the Role of Affect in Social Thinking and Behavior. *Psychological Inquiry*, 13(1), 90–102. https://doi.org/10.1207/S15327965PLI1301_03
- Gajewska, E., & Konat, B. (in preparation). Influencing hearts: Investigating conversational dynamics of emotional appeals in online discussions.
- Gajewska, E., & Konat, B. (2023). Text Classification for Subjective Phenomena on Disaggregated Data and Rater Behaviour (73-78). In Z. Vetulani, & P. Paroubek (eds.). *Human Language Technologies as a Challenge for Computer Science and Linguistics*. Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza. <https://doi.org/10.14746/amup.9788323241775>
- Gawronski, B. & Walther, E. (2008). The TAR effect: When the ones who dislike become the ones who are disliked. *Personality and Social Psychology Bulletin*, 34(9), 1276–1289. <https://doi.org/10.1177/0146167208318952>
- Grofman, B. (2005). Comparisons among electoral systems: Distinguishing between localism and candidate-centered politics. *Electoral Studies*, 24(4), 735-740. <https://doi.org/10.1016/j.electstud.2005.03.007>
- Holleman, B., Kamoen, N., & Struiksmā, M. (2021). The affect of negativity: testing the Foreign Language Effect in three types of valence framing and a moral dilemma.

Cognition and Emotion, 35(4), 690-704.

<https://doi.org/10.1080/02699931.2021.1889472>

Imbir, K. K. (2016). Affective norms for 4900 Polish words reload (ANPW_R): Assessments for valence, arousal, dominance, origin, significance, concreteness, imageability and age of acquisition. *Frontiers in psychology*, 7, 174568. <https://doi.org/10.3389/fpsyg.2016.01081>

Kite, M. E., Whitley, B.E., Buxton, K., & Ballas, H. (2021). Gender differences in anti-gay prejudice: Evidence for stability and change. *Sex Roles: A Journal of Research*, 85(11-12), 721-750. <http://dx.doi.org/10.1007/s11199-021-01227-4>

Konat, B., Gajewska, E., & Rossa, W. (2024). Pathos in natural language argumentation: Emotional appeals and reactions. *Argumentation*, 38(2), 369–403.

Konat, B., Obrębska, M., Gajewska, E., & Dembska, N. (submitted). Do emotional appeals make arguments seem stronger? A psycholinguistic experiment on pathos and audience involvement.

Leets, L., & Giles H. (1999). Harmful speech in intergroup encounters: An organizational framework for communication research. *Annals of the International Communication Association*, 22(1), 91–137. <https://doi.org/10.1080/23808985.1999.11678960>

Ling, A., George, N., Shiv, B., & Plassmann, H. (2023). Altering experienced utility by incidental affect: The interplay of valence and arousal in incidental affect infusion processes. *Emotion*, 23(8), 2179-2193. <https://doi.org/10.1037/emo0001241>

McLaughlin B, Thompson BA. (2016). Conditioned by Race: How Race and Religion Intersect to Affect Candidate Evaluations. *Politics and Religion*. 9(3), 605-629. doi:10.1017/S1755048316000213

Mullen, B., Smyth, J.M. (2004). Immigrant suicide rates as a function of ethnophobias: Hate speech predicts death. *Psychosomatic Medicine*, 66, 343–348. <https://psycnet.apa.org/doi/10.1097/01.psy.0000126197.59447.b3>

- Ntontis, E., Jurstakova, K., Neville, F., Haslam, S. A., & Reicher, S. (2024). A warrant for violence? An analysis of Donald Trump's speech before the US Capitol attack. *The British Journal of Social Psychology*, 63(1), 3–19. <https://doi.org/10.1111/bjso.12679>
- Obreńska, M. (2020). Contempt Speech and Hate Speech: Characteristics, Determinants and Consequences. *Annales Universitatis Mariae Curie-Skłodowska*, 33(3), 9-20, <http://dx.doi.org/10.17951/j.2020.33.3.9-20>
- Ott, B. L. (2017). The age of Twitter: Donald J. Trump and the politics of debasement. *Critical Studies in Media Communication*, 34(1), 59–68. <https://doi.org/10.1080/15295036.2016.1266686>
- Papcunová, J., Martončík, M., Fedáková, D., Kentoš, M., & Adamkovič, M. (2023). Perception of Hate Speech by the Public and Experts: Insights into Predictors of the Perceived Hate Speech Towards Migrants. *Cyberpsychology, Behavior, and Social Networking*, 26(7), 489-498. <http://dx.doi.org/10.1089/cyber.2022.0191>.
- Peresie, J. L. (2004). Female judges matter: Gender and collegial decisionmaking in the federal appellate courts. *The Yale Law Journal*, 114(7), 1759.
- Pérez-Fernández, H., Delgado-García, J.B., Martín-Cruz, N., & Rodríguez-Escudero, A.I.(2022). The Role of Affect in the Development of Entrepreneurial Intentions. *Entrepreneurship Research Journal*, 12(2), 19. <https://doi.org/10.1515/erj-2019-0124>
- LaPiere, R. T. (1934). *Attitudes vs. Actions*. Social Forces.
- Riegel, M., Wierzba, M., Wypych, M., Żurawski, Ł., Jednoróg, K., Grabowska, A., & Marchewka, A. (2015). Nencki affective word list (NAWL): the cultural adaptation of the Berlin affective word list–reloaded (BAWL-R) for Polish. *Behavior Research Methods*, 47, 1222-1236. <https://doi.org/10.3758/s13428-014-0552-1>
- Rosenblum, M., Schroeder, J., & Gino, F. (2019). Tell it like it is: When politically incorrect language promotes authenticity. *Journal of Personality and Social Psychology*, 119(1), 75–103. <https://doi.org/10.1037/pspi0000206>

- Rossini, P., Sturm-Wikerson, H., & Johnson, T. J. (2020). A wall of incivility? Public discourse and immigration in the 2016 U.S. Primaries. *Journal of Information Technology & Politics*, 0(0), 1–15. <https://doi.org/10.1080/19331681.2020.1858218> [
- Rozin, P., Berman, R., & Royzman, E. (2010). Biases in use of positive and negative words across twenty natural languages. *Cognition and Emotion*, 24(3), 536–548. <https://doi.org/10.1080/02699930902793462>
- Salimi, E. A., & Mortazavi, S. M. (2024). Impoliteness in Twitter Discourse: a Case Study of Replies to Donald Trump and Greta Thunberg. *Baltic Journal of English Language, Literature and Culture*, 14, 86–10. <https://doi.org/10.22364/BJELLC.14.2024.06>
- Shin, L.J., Layous, K., Choi, I., Na, S., & Lyubomirsky, S. (2020). Good for self or good for others? The well-being benefits of kindness in two cultures depend on how the kindness is framed. *Journal of Positive Psychology*, 15(6), 795–805. <https://doi.org/10.1080/17439760.2019.1651894>
- Skytte, R. (2021). Dimensions of elite partisan polarization: Disentangling the effects of incivility and issue polarization. *British Journal of Political Science*, 51(4), 1457–1475. <http://dx.doi.org/10.1017/S0007123419000760>
- Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior*, 44, 136–146. <https://doi.org/10.1002/ab.21737>
- Schultz, C., Pancer, S.M. (1997). Character attacks and their effects on perceptions of male and female political candidates. *Political Psychology*, 18(1), 93–102. <http://dx.doi.org/10.1111/0162-895X.00046>
- Swim, J. K., Johnston, K., & Pearson, N. B. (2009). Daily experiences with heterosexism: Relations between heterosexist hassles and psychological well-being. *Journal of Social and Clinical Psychology*, 28(5), 597–629. <https://doi.org/10.1521/jscp.2009.28.5.597>

- Theocharis, Y., Cardenal, A., & Štětka, V. (2024). Does the platform matter? Social media and COVID-19 conspiracy theory beliefs in 17 countries. *New Media & Society*, 25(12). <https://doi.org/10.1177/14614448211045666>
- Van't Riet, J., & Van Stekelenburg, A. (2022). The effects of political incivility on political trust and political participation: A meta-analysis of experimental research. *Human Communication Research*, 48(2), 203–229. <https://doi.org/10.1093/hcr/hqab022>
- Vieira de Figueiredo, C., & Pereira, C. R. (2021). The Effect of Gender and Male Distinctiveness Threat on Prejudice Against Homosexuals. *Journal of Personality and Social Psychology*, 121(6), 1241-1257. <https://doi.org/10.1037/pspi0000269>
- Wattenberg, M. P. (1991). *The rise of candidate-centered politics*. Harvard University Press
- Wierzba, M., Riegel, M., Wypych, M., Jednoróg, K., Turnau, P., Grabowska, A., & Marchewka, A. (2015). Basic emotions in the Nencki Affective Word List (NAWL BE): New method of classifying emotional stimuli. *PLOS ONE*. <https://doi.org/10.1371/journal.pone.0132305>
- Wierzba, M., Riegel, M., Kocoń, J., Miłkowski, P., Janz, A., Klessa, K., Juszczyk, K., Konat, B., Grimling, D., Piasecki, M. & Marchewka, A. (2021). Emotion norms for 6000 polish word meanings with a direct mapping to the polish wordnet. *Behavior Research Methods*, 1-16. <https://doi.org/10.3758/s13428-021-01697-0>
- Williams, P.B., Poljacik, G., Decety, J., & Nusbaum, H.C. (2018). Loving-kindness language exposure leads to changes in sensitivity to imagined pain. *The Journal of Positive Psychology*, 4(13), 429–433. <https://doi.org/10.1080/17439760.2017.1315648>
- Winiewski, M., Hansen, K., Bilewicz, M., Soral, W., Świdorska, A., & Bulska, D. (2017). *Mowa nienawiści, mowa pogardy. Raport z badania przemocy werbalnej wobec grup mniejszościowych* [Contempt speech, hate speech. Report from research on verbal violence against minority groups]. Wydawnictwo Sowa.

- Winkielman, P., Knutson, B., Paulus, M., & Trujillo, J. L. (2007). Affective influence on judgments and decisions: Moving towards core mechanisms. *Review of General Psychology, 11* (2), 179–192. <https://doi:10.1037/1089-2680.11.2.179>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. (2020, October). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 38-45). Association for Computational Linguistics. Online. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>
- Wypych, M. & Bilewicz, M. (2024). Psychological toll of hate speech: The role of acculturation stress in the effects of exposure to ethnic slurs on mental health among Ukrainian immigrants in Poland. *Cultural Diversity and Ethnic Minority Psychology, 30*(1), 35-44. <https://doi.org/10.1037/cdp0000522>