Journal of Behavioral and Experimental Economics 110 (2024) 102203

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Diffusion of tax-related communication on social media

Žiga Puklavec[a,*], Olga Stavrova[a,b], Christoph Kogler[a], Marcel Zeelenberg[a,c]

[a] *Department of Social Psychology, and Tilburg Institute for Behavioral Economics Research, (TIBER), Tilburg University, the Netherlands*
[b] *Department of Psychology, University of Lübeck, Germany*
[c] *Department of Marketing, Vrije Universiteit Amsterdam, the Netherlands*

ARTICLE INFO

Keywords:
Taxation
Information diffusion
Social media
Emotions
Moral language

ABSTRACT

Taxation is a recurrent topic in people's conversations, also on social media. Yet, informal channels such as social media have been widely neglected in studies that examined how information about taxation spreads across social networks. Using posts on Twitter (currently called "X") with taxation related hashtags from 2010 to 2020, we examined what linguistic cues are associated with information diffusion, that is, the number of retweets a message receives. The use of emotional, moral, and moral-emotional language in a tweet was associated with greater diffusion (i.e., more retweets). In contrast to the negativity bias literature, positive emotional words were more strongly associated with information diffusion than negative emotional words. Among the specific emotions that taxation research has focused on, only the use of anger (but not anxiety) words was associated with more retweets. The study contributes to the literature by examining individuals' reasoning about taxes.

1. Diffusion of tax-related communication on social media

Taxation is crucial for any functioning society, as it contributes towards public goods and finances governmental services. Yet, attitudes towards taxation vary widely (Braithwaite, 2003; Kirchler, 1998). Not surprisingly, taxation is a recurrent topic in people's conversations, ranging from discussions about tax rates to even questioning the sole purpose of the taxation system. Such conversations provide important information about other people's views about taxes, and as demonstrated by research on social norms, shape people's attitudes and behavior (e.g., Cialdini et al., 1990), including tax compliance decisions (Wenzel, 2005a, 2005b).

Research found that people communicate about taxes with others (e.g., Onu & Oats, 2016). Hence, information (including sentiments and behaviors) spreads across individuals' social networks – a phenomenon referred to as "network effects".[1] For example, neighborhood communication effectively diffuses information on enforcement, ultimately resulting in increased tax compliance (Drago et al., 2020; Rincke & Traxler, 2009). Network effects may also operate via professional tax accountants, who share information about the procedures of the tax administration obtained from one client, with other clients (Boning et al., 2020). Information about taxes is shared among family members and friends as well. For example, when individuals adopt legal tax avoidance strategies, their family members are more likely to do so as well (Alstadsæter et al., 2019).

While the network effects literature focused on information diffusion across individuals in close contact or geographical proximity, little is known about how the information about taxes is shared online. Importantly, people recurrently talk about taxes on social media, such as Twitter. With over 4.2 billion users in 2021 (DataReportal, 2021), social media are becoming an increasingly important platform for sharing news and information, including taxes (e.g., groups for tax-related questions on Facebook, discussion threads on Reddit). For example, since 2010, over 5 million tweets under tax-related hashtags have been posted on Twitter (called "X" since 2023). This makes social media platforms a great tool to study network effects for taxation related topics (see, e.g., Puklavec et al., 2023).

We examined what makes a message about taxes more or less likely to spread across users' networks on one of the most popular social media platforms: Twitter.[2] We focused on the linguistic features of the message. Previous taxation research suggests that linguistic analysis might be useful in categorizing types of online interactions about taxes (Onu & Oats, 2016, 2018) and getting insight about national tax climates (Lozza & Castiglioni, 2018). Yet, no study thus far looked at the effect of

* Corresponding author at: Department of Social Psychology, Tilburg University, PO Box 90153, 5000 LE, Tilburg, the Netherlands.
E-mail address: Z.Puklavec@tilburguniversity.edu (Ž. Puklavec).

[1] Note that sometimes the effect of taxation information on tax paying behavior is referred to as "network effect" as well.
[2] https://gs.statcounter.com/social-media-stats

https://doi.org/10.1016/j.socec.2024.102203
Received 1 June 2023; Received in revised form 9 February 2024; Accepted 16 March 2024
Available online 18 March 2024

language use on information diffusion of tax-related content via social media.

We have focused on the effect of moral and emotional language. The role of moral and emotional language has been discussed as a potential accelerator of the spread of socio-political information before. Yet, research failed to reach consensus, with some studies suggesting that these linguistic features can both contribute to or undermine diffusion (e.g., Brady et al., 2017; Burton et al., 2019). Using a dataset of tweets posted under tax-related hashtags within the last 10 years, we found that tweets with more moral and emotional words were more likely to get shared (i.e., retweeted). These findings therefore contribute to the discussion regarding the role of moral and emotional language in information diffusion (Brady et al., 2017; Burton et al., 2019). Comparing these findings to previous studies using similar methodology reveals that the effects of the linguistic cues are similar in size to the effects reported in other studies (Brady et al., 2017, 2019; Burton et al., 2019). Interestingly, diffusion patterns (i.e., hindering or promoting diffusion) associated with emotional, moral, and moral-emotional language in taxation-related tweets differ compared to other topics (e.g., U.S political issues, MeToo movement, Brexit, etc. – as seen in Burton et al., 2019).

Furthermore, these findings provide a novel take on network effects in the taxation literature. Network effects describe information diffusion across social networks, for example, in response to governmental campaigns. While previous studies showed that information spreads via social networks (e.g., Castro & Scartascini, 2019), the present study addresses the question of what makes taxation-specific information spread, by demonstrating that the use of certain linguistic features in a message is associated with a higher probability of the message being shared with others. Considering how common online information exchange is nowadays, this is an important addition to the field of taxation research. Finally, our findings are also particularly interesting in light of the scarce literature on the use of language in online media related to taxation (Onu & Oats, 2016, 2018) and in print media (Lozza & Castiglioni, 2018).

1.1. Information diffusion on social media

Research on information spread on social media platforms suggests that diffusion can be affected by the linguistic features of the message. For example, negative emotional language in the Centers' for Disease Control and Prevention (CDC) tweets contributes to diffusion, while positive emotional language suppresses it (Zhu et al., 2020). This finding is consistent with the negativity bias–individuals giving more weight to negative, rather than positive experiences when making judgements (Rozin & Royzman, 2001). Berger and Milkman (2012) suggest that focusing solely on the valence of the message might be misleading, since information diffusion is also partly driven by specific emotions. Their analyses of various New York Times articles showed that some emotional words (e.g., awe, anger and anxiety) were associated with wider information diffusion (i.e., inclusion in mailing lists) than others (e.g., sadness). Furthermore, information diffusion was also more likely for content that was practically useful, interesting, and surprising.

When it comes to taxes, research suggests that some emotions might be more important than others. Enachescu et al. (2019) analyzed focus group discussions with taxpayers and tax auditors and found that among the most frequent emotions mentioned are anger and fear. This converges with surveys by the Internal Revenue Service of the United States (IRS) revealing that after being audited, self-employed taxpayers reported anger and fear (anxiety) (Erard et al., 2018). Thus, besides examining the valence of the emotional words, we additionally explored the role of two specific emotions: anger and anxiety.[3]

Besides emotional language, moral language has been associated with information diffusion as well. Specifically, messages containing both moral and moral-emotional language (i.e., language that reflects emotions linked to the welfare of other people; Haidt, 2003) are shared more often on social media (Brady et al., 2017). Moral-emotional language was associated with wider diffusion of messages in online discussions of polarizing political issues, such as gun control, same-sex marriage, or climate change (Brady et al., 2017). Also tweets from US politicians (Brady et al., 2019) and members of the US Congress (Wang & Inbar, 2021) containing more moral and moral-emotional language were associated with wider diffusion. Additional analyses, using a combination of dictionaries and distributed language models, revealed that this was especially true for negative moral language (Wang & Inbar, 2022). In contrast, for less polarizing topics (e.g., mathematics), the use of moral and emotional language might appear "unnecessary" and jeopardize rational debate, ultimately leading to less diffusion (Burton et al., 2019). Also, when looking at online news, moral language has been found to hinder consumption (Robertson et al., 2023). Given these different findings in prior research, we explored the role of moral and moral-emotional language in information diffusion about taxes.

1.2. The present research

The aim of our study was to identify whether the use of emotional, moral, and moral-emotional language in tweets with taxation-related hashtags is associated with more diffusion. In light of the literature on negativity bias (Rozin & Royzman, 2001), we additionally examined the role of valence by considering the effect of positive vs. negative emotional and moral-emotional language separately. Lastly, we tested whether specific emotions identified as important in existing taxation research (e.g., anger and anxiety) are related to information diffusion.

2. Method

2.1. Data collection

We gathered English language (independent of country of origin) tweets with the hashtags #tax, #taxes, and #taxation, posted from 2010 to 2020. The selected hashtags were meant to be as general as possible, so the data would not disproportionally include tweets from a specific country or timeframe, while also including the largest possible number of relevant tweets. We used a python-based scraper for social networking services (Snscrape; JustAnotherArchivist, 2021) and an R script in combination with Twitter's API to pull the tweet messages and associated metadata (e.g., retweet count, number of followers). The data were collected in January 2021. All data and corresponding scripts are available at https://osf.io/4ym7p/.

2.2. Data processing and dataset

After pulling the tweets, we first removed tweets with identical text.[4] Following the common practice in the field (e.g., Brady et al., 2017), we processed the remaining tweets in the following way: we removed any URL links, mentions of users (beginning with "@"), converted all letters to lower case, removed specific words (i.e., "rt" and "re"), English stop words, punctuation, numbers, and excessive whitespaces from the text corpus. Our final dataset consisted of 4,080,686 unique tweets from 560,778 different users that tweeted under the taxation hashtags from 2010 to 2020. The data included the text of the message, the retweet count, the follower count of the respective users, and information regarding whether the tweets included an URL link or other media (pictures, gifs, etc.).

[3] We rely on a dictionary-based approach where fear and anxiety are considered synonyms.

[4] Duplicated messages most probably occur due to bots which post the same message multiple times. In total, there were ~20% of duplicated messages.

2.3. Measures

2.3.1. Information diffusion

Following the established practice (e.g., Brady et al., 2017), as a measure of information diffusion, we used the retweet-count[5] of each tweet (mean = 0.48, variance = 71.48). The retweet count was heavily skewed (over-dispersed), containing a large number of tweets that got zero retweets (0 = 84 %). For the distribution see Fig. 1.

2.3.2. Emotional and moral language

To measure emotional and moral language we used a dictionary-based approach. That is, we compared the words in our text corpus to a dictionary defining a list of words that fall into specific categories (e.g., emotions or morality), returning the prevalence of the dictionary word occurring within each individual tweet. To identify moral language, we used the Moral Foundations Dictionary 2.0 (Firmer et al., 2019). This dictionary contains words that capture the content of the five moral foundations (care, fairness, loyalty, authority, and sanctity; Haidt & Graham, 2007; Haidt & Joseph, 2004). To measure emotional language, we used the Linguistic Inquiry and Word Count software (LIWC; Pennebaker et al., 2015). Specifically, we used the LIWC category for affect (to measure emotional language) and the categories for positive and negative emotions (to measure positive and negative emotional language). LIWC was also used to determine the frequencies of the specific emotions: anger and anxiety.

Following Brady et al. (2017), we also identified words that were part of both, moral foundations and affect dictionaries, to classify moral-emotional language. The words from the moral-emotional category were excluded from the other two categories (emotional and moral language). Overall, we computed the frequencies of three word categories: moral, emotional, and moral-emotional (i.e., words that were part of both emotional and moral dictionaries) words.

For further analyses, we split the emotional and moral-emotional categories by valence (positive vs. negative), resulting in four separate variables: positive emotional, negative emotional, positive moral-emotional, and negative moral-emotional language. Examples of words from all categories / dictionaries are shown in Table 1.

Before analyzing the data, we inspected the 20 most frequent words from each dictionary for face validity. We detected two words that could have a different meaning in the taxation context than the meaning intended by the dictionary. Specifically, the word "cut" in the negative emotions dictionary could be part of the expression "tax cut" and the word "avoid" in the anxiety dictionary could be part of the expression "tax avoidance". Given the ambivalence of the valence of these expressions, we classified all occurrences where the word stem "tax" and the word "cut" appear one after another (in both directions) and subtracted this score from the initial negative emotional category score. We applied the same procedure for "tax avoidance".

We also accounted for negations in the word counts for the valence-split emotional and moral-emotional language categories. Negations can invert the categorization of a specific word as either positive or negative. For example, the word "good" would be coded as positive, but not, if it was negated before (e.g., "not good"). The negation handling was done using the "sentimentr" R package by looking at negations within 5 words before and 2 words after the (moral-)emotional word (Rinker, 2021).

2.3.3. Control variables

Previous research showed that the follower count, and whether the tweet included a URL link or other media were associated with more information diffusion (Brady et al., 2019; Wang & Inbar, 2021). Therefore, we included those variables as controls in the analysis.

[5] The retweet-count does not include quote retweets.

2.4. Analytic strategy

The dependent variable (retweet count) was an over-dispersed count variable. Therefore, we ran a series of negative binomial models, while controlling for several control variables. We started by estimating the first model specification as following (Table 2):

$$Y_i \sim NB(\mu_i, \theta)$$

$$\mu_i = \exp(\beta_0 + \beta_1 \mathrm{M}_i + \beta_2 E_i + \beta_3 \mathrm{ME}_i + \gamma_1 X_i)$$

where Y_i is the retweet count of message *i*. M_i, E_i, and ME_i represent the count of moral, emotional, and moral-emotional words within message *i*, respectively. X_i represents a set of control variables for message *i*, specifically, the standardized follower count (achieved by subtracting the mean and dividing by the standard deviation of the sample) of the Twitter user that posted the message, as well as two dummy variables indicating the presence of either an URL link, or media content in the message. θ represents the overdispersion parameter of a quadratic parameterization, where the variance is given as $\mu(1 + \mu/\theta)$ (Brooks et al., 2017).

The second model specification split the emotional and moral language by valence (Table 3):

$$\mu_i = \exp(\beta_0 + \beta_1 \mathrm{M}_i + \beta_2 \mathrm{EP}_i + \beta_3 \mathrm{EN}_i + \beta_4 \mathrm{MEP}_i + \beta_5 \mathrm{MEN}_i + \gamma_1 X_i)$$

where M_i, EP_i, EN_i, MEP_i, and MEN_i represent the count of moral, positive emotional, negative emotional, positive moral-emotional, and negative moral-emotional words within message *i*, respectively. The set of control variables X_i remains the same.

Lastly, the third specification accounts for the specific emotions (Table 4):

$$\mu_i = \exp(\beta_0 + \beta_1 \mathrm{ANG}_i + \beta_2 \mathrm{ANX}_i + \gamma_1 X_i)$$

where ANG_i and ANX_i represent the count for anger and anxiety words within message *i*, respectively. The set of control variables X_i remains the same. We ran each of the three specifications in separate models, either excluding or including the control variables.

For the main analysis, we selected tweets from users that posted only once under the tax related hashtags, which accounted for ~64 % users. We did this because most users posted only once, and as tweets could come from any source that used the tax related hashtags, they may have also contained noise (e.g., advertisements for tax advice, tweets from bots). Previous findings support the idea that bots are often "spamming", thus having significantly higher amounts of tweets than human users (e.g., Anwar & Yaqub, 2020; Chavoshi et al., 2016).

To ensure the robustness of the results, we ran additional analyses with tweets from users with up to 15 tweets (that constituted 95 % of users), and with the full sample. As those datasets included users with multiple tweets per user, we ran mixed negative binomial models with random intercepts for users to control for the nested structure of the data (tweets nested within users). The results of these robustness checks support the main conclusions and are reported in the supplemental materials (Tables S2–S7).

To additionally control for potential influences of bots within our dataset, we looked at other possible indicators of bot activity. For example, bots try to reach the broadest possible audience by including many diverse hashtags (e.g., Grier et al., 2010). Another indication of bots is that they have a highly skewed follower-to-following ratio, meaning they follow a large number of users, but have few or no followers themselves (e.g., Chu et al., 2010). We conducted additional robustness checks where we excluded tweets with high amounts of hashtags (in the top 10th and 15th percentile of the sample) and excluded users with no followers. Results of hashtag-related (Tables S8–S13) and follower-related (Tables S14–S16) robustness checks further support the main conclusions and are reported in the supplemental materials. Lastly, due to the skewed distribution of

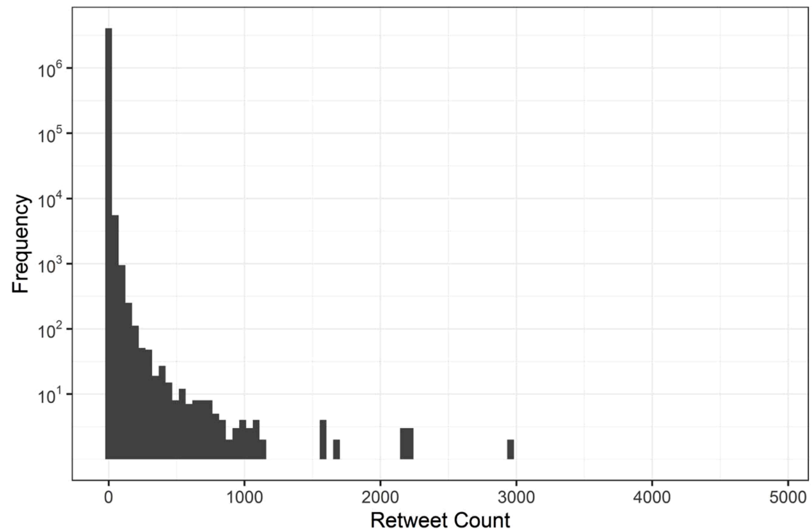

Fig. 1. Frequency of Retweets in the Tweet Dataset
Note. The y-axis is on an exponential scale, as otherwise the frequency of zero retweets would render the other retweet counts invisible. The x-axis is non-truncated, which means that there are single occurrences of retweet counts beyond 3000, yet not enough to be visible on the plot.

Table 1
Example words for each dictionary.

Dictionary	Valence	Example words
Moral		*betray, law, health, protect, guide, fraud, compliance, …*
Emotional	*Positive:*	*like, good, free, great, support, best, hope, profit, …*
	Negative:	*damn, hate, bad, shit, poor, problem, suck, kill, …*
Moral-emotional	*Positive:*	*help, love, care, share, fair, trust, safe, honest, …*
	Negative:	*fuck, hell, war, pain, stealing, lying, sin, thieves, …*
Anger		*hate, suck, shit, damn, bitch, stupid, mad, crap, …*
Anxiety		*stress, risk, fear, struggle, threat, pressure, dread, …*

Note. $N_{\text{Moral}} = 1{,}713$, $N_{\text{Emotional}} = 1{,}270$, $N_{\text{Moral-Emotional}} = 385$, $N_{\text{Anger}} = 230$, $N_{\text{Anxiety}} = 116$.

retweets (see Fig. 1) there are single cases which are extreme outliers. To check whether our results are robust to excluding these outliers, we re-ran the main analysis on a dataset excluding cases above the threshold of 1000 retweets (based on the distribution in Fig. 1). The results remained robust and a are reported in the supplemental materials (Tables S17–S19).

3. Results

Descriptive statistics and zero-order correlations among the predictor variables are reported in Table S-1 (see supplemental materials). Tweets that contain moral-emotional words are somewhat more likely to contain words from uniquely emotional and moral dictionaries, indicating that some tweets contain multiple word categories. Anger and anxiety are also strongly correlated with negative emotional words, as both categories are (almost entirely) part of the negative emotional category.

Table 2
Negative binomial regression for moral, emotional, and moral-emotional words explaining retweet count.

	Model 1		Model 2		Model 3	
Predictors	*IRR*	*95 % CI*	*IRR*	*95 % CI*	*IRR*	*95 % CI*
(Intercept)	0.36***	0.35 – 0.36	0.25***	0.24 – 0.25	0.20***	0.19 – 0.20
Moral Language	1.25***	1.22 – 1.27			1.23***	1.20 – 1.25
Emotional laguage	1.17***	1.15 – 1.18			1.16***	1.15 – 1.18
Moral-emotional language	1.10***	1.06 – 1.13			1.09***	1.06 – 1.12
Follower count			20.19***	18.11 – 22.50	18.52***	16.65 – 20.59
Media [1]			3.16***	3.06 – 3.26	3.16***	3.06 – 3.27
URL [1]			1.25***	1.22 – 1.28	1.29***	1.26 – 1.32
N_{tweets}	357,592		357,592		357,592	
AIC	456,637.539		439,951.062		438,499.721	
log-Likelihood	−228,313.770		−219,970.531		−219,241.860	

Note. *** $p < .001$. IRR = Incidence rate ratio. CI = Confidence interval. Moral language, emotional language, and moral-emotional language are counts representing the number of words of the respective category used in a tweet. The follower count has been standardized (z-transformed) before the analyses, so that the respective estimates reflect a one standard deviation change. Media: 1 = tweet contains media, 0 = tweet does not contain media. URL: 1 = tweet contains URL link, 0 = tweet does not contain URL link.

Table 3
Negative binomial regression for moral, emotional, and moral-emotional words (split by valence) explaining retweet count.

	Model 1		Model 2		Model 3	
Predictors	*IRR*	*95 % CI*	*IRR*	*95 % CI*	*IRR*	*95 % CI*
(Intercept)	0.36***	0.35 – 0.36	0.25***	0.24 – 0.25	0.20***	0.19 – 0.20
Moral language	1.27***	1.25 – 1.30			1.24***	1.22 – 1.27
Emotional language [Negative]	0.99	0.97 – 1.02			1.04***	1.01 – 1.06
Emotional language [Positive]	1.25***	1.23 – 1.27			1.22***	1.21 – 1.24
Moral-emotional language [Negative]	1.34***	1.26 – 1.42			1.13***	1.07 – 1.19
Moral-emotional language [Positive]	1.04 *	1.00 – 1.08			1.10***	1.06 – 1.13
Follower count			20.19***	18.11 – 22.50	18.94***	17.03 – 21.06
Media [1]			3.16***	3.06 – 3.26	3.11***	3.02 – 3.21
URL [1]			1.25***	1.22 – 1.28	1.28***	1.25 – 1.31
N_{tweets}	357,592		357,592		357,592	
AIC	456,384.326		439,951.062		438,361.059	
log-Likelihood	−228,185.163		−219,970.531		−219,170.530	

Note. * $p < .05$, *** $p < .001$. IRR = Incidence rate ratio. CI = Confidence interval. Moral language, emotional language (positive and negative), and moral-emotional language (positive and negative) are counts representing the number of words of the respective category used in a tweet. The follower count has been standardized (z-transformed) before the analyses, so that the respective estimates reflect a one standard deviation change. Media: 1 = tweet contains media, 0 = tweet does not contain media. URL: 1 = tweet contains URL link, 0 = tweet does not contain URL link.

3.1. Effects of emotional and moral language on information diffusion

The results of negative binomial models featuring retweet count as dependent variable are shown in Table 2. The use of moral, emotional, and moral-emotional language was associated with a higher retweet count. Specifically, an additional moral word was associated with a 25 % higher retweet probability, an additional emotional word was associated with a 17 % higher retweet probability, and an additional moral-emotional word was associated with an 10 % higher retweet probability. The results were robust after including the control variables.

Next, we examined whether the effects of emotional and moral-emotional language differ by valence (see Table 3). Splitting the emotional language into positive and negative valence revealed that information diffusion (by emotional words) was driven primarily by positive emotional words, which were associated with a 25 % increase in retweet probability for each additional word. Negative emotional words were not significantly associated with retweet probability. In Model 3 (that included control variables), the effect of negative emotional words was associated with a slight increase in retweet count, however, it was significantly smaller compared to the effect of positive emotional words ($z = -13.12$, $p < .001$).

A different pattern emerged for moral-emotional words, where the use of negative valence words was associated with a stronger (34 %) increase in retweet probability than the use of positive valence words (only a 4 % increase). However, adding the control variables in Model 3 rendered this difference much smaller and non-significant ($z = 0.86$, $p = .388$). Note that the effect of purely (non-emotional) moral language explaining retweet count remained robust across all models.

3.2. Effect of anger and anxiety words on information diffusion

In the last series of models, we examined whether the use of anger and anxiety words were associated with retweet count (see Table 4). In Model 1, both anger and anxiety words were associated with a higher retweet count, with a 4 % and an 8 % increase for each additional word, respectively. When adding the control variables in Model 3, the effect of anger words remained significant, with an 8 % increase in retweet count per word; while the effect of anxiety words turned non-significant. A stepwise inclusion of control variables in the equation revealed that including the follower count rendered the effect of anxiety words non-significant.

The effects of the language predictors on retweet probability are illustrated in Fig. 2. This figure is a visual representation of the main regression model estimates, specifically, the association between the language variables and retweet count. For example, results from Table 1 are represented in Panel A of Fig. 2. The x-axis represents the count of words of the respective language category contained in a tweet, while the y-axis represents the predicted retweet count of the regression model.

4. Discussion

In summary, the present study finds evidence that the use of emotional, moral, and moral-emotional language is associated with information diffusion of taxation related messages on social media. This finding is a novel contribution to the literature on tax psychology, as it shows the usefulness of a previously unexplored, but widely used,

Table 4
Negative binomial regression for anger and anxiety words explaining retweet count.

	Model 1		Model 2		Model 3	
Predictors	*IRR*	*95 % CI*	*IRR*	*95 % CI*	*IRR*	*95 % CI*
(Intercept)	0.45***	0.44 – 0.45	0.25***	0.24 – 0.25	0.24***	0.24 – 0.25
Anger	1.04 *	1.00 – 1.08			1.08***	1.05 – 1.12
Anxiety	1.08 *	1.02 – 1.15			1.01	0.95 – 1.06
Follower count			20.19***	18.11 – 22.50	20.03***	17.98 – 22.33
Media [1]			3.16***	3.06 – 3.26	3.17***	3.07 – 3.27
URL [1]			1.25***	1.22 – 1.28	1.26***	1.23 – 1.29
N_{tweets}	357,592		357,592		357,592	
AIC	458,012.579		439,951.062		439,932.695	
log-Likelihood	−229,002.289		−219,970.531		−219,959.347	

Note. * $p < .05$, *** $p < .001$. IRR = Incidence rate ratio. CI = Confidence interval. Anger and anxiety are counts representing the number of words of the respective category used in a tweet. The follower count has been standardized (z-transformed) before the analyses, so that the respective estimates reflect a one standard deviation change. Media: 1 = tweet contains media, 0 = tweet does not contain media. URL: 1 = tweet contains URL link, 0 = tweet does not contain URL link.

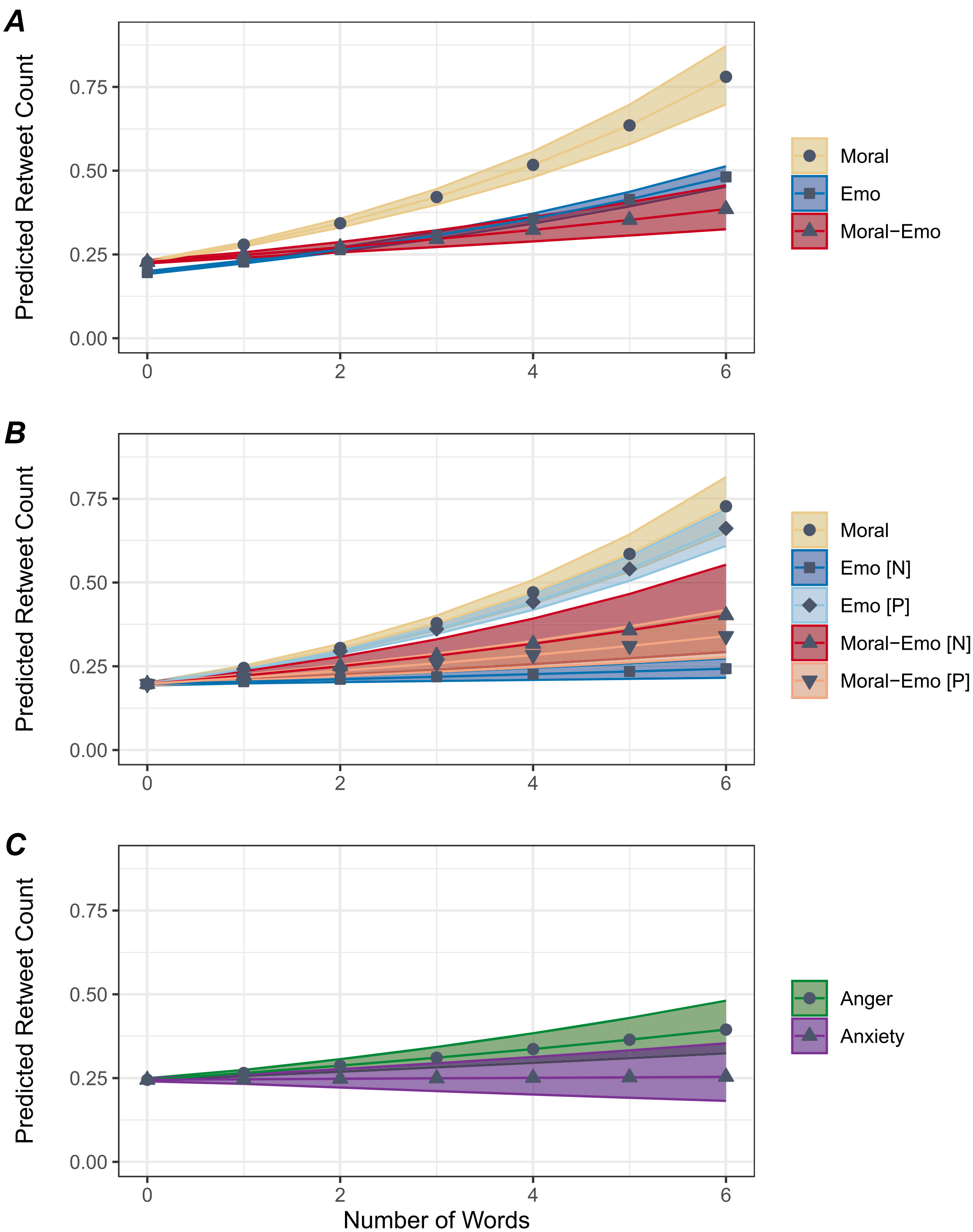

Fig. 2. Predicted Retweet Counts for Different Word Categories
Note. The x-axis represents the number of words from the respective language category contained in a tweet. The y-axis represents the predicted retweet count. Predicted values are based on models including control variables. [P] = positive. [N] = negative. The plots depicted are winsorized on the right part of the x-axis, as there was a small number of tweets containing a very high number of words. Importantly, the analyses were done on non-winsorized data.

medium for communicating information and exchanging opinions about taxes.

Importantly, even though the use of emotional language in general was associated with greater information diffusion, we also found that this effect was stronger for positive (compared to negative) emotional language. This finding is surprising in the light of research on negativity bias (e.g., Baumeister et al., 2001; Kahneman & Tversky, 1979; Rozin & Royzman, 2001). We speculate that a stronger impact of positive (compared to negative) emotional language could be a reflection of the Expectancy Violation Effect (e.g., Jussim et al., 1987): people might react stronger to informational cues that contradict their prior expectations. Given that people hold generally rather negative attitudes towards taxes (Kogler & Kirchler, 2020), tweets about taxes containing positive cues might draw more attention and hence receive more retweets (e.g., Berger & Milkman, 2012). Importantly, it has to be noted that by simply counting the word occurrences of each category is not necessarily a reflection of how people actually feel while reading these messages (Kross et al., 2019). Our results only indicate that the presence of certain word categories is associated with information diffusion. Furthermore, the interpretations of differences between positive and negative emotional language should be taken cautiously, as the discrepancy between the two categories did not consistently emerge across the different robustness checks (see supplemental materials) and was restricted to the use of purely emotional (and not moral-emotional) words.

We also found evidence that content containing anger words was associated with more information diffusion. While previous research found that anger and anxiety are often discussed within the context of taxation (Enachescu et al., 2019; Erard et al., 2018), it seems that only expressions of anger are associated with information diffusion.

As the results suggest that moral and emotional language is associated with information diffusion of tax related content on social media, we assume that people take moral and emotional expressions into account, when thinking about taxation. This idea converges with studies showing that emotions play a role in peoples' judgements about taxes (Enachescu et al., 2019; Olsen et al., 2018), as well as morals appeals being able to increase compliance behavior (e.g., Bott et al., 2020). In a broader fashion, the present findings contribute to the literature of how people reason about taxes. For example, Stantcheva (2021) identifies different profiles of people who (don't) see inequality as a major issue, and therefore emphasize the (un-)fairness of the tax system. These fairness concerns are in line with the present finding of moral language being the main driver of information diffusion.

We used data from Twitter, a single social media platform where messages have specific and unique restrictions. For example, the character limit of 280 (at the time of the study) can cause adaptations in language use, such as using more abbreviations (Boot et al., 2019). While this does not diminish the relevance of the effect of moral and emotional language on information diffusion demonstrated here, it would be interesting to explore whether this effect extends to different social media platforms and other languages.

One important challenge of using data from social media is to clearly distinguish between human and computer-generated content. It is important to identify which content was generated by bots to make sure that the study conclusions apply to human (not bot) social media behavior (e.g., use of specific language to increase engagement). In the present analysis several robustness checks were conducted to make sure that our conclusions are not driven by bot activity (excluding duplicate messages, examining separate subsamples, excluding tweets with many hashtags, and excluding users with no followers). While bot detection comes with many different challenges, given the results of the several robustness checks, we believe that bot activity is unlikely to have biased our results.

Governments and tax authorities invest great effort in reaching out to citizens to promote their positive image and increase tax morale. We found that messages about taxation that contain emotional and moral language are more likely to spread on social media. Hence, by simply twitching message language, tax office image management initiatives could be more fruitful than before. Also, considering the mere volume of information being shared on social media, the tax officials might reach broader audiences compared to other information mediums. Finally, the exposure to the information shared online is likely to impact attitudes and behaviors. Indeed, existing research has shown that sharing taxation related information has implications for tax compliance (e.g., Castro & Scartascini, 2019; Drago et al., 2020), suggesting that the use of moral and emotional words in Twitter messages might have downstream consequences for tax attitudes and tax paying behaviors.

Moreover, tracking changes in negative and positive emotional language in tax-related tweets may allow for an unobtrusive assessment of tax morale or tax climate (e.g., a Twitter Tax Thermometer). Existing research has mostly relied on self-report measures of tax attitudes in nationally representative surveys, such as the World Values Survey (Haerpfer et al., 2020) or the Eurobarometer (European Commission, 2021). This method involves high costs and effort, provides estimates with time delays, and offers a limited possibility to study change over time. Analyzing the sentiment about taxation on social media could offer a real-time assessment of the tax climate. In addition, identifying the most frequently discussed tax related topics could help policy makers determine where action or reform is needed most urgently. Finally, by tracking changes in sentiment over time and/or across regions, one could also examine the effects of certain changes in tax policies and procedures.

CRediT authorship contribution statement

Žiga Puklavec: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. **Olga Stavrova:** Conceptualization, Formal analysis, Methodology, Writing – review & editing. **Christoph Kogler:** Conceptualization, Writing – review & editing. **Marcel Zeelenberg:** Conceptualization, Writing – review & editing.

Declaration of Competing Interest

We have no conflict of interest to disclose. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

All data that we are allowed to share can be found at https://osf.io/4ym7p/. To retrieve the used messages, it is necessary to "rehydrate" the tweets with the provided identification numbers.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.socec.2024.102203.

References

Alstadsæter, A., Kopczuk, W., & Telle, K. (2019). Social networks and tax avoidance: Evidence from a well-defined Norwegian tax shelter. *International Tax and Public Finance, 26*(6), 1291–1328. https://doi.org/10.1007/s10797-019-09568-3

Anwar, A., & Yaqub, U. (2020). Bot detection in twitter landscape using unsupervised learning. In *The 21st Annual International Conference on Digital Government Research* (pp. 329–330). https://doi.org/10.1145/3396956.3401801

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.org/10.1037/1089-2680.5.4.323

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research, 49*(2), 192–205. https://doi.org/10.1509/jmr.10.0353

Boning, W. C., Guyton, J., Hodge, R., & Slemrod, J. (2020). Heard it through the grapevine: The direct and network effects of a tax enforcement field experiment on firms. *Journal of Public Economics, 190*. https://doi.org/10.1016/j.jpubeco.2020.104261, 104261-104261.

Boot, A. B., Tjong Kim Sang, E., Dijkstra, K., & Zwaan, R. A. (2019). How character limit affects language usage in tweets. *Palgrave Communications, 5*(1). https://doi.org/10.1057/s41599-019-0280-3

Bott, K. M., Cappelen, A. W., Sørensen, E.Ø., & Tungodden, B. (2020). You've got mail: A randomized field experiment on tax evasion. *Management Science, 66*(7), 2801–2819. https://doi.org/10.1287/mnsc.2019.3390

Brady, W. J., Wills, J. A., Burkart, D., Jost, J. T., & Van Bavel, J. J. (2019). An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General, 148*(10), 1802–1813. https://doi.org/10.1037/xge0000532

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., Van Bavel, J. J., & Fiske, S. T. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences of the United States of America, 114*(28), 7313–7318. https://doi.org/10.1073/pnas.1618923114

Braithwaite, V. (2003). Dancing with tax authorities: Motivational postures and non-compliant actions. In V. Braithwaite (Ed.), *Taxing democracy* (pp. 15–39).

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., et al. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal, 9*(2), 378–400. https://doi.org/10.32614/RJ-2017-066.].

Burton, J. W., Cruz, N., & Hahn, U. (2019). *How real is moral contagion in online social networks? CogSci 2019.* July 24-27 https://cogsci.mindmodeling.org/2019/papers/0051/0051.pdf.

Castro, E., & Scartascini, C. (2019). *IDB Discussion Paper No. 665*. Inter-American Development Bank. https://doi.org/10.18235/0001661

Chavoshi, N., Hamooni, H., & Mueen, A. (2016). Identifying correlated bots in Twitter. In E. Spiro, & Y.-Y. Ahn (Eds.), *Social informatics* (pp. 14–21). Springer International Publishing.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2010). Who is tweeting on Twitter: Human, bot, or cyborg?. In *Proceedings of the 26th Annual Computer Security Applications Conference* (pp. 21–30). https://doi.org/10.1145/1920261.1920265

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology, 58*(6), 1015–1026. https://doi.org/10.1037/0022-3514.58.6.1015

DataReportal. (2021). *Digital 2021: Global overview report*. https://datareportal.com/reports/digital-2021-global-overview-report.

Drago, F., Mengel, F., & Traxler, C. (2020). Compliance behavior in networks: Evidence from a field experiment. *American Economic Journal: Applied Economics, 12*(2), 96–133. https://doi.org/10.1257/app.20170690

Enachescu, J., Olsen, J., Kogler, C., Zeelenberg, M., Breugelmans, S. M., & Kirchler, E. (2019). The role of emotions in tax compliance behavior: A mixed-methods approach. *Journal of Economic Psychology, 74*, Article 102194. https://doi.org/10.1016/j.joep.2019.102194

Erard, B., Kasper, M., Kirchler, E., & Olsen, J. (2018). What influence do IRS audits have on taxpayer attitudes and perceptions? Evidence from a national survey. *National Taxpayer Advocate Annual Report to Congress, 2*, 77–130.

European Commission, B. (2021). *Eurobarometer 93.1 (2020)* GESIS Datenarchiv, Köln. ZA7649 Datenfile Version 1.2.0, 10.4232/1.13746.

Firmer, J. A., Boghrati, R., Haidt, J., Graham, J., & Dehgani, M. (2019). *Moral foundations dictionary for linguistic analyses 2.0*. Charlottesville, VA: Center for Open Science. https://doi.org/10.17605/OSF.IO/EZN37

Grier, C., Thomas, K., Paxson, V., & Zhang, M. (2010). @spam: The underground on 140 characters or less. In *Proceedings of the 17th ACM Conference on Computer and Communications Security* (pp. 27–37). https://doi.org/10.1145/1866307.1866311

Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., et al. (2020). *World values survey: Round seven – country-pooled datafile*. Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat. https://doi.org/10.14281/18241.1

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852–870). Oxford University Press.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that Liberals may not recognize. *Social Justice Research, 20*(1), 98–116. https://doi.org/10.1007/s11211-007-0034-z

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus, 133*(4), 55–66. https://doi.org/10.1162/0011526042365555

Jussim, L., Coleman, L. M., & Lerch, L. (1987). The nature of stereotypes: A comparison and integration of three theories. *Journal of Personality and Social Psychology, 52*(3), 536–546. https://doi.org/10.1037/0022-3514.52.3.536

JustAnotherArchivist. (2021). *Snscrape [Computer software]*. https://github.com/JustAnotherArchivist/snscrape.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica : Journal of the Econometric Society, 47*(2), 263–291. https://doi.org/10.2307/1914185

Kirchler, E. (1998). Differential representations of taxes: Analysis of free associations and judgments of five employment groups. *Journal of Socio-Economics, 27*(1), 117–131. https://doi.org/10.1016/s1053-5357(99)80080-8

Kogler, C., & Kirchler, E. (2020). Taxpayers' subjective concepts of taxes, tax evasion, and tax avoidance. In R. F. van Brederode (Ed.), *Ethics and taxation* (pp. 191–205). Springer. https://doi.org/10.1007/978-981-15-0089-3_8.

Kross, E., Verduyn, P., Boyer, M., Drake, B., Gainsburg, I., Vickers, B., et al. (2019). Does counting emotion words on online social networks provide a window into people's subjective experience of emotion? A case study on Facebook. *Emotion (Washington, D.C.), 19*(1), 97–107. https://doi.org/10.1037/emo0000416

Lozza, E., & Castiglioni, C. (2018). Tax climate in the national press: A new tool in tax behaviour research. *Journal of Social and Political Psychology, 6*(2), 401–419. https://doi.org/10.5964/jspp.v6i2.937

Olsen, J., Kasper, M., Enachescu, J., Benk, S., Budak, T., & Kirchler, E. (2018). Emotions and tax compliance among small business owners: An experimental survey. *International Review of Law and Economics, 56*, 42–52. https://doi.org/10.1016/j.irle.2018.05.004

Onu, D., & Oats, L. (2016). Paying tax is part of life": Social norms and social influence in tax communications. *Journal of Economic Behavior and Organization, 124*, 29–42. https://doi.org/10.1016/j.jebo.2015.11.017

Onu, D., & Oats, L. (2018). Tax talk: An exploration of online discussions among taxpayers. *Journal of Business Ethics, 149*(4), 931–944. https://doi.org/10.1007/s10551-016-3032-y

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.

Puklavec, Ž., Kogler, C., Stavrova, O., & Zeelenberg, M. (2023). What we tweet about when we tweet about taxes: A topic modelling approach. *Journal of Economic Behavior & Organization, 212*, 1242–1254. https://doi.org/10.1016/j.jebo.2023.07.005

Rincke, J., & Traxler, C. (2009). *CESifo Working Paper Series No. 2549*. Center for Economic Studies and Ifo Institute. https://ssrn.com/abstract=1347956.

Rinker, T. W. (2021). *sentimentr: Calculate Text Polarity Sentiment version 2.9.0*. https://github.com/trinker/sentimentr.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*(4), 296–320. https://doi.org/10.1207/2FS15327957PSPR0504_2

Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., & Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature Human Behaviour, 7*(5), 812–822. https://doi.org/10.1038/s41562-023-01538-4

Stantcheva, S. (2021). Understanding tax policy: How do people reason? *The Quarterly Journal of Economics, 136*(4), 2309–2369. https://doi.org/10.1093/qje/qjab033

Wang, S.-Y. N., & Inbar, Y. (2021). Moral-language use by U.S. political elites. *Psychological Science, 32*(1), 14–26. https://doi.org/10.1177/0956797620960397

Wang, S.-Y. N., & Inbar, Y. (2022). *Re*-examining the spread of moralized rhetoric from political elites: Effects of valence and ideology. *Journal of Experimental Psychology: General, 151*(12), 3292–3303. https://doi.org/10.1037/xge0001247

Wenzel, M. (2005a). Misperceptions of social norms about tax compliance: From theory to intervention. *Journal of Economic Psychology, 26*(6), 862–883. https://doi.org/10.1016/j.joep.2005.02.002

Wenzel, M. (2005b). Motivation or rationalisation? Causal relations between ethics, norms and tax compliance. *Journal of Economic Psychology, 26*(4), 491–508. https://doi.org/10.1016/j.joep.2004.03.003

Zhu, X., Kim, Y., & Park, H. (2020). Do messages spread widely also diffuse fast? Examining the effects of message characteristics on information diffusion. *Computers in Human Behavior, 103*, 37–47. https://doi.org/10.1016/j.chb.2019.09.006