Data & Policy (2025), 7: e22
doi:10.1017/dap.2024.98

RESEARCH ARTICLE

Count on Me: Moral Language in Social Media and Policy Discourse during the Ukraine-Russia Conflict

Eimon Amjadi[1] and Richard S. John[2]

[1]Dornsife College of Letters, Arts, and Sciences, University of Southern California, Los Angeles, CA, USA
[2]Department of Psychology, University of Southern California, Los Angeles, CA, USA
Corresponding author: Eimon Amjadi; Email: eamjadi@usc.edu

Received: 01 October 2024; **Accepted:** 20 November 2024

Keywords: LIWC; Moral Foundations Theory; Social media; Text analysis; Twitter

Abstract

We apply moral foundations theory (MFT) to explore how the public conceptualizes the first eight months of the conflict between Ukraine and the Russian Federation (Russia). Our analysis includes over 1.1 million English tweets related to the conflict over the first 36 weeks. We used linguistic inquiry word count (LIWC) and a moral foundations dictionary to identify tweets' moral components (care, fairness, loyalty, authority, and sanctity) from the United States, pre- and post-Cold War NATO countries, Ukraine, and Russia. Following an initial spike at the beginning of the conflict, tweet volume declined and stabilized by week 10. The level of moral content varied significantly across the five regions and the five moral components. Tweets from the different regions included significantly different moral foundations to conceptualize the conflict. Across all regions, tweets were dominated by loyalty content, while fairness content was infrequent. Moral content over time was relatively stable, and variations were linked to reported conflict events.

Policy Significance Statement

Our study reveals the critical need to customize news headlines to align with the moral frameworks in specific regions for various events. A key policy implication is whether to match these moral concepts to resonate and be perceived as relevant or to introduce new moral perspectives that challenge existing norms. This approach could enhance the effectiveness of communications, fostering better understanding and engagement. Our findings advocate for a strategic evaluation of messaging techniques, emphasizing the importance of context-sensitive communications to optimize the global impact and relevance of policy initiatives. This tailored approach could significantly improve the reception and effectiveness of messages across diverse cultural landscapes.

1. Introduction

This article explores the moral foundations of social media discourse related to the armed conflict between Ukraine and the Russian Federation (Russia). Rather than conducting a topical analysis of social media content, we focus on five components of Moral Foundations Theory to better understand the moral framework used by individuals from different regions of the world to conceptualize the first 36 weeks of the conflict (Atari et al., 2023). Previous research has demonstrated the efficacy of the moral foundations framework for understanding fundamental attitudinal differences among groups on particular topics.

Research has shown that Moral Foundations Theory can be applied to reveal moral content in policy debates, e.g., the death penalty and same-sex marriage (Tatalovich and Wendell, 2018; Wendell and Tatalovich, 2021). By examining social media discourse, we aim to understand how the public uses moral language to describe the first eight months of the Ukraine-Russian conflict and how these moral frameworks influence public perception and response (Parmelee et al., 2024). This understanding can help identify the moral content in social media discussions, allowing policymakers to differentiate between pure and mixed morality policies (Ronzhyn and Wimmer, 2021; Wendell and Tatalovich, 2021). One policy implication of our study is the need to tailor messaging to the moral framework utilized in particular world regions for particular types of events. This tailored approach could significantly improve the effectiveness of messages across different regions and cultural groups, aligning with broader trends toward data-driven policy innovation (Lämmerhirt et al., 2024) and evidence-based policy assessment (Liu and Dijk, 2022).

Prior studies have examined Twitter data about the Ukraine-Russian war. Building on established methods for analyzing Twitter sentiment and influence patterns (Bae and Lee, 2012), studies have examined a single week of the conflict and the role of Online Social Networks in disseminating information (Haq et al., 2022). Moreover, similar efforts have examined Twitter data during the Ukraine--Russia conflict to identify false and unverified claims about the war (La Gatta et al., 2023). Other studies have collected data from the conflict and reported the volume of tweets over time (Chen and Ferrara, 2023). This growing body of conflict-related social media research complements broader efforts to forecast and analyze conflicts using large-scale datasets (Mueller et al., 2024) and machine learning approaches (Murphy et al., 2024). Beyond conflict analysis, researchers have leveraged Twitter data to examine public discourse during various global events, from natural disasters (Bruns and Liang, 2012) to public health crises like COVID-19 (Biddle et al., 2022; Rowe et al., 2021), demonstrating the platform's versatility as a lens for understanding collective moral responses to significant events. Our study examines Twitter data about the war over 36 weeks and compares different regions' use of different moral language to describe individual thoughts and feelings related to the war.

As suggested by Graham et al. (2018), "Moral Foundations Theory (MFT; Graham et al., 2013; Haidt and Joseph, 2004) was designed to explain both the variety and universality of moral judgments." MFT is based on four propositions about morality, two of which are central to the current study: (1) *Intuitions come first*, and (2) *There are many psychological foundations of morality* (Chung and Pennebaker, 2018; Graham et al., 2018; Wang and Inbar, 2021). By focusing on moral language, we aim to capture social media users' visceral, intuitive reactions to the war in Ukraine. We utilize the MFT framework to characterize moral commentary into five components: care, fairness, loyalty, authority, and sanctity.

By examining public discourse through the lens of MFT, we can better understand how different regions react differently to the same event and how cultural context can influence the morality judgments of individuals and the subsequent effects on adapted policies. This builds on previous research demonstrating how moral values fundamentally shape foreign policy attitudes (Kertzer et al., 2014). Prior studies have shown that rhetoric strengthens the link between individuals' moral foundations and their political attitudes (Clifford and Jerit, 2013), particularly persuading those who endorse the relevant moral beliefs (Clifford et al., 2015). This study also explores how the five regions view the war differently over the first 36 weeks of conflict. We aim to gain insight into how regional history and culture can impact the moral lens through which people conceptualize reported war events.

Psychologists developed MFT to create a framework for organizing human morality. Prioritization of the five moral foundations (care, fairness, loyalty, authority, and sanctity) is useful for predicting and understanding group differences. For example, conservative and liberal political ideology in the United States has been linked to different moral priorities (Graham et al., 2009; Haidt et al., 2009). Similarly, priorities for moral foundations have been shown to predict attitudes toward culture war issues (Koleva et al., 2012), vaccine hesitancy (Amin et al., 2017), needle exchange (Christie et al., 2019), and the use of

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

nuclear, chemical, and conventional military strikes (Smetana and Vranka, 2021). Reimer et al. (2022) reported that moral values predict county-wide COVID-19 vaccination rates.

MFT has recently been used to characterize social media posts. Chen et al. (2022) provide an overview of the use of Twitter as a research tool for understanding public concerns. For example, Sagi and Dehghani (2014) used moral foundations in Twitter data to characterize attitudes about the U.S. federal shutdown in 2013. Sylwester and Purver (2015) used tweet data to describe links between moral foundations and U.S. political orientation.

MFT has been applied to Twitter data on various moral and social issues (abortion, homosexuality, immigration, religion, and immorality) (Kaur and Sasahara, 2016). Priorities over the five moral foundations have been linked to positions taken in tweets related to extremist politics (Alizadeh et al., 2019), immigration policy (Grover et al., 2019), Asian hate crime during COVID-19 (Kim et al., 2022), bushfires in Australia (Nasim et al., 2022), and COVID-19 vaccination (Borghouts et al., 2023; Schmitz et al., 2023).

2. Current study

The current study has the following five aims:

1. Estimate the total volume of English-language tweets related to the Ukraine and Russian conflict over the first 36 weeks in the United States, NATO countries, Ukraine, and Russia.
2. Compare the rates of care, fairness, loyalty, authority, and sanctity to determine the dominant moral foundations used in tweets about the conflict.
3. Compare the rates of moral foundations across the five regions: the U.S., pre-Cold War NATO countries, post-Cold War NATO countries, Ukraine, and Russia.
4. Characterize each region in terms of a profile describing the moral discourse observed in English-language tweets.
5. Identify trends over the first 36 weeks of the conflict for each of the five moral foundations within each region.
6. Evaluate the implications of moral discourse for policy communication strategies and how aligning or challenging dominant moral frameworks can affect the effectiveness of messaging in different regions. We seek to provide actionable insights for policymakers on tailoring communication strategies to resonate with regional moral norms or to introduce new perspectives that may influence public engagement and discourse.

In contrast to other social media text analyses, our analysis does not attempt to characterize beliefs or attitudes expressed in Twitter content. Instead, by focusing on moral foundations, we aim to understand better how social media users conceptualize the conflict morally. Our analysis allows us to understand how moral discourse related to conflict varied across regions overtime. Our study is novel in that it addresses moral discourse about an international armed conflict and compares five different stakeholder regions.

By analyzing which of the five components of MFT were particularly relevant for Twitter users in discussing the conflict, we provide insights that can assist policymakers in constructing effective strategies for tailoring their messaging to resonate with the specific moral concerns or values of different regions.

We focus on which of the five components of MFT were particularly relevant for social media users in discussing the conflict between Ukraine and Russia. The moral foundation categories were each partitioned into two categories: virtue and vice. Virtue content has a positive connotation related to the moral component, while vice has a negative sense. For example, care (virtue) and harm (vice) are related to the same MFT category.

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

3. Methods

3.1 Overview

The methodology for this current study consisted of four steps:

1. Collect tweets about the conflict between Ukraine and Russia from the target countries.
2. Preprocess the data to ensure the data collected is relevant to the conflict.
3. Using a moral foundations dictionary, perform text analysis of the Twitter data using LIWC (Linguistic Inquiry and Word Count).
4. Conduct statistical analysis on the moral foundation scores output from LIWC to compare moral content over time and region.

3.2 Twitter search

We utilized the Twitter API website (developer.twitter.com/apitools/downloader) to download tweets. Figure 1 provides the filter used for U.S. tweets. Similar filters were used to obtain English-language tweets for Russia and Ukraine, replacing the U.S. with Russia or Ukraine. We partitioned NATO countries into pre- and post-Cold War member nations. Figure 2 displays the nations used to filter for each of the groups of NATO nations.

We grouped nations that entered NATO following the end of the Cold War since they have a different historical relationship with Russia and the former Soviet Union than nations that entered NATO before the end of the Cold War. We did not include non-European NATO members (Canada) or nations recently admitted to NATO after February 2022 (Finland, Sweden). Our analysis focuses on nation-group similarities and differences in moral foundations characterizing social media discourse about the war.

Tweets were collected for 36 weeks, beginning with the first week of February and ending with the last week of September. We focused on the first 36 weeks of the conflict to capture the beginning of the war when there were intense public reactions. The first 8 months of the war included many evolving events that were key to creating tension between the regions and drawing public attention. The first 36 weeks of conflict comprise the first phase, including the Russian invasion and Ukraine's counter-offensive in the north, ending in a temporary stalemate by the end of September 2022. Our analysis focused on the first 36 weeks of the conflict, encompassing significant military developments and culminating in a relative strategic stalemate. This timeframe was selected because it captured both the initial dynamic phase of the conflict, marked by major territorial changes and military operations, as well as the transition to a more static phase characterized by reduced territorial shifts and stabilized front lines. By week 36, the conflict had reached a steady state, making it a suitable endpoint for our temporal analysis.

Figure 1. *Filter for U.S. tweets*

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

Table 1. *The number of tweets analyzed for each region*

	USA	Early members of NATO	NATO's post-cold war members	Ukraine	Russia
Number of Tweets	710,495	361,058	36,914	61,152	9796
Average Word Count	29.02	28.79	28.25	29.66	27.76
Median Word Count	26	27	27	30	25
Average Words in Dictionary	4.73	4.09	4.16	3.84	2.21
Median Words in Dictionary	3.85	3.13	2.86	2.86	0.00
Percentage of 0 words in Dictionary	33.93%	38.01%	39.61%	39.28%	63.40%

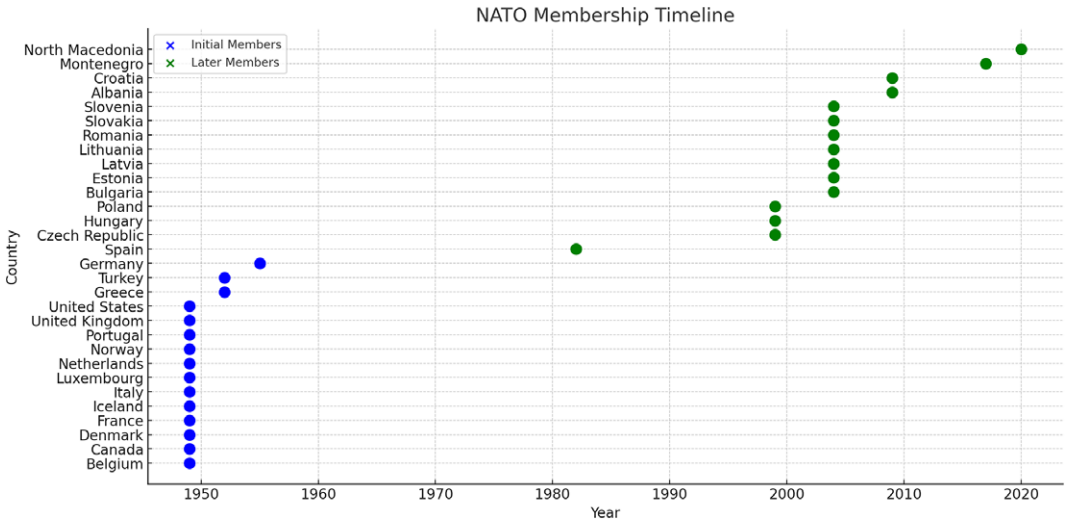

Figure 2. *Countries by year of entry to NATO*

We chose the API downloader program option that excluded retweets. Data were collected in blocks exactly one week long. Table 1 displays the counts of tweets downloaded for the five regions over the entire 36-week period. Table 1 also shows the average and median word count of Tweets, average and median words in the MF dictionary, and percentage of Tweets that did not contain any words in the MF dictionary for each region. By examining the extent to which moral language is present across different regions, we can better identify regional variations in using such language.

We randomly sampled 1000 tweets and examined the content to confirm the filter excluded tweets not about the war. We found that over 97% of the tweets sampled were about the war. Hence, we chose not to conduct post-processing of the tweets.

3.3 LIWC and seed words

Tweets were analyzed using the Moral Foundations dictionary and the Linguistic Inquiry Word Count (LIWC) software (Frimer et al., 2019; Hopp et al., 2021). We used LIWC-22, a software tool

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

Table 2. *LIWC seed words*

	Care	Fairness	Loyalty	Authority	Sanctity
Virtue	Kindness	Fairness	Loyal	Authority	Purity
	Compassion	Equality	Team Player	Obey	Sanctity
	Nurture	Justice	Patriot	Respect	Sacred
	Empathy	Rights	Fidelity	Tradition	Wholesome
Vice	Suffer	Cheat	Betray	Subversion	Impurity
	Cruel	Fraud	Treason	Disobey	Depravity
	Hurt	Unfair	Disloyal	Disrespect	Degradation
	Harm	Injustice	Traitor	Chaos	Unnatural

(LIWC.app, n.d.) commercially available for a nominal licensing fee. The only code utilized in our study to process the 1.1 million tweets is contained in the LIWC-22 software. LIWC is a standard software tool used to gain insight into natural language samples. An MFT dictionary is available for LIWC to analyze text regarding the five moral components.

Seed words for the Moral Foundations dictionary are presented in Table 2. The seed words listed in Table 2 are not the only words comprising the dictionary; they are the base or root words defining the dictionary content domain. The seed words in the table are exemplars for each specific category. LIWC uses text matching as a default strategy, matching words to the appropriate categories based on the dictionary specified by the user. The dictionary we used was Moral Foundations Dictionary 2.0. LIWC utilizes two different strategies to match words:*

(1) Finding exact matches to the seed words that are a part of the LIWC Moral Foundations dictionary
(2) Pattern-based matching to identify synonyms of seed words and words with prefix differences (Bahgat et al., 2022)

LIWC produces moral foundation scores based on the occurrence of words in a tweet that belong to each moral foundation component, i.e., care, fairness, loyalty, authority, and sanctity (Hopp et al., 2021). The five moral foundations represent different frameworks used to discuss the conflict in English-language social media. The MFT foundations have been partitioned into individuating moral concerns (care and fairness) and binding (loyalty, authority, and sanctity) foundations (Graham et al., 2011). Care/harm and fairness/injustice refer to moral issues related to individuals, while loyalty, authority, and sanctity relate to ingroup versus outgroup morality concerns. In an armed invasion of one nation by a neighboring nation, both individuating and binding moral foundations are potentially relevant to public discourse. In particular, public discourse would be expected to be characterized in terms of care/harm to individuals during an armed conflict between nations. War also emphasizes ingroup versus outgroup boundaries; hence, moral discourse would also be expected to relate to binding foundations, such as loyalty and authority. The sanctity foundation, which is usually associated with a deity, would be expected to be part of the discourse if the armed conflict involved a religious component, which is not a prominent motivation for the conflict between Russia and Ukraine.

3.4 Number of tweets analyzed over time

Figure 3 plots the number of English-language tweets per week on a log scale over 36 weeks for the US, Russia, Ukraine, Early NATO Members, and NATO's Post-Cold War Members. We use a log scale because the weekly tweet count by region ranges from about 100 to 100,000. As expected, the overall volume of tweets varies by nation(s) and is associated with population size and the prevalence of English speakers on Twitter.

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

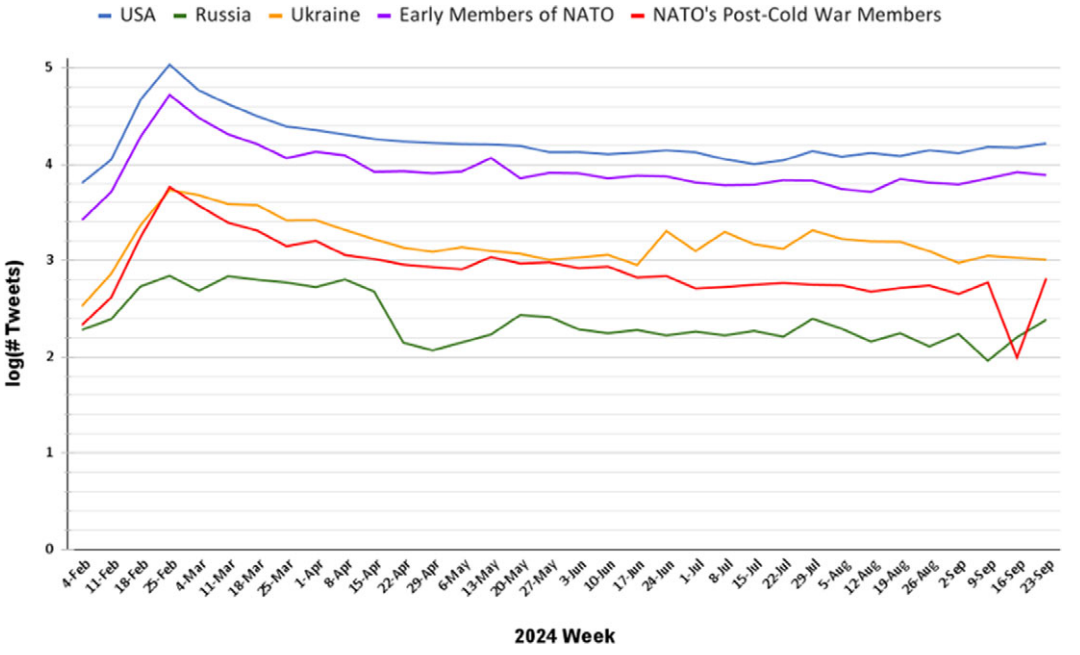

Figure 3. *Log tweet counts by region and week.*

Twitter activity rises quickly during February and peaks during the beginning of the conflict in the fourth week of February. The volume of tweets falls off during March and stabilizes to a value in April that continues until the end of September. While Twitter activity in Ukraine shows an elevated and oscillating pattern throughout the summer, Twitter activity in Russia is nearly extinguished by the end of April. It is likely that Russian Twitter users were inhibited by actions taken by the Russian government to censor social media discourse, or in some cases, by Twitter based on the content of the tweets posted from particular accounts.

4. Results collapsed over time

We conducted a 5 (moral foundations) X 5 (regions) X 2 (MF valence, virtue versus vice) mixed model Analysis of Variance (ANOVA) on the LIWC average weekly scores. Regions were treated as a between-groups factor, and moral foundations and valence were within (repeated) factors. Figure 4 displays mean LIWC scores collapsed across weeks by region and moral foundation for virtue (top plot) and vice (bottom plot). Mean LIWC scores varied across the five moral foundations, $F(4,16) = 1472.88$, $p < .001$, partial eta-squared = .97. Tweet content was greatest for loyalty words, followed by care words. Fairness words were hardly used in tweets related to the conflict.

Use of moral foundation words also varied significantly by nation, $F(4,165) = 115.96$, $p < .001$, partial eta-squared = .37. Note that the LIWC scores are per tweet; hence, scores are independent of the overall volume of tweets for a particular nation or group of nations. The U.S. and Cold War NATO nations produced the greatest rate of moral foundation words. Russia produced the lowest rate of moral foundation words.

A significant interaction between region and moral foundation indicated that regions used different moral foundations to conceptualize the conflict, $F(16, 660) = 23.77$, $p < 001$, partial eta-squared = .37. Both loyalty virtue and care vice reveal differences in the prevalence of these two moral foundations across different regions.

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

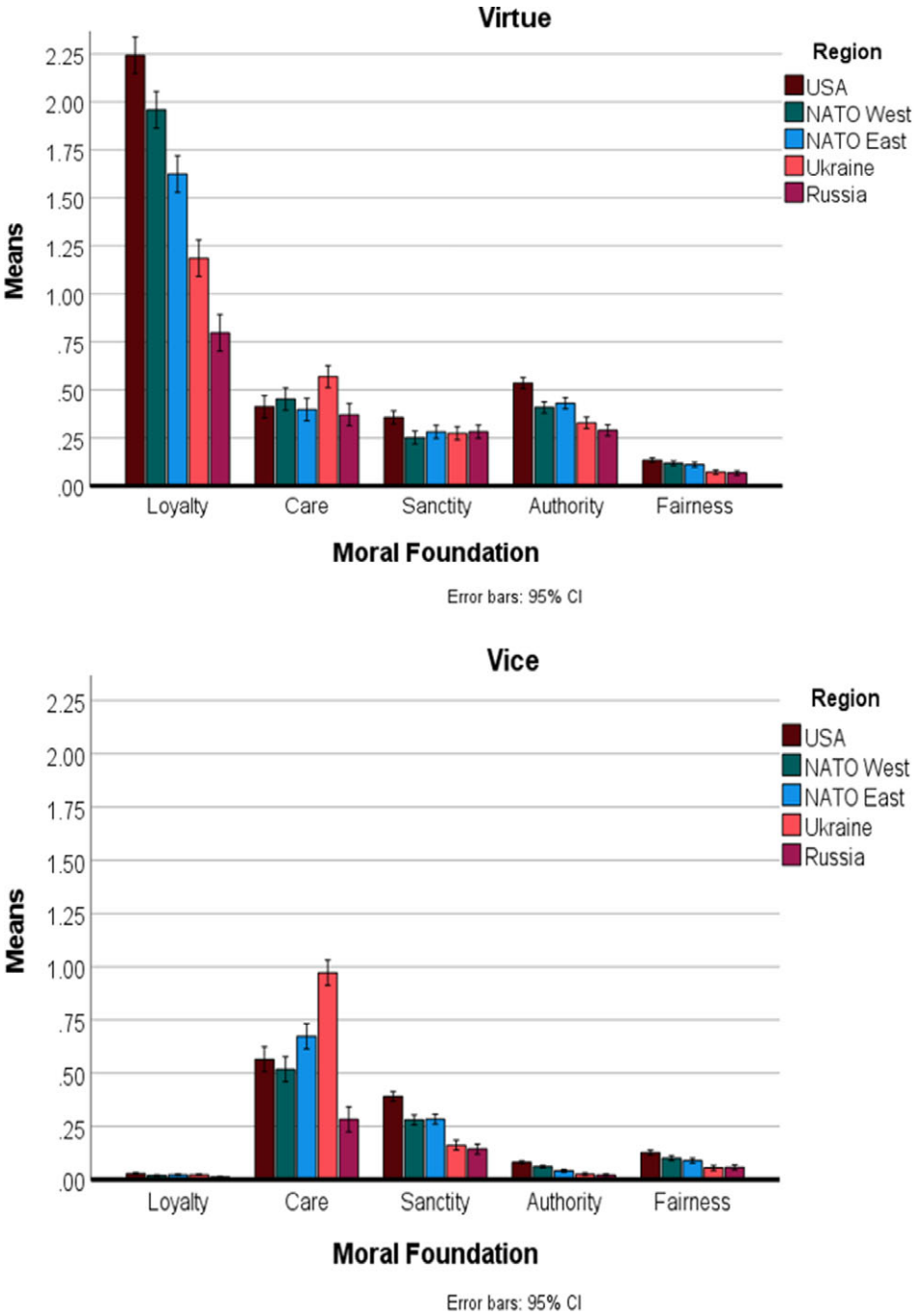

Figure 4. *Mean virtue and vice moral foundation rates by region.*

There was greater virtue content than vice content in the conflict-related tweets, $F(1,165) = 3129.89$, $p < 001$, partial eta-squared = .95. However, there was a significant interaction between moral foundation and valence, $F(4,162) = 62.69$, $p < .001$, partial eta-squared = .60. While virtue tweet content was substantially greater than vice for loyalty and somewhat greater for authority, the reverse was true for care. Virtue and vice content were about the same for Sanctity and Fairness.

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

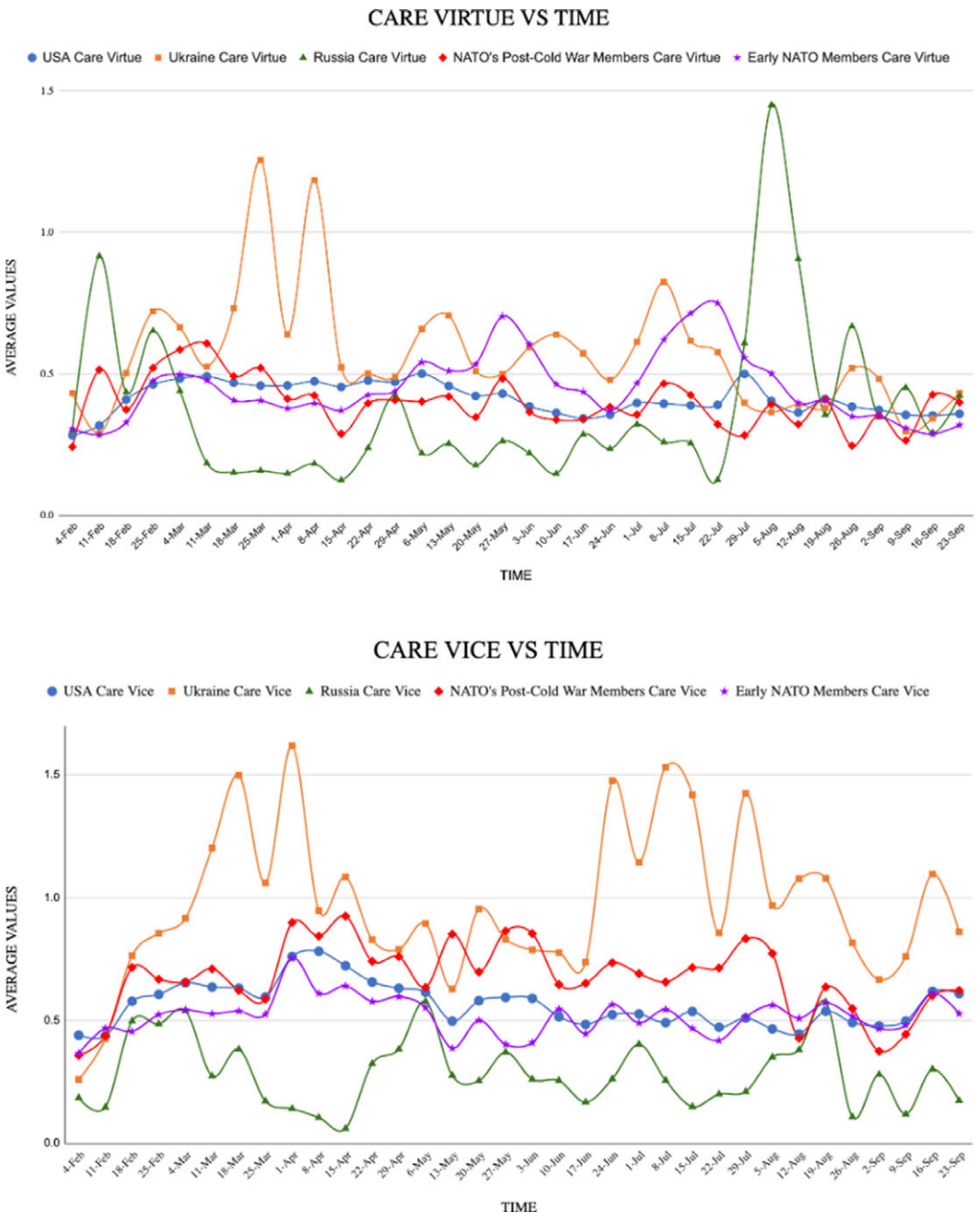

Figure 5. *Care virtue (top) and care vice (bottom) by region over 36 weeks.*

5. Results over time

LIWC average weekly scores are plotted over time in Figures 5–9 for each of the five moral components. In each plot, average scores by region are plotted for virtue on the top and vice on the bottom.

Figure 5 (top, care virtue) shows distinct peak periods for each region. Russia peaked on February 8th and August 5th, while Ukraine peaked at the end of March and the beginning of April. Notably, there appears to be a reciprocal correlation between Ukraine and Russia. The US and early members of NATO

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

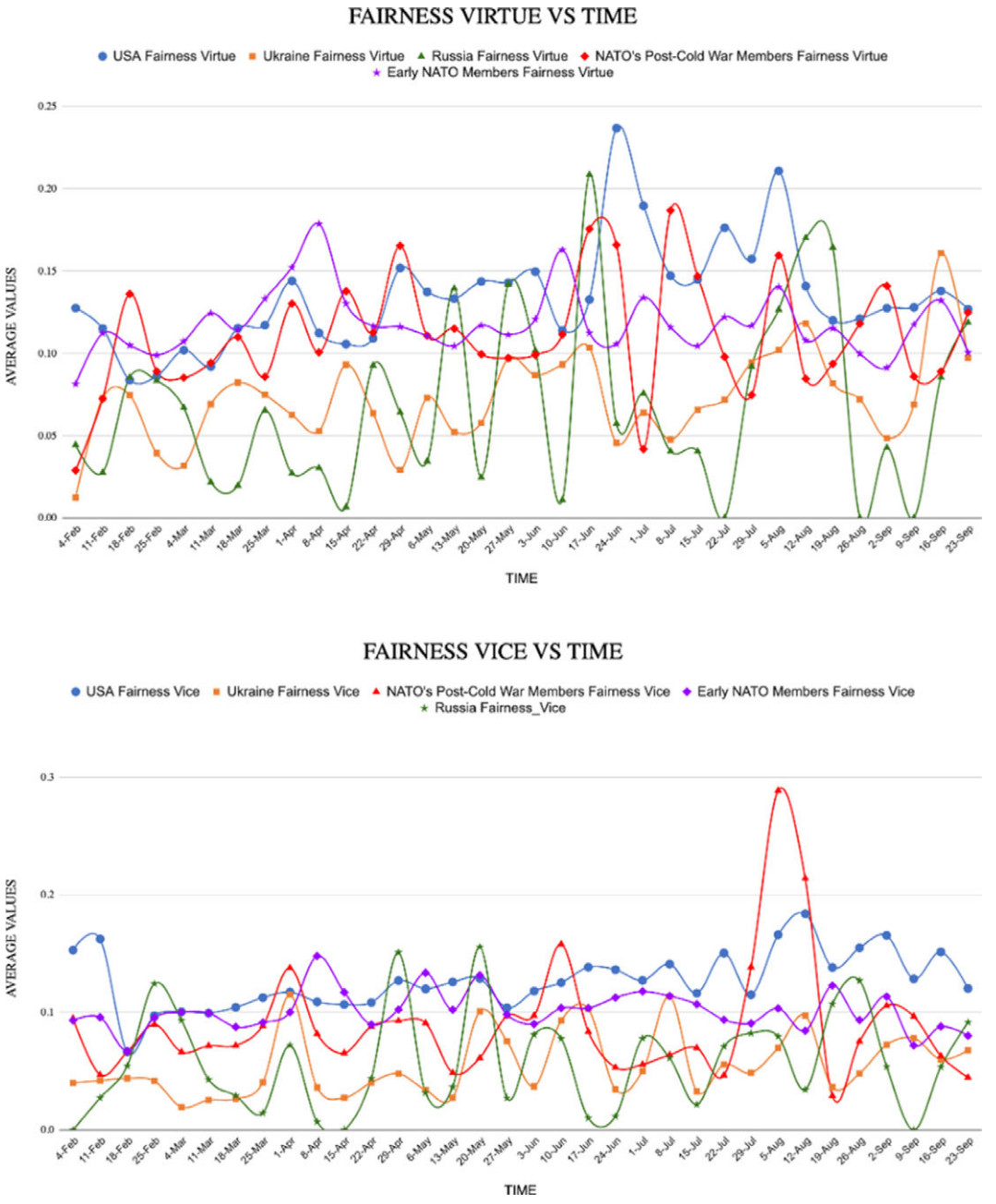

Figure 6. *Fairness virtue (top) and fairness vice (bottom) by region over 36 weeks.*

countries have similar patterns, with peaks on May 19th and the beginning of July, while the US does not have such peaks. On the other hand, post-Cold War NATO members followed a consistent pattern throughout the year, with a much higher peak at the beginning of January.

For care-vice (Figure 5, bottom), there are similar patterns between NATO's post-Cold War members and Ukraine, although the intensity of the peaks is different. Ukraine experiences multiple peaks throughout the year, with one occurring on March 18th and then going down before another peak on April 1st. Additional peaks are observed on June 24th, July 8th, and July 29th. However, Russia has low

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

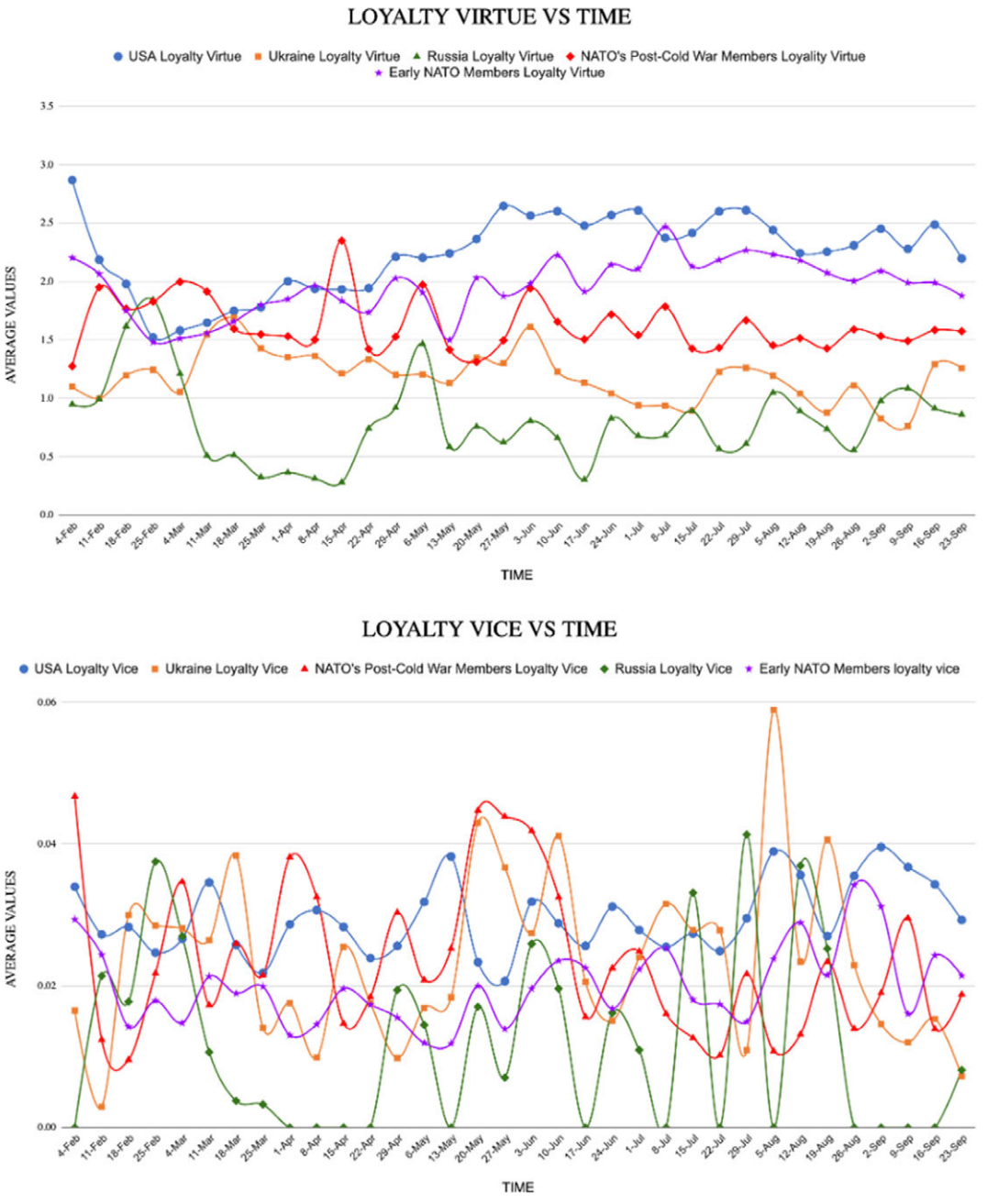

Figure 7. *Loyalty virtue (top) and loyalty vice (bottom) by region over 36 weeks.*

levels, with only small peaks occurring on May 6th, August 19th, and mid-February, almost an inverse of Ukraine's pattern. The rest of the plot exhibits similar patterns.

For fairness virtue (Figure 6, top), the United States has the highest values throughout the data set compared to Russia and Ukraine. However, the U.S's scores are consistently lower than those of Ukraine, although they are still higher than Russia's. Figure 6 (top) displays many peaks throughout the year, with numerous peaks observed before July.

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

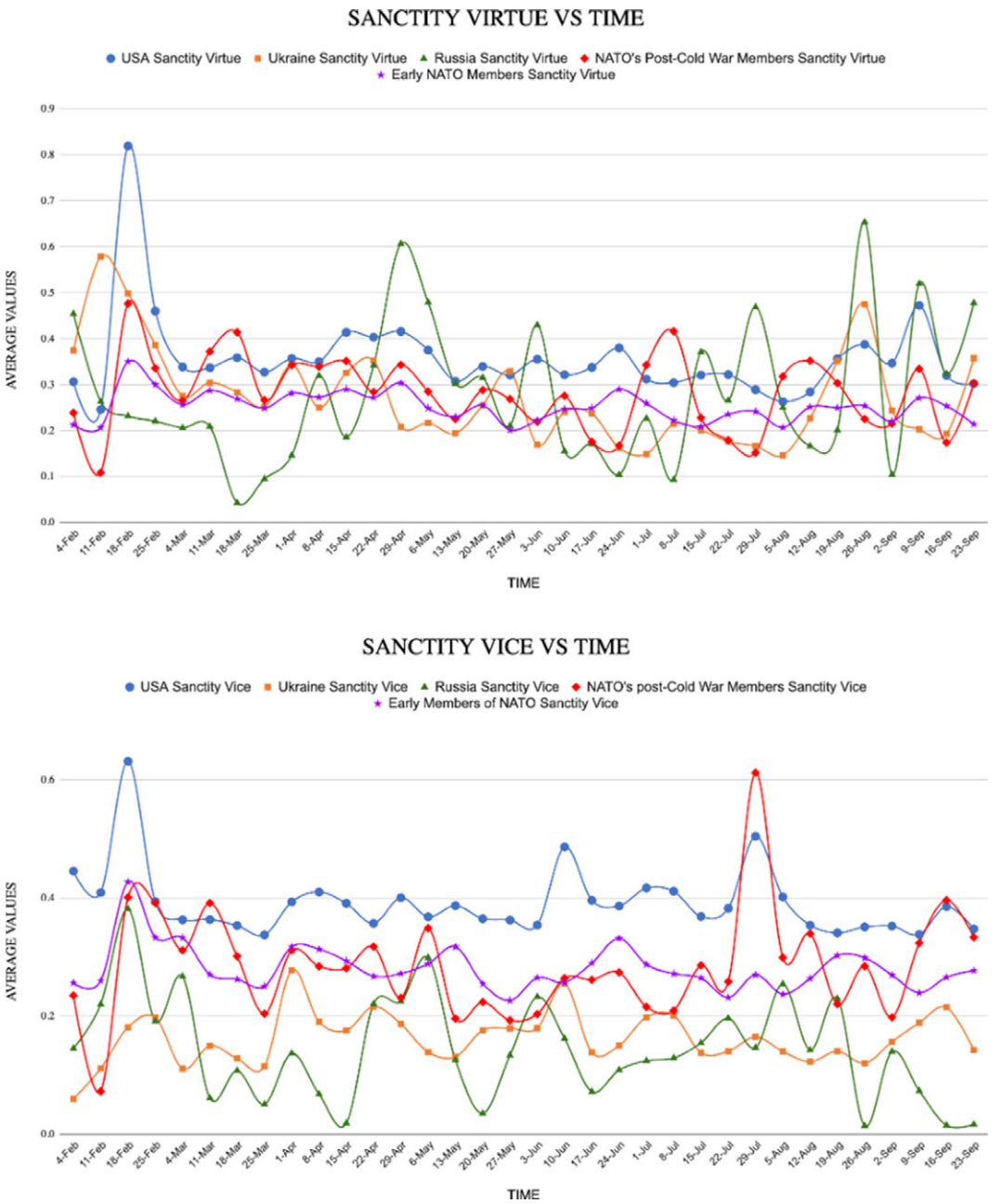

Figure 8. *Sanctity virtue (top) and sanctity vice (bottom) by region over 36 weeks.*

When examining the fairness vice (Figure 6, bottom), it appears that NATO's post-Cold War region has a peak on August 5th. Interestingly, when Russia's scores are higher and peak, Ukraine tends to have lower scores, such as on April 29th.

For loyalty virtue, (Figure 7, top) shows that the United States had the highest values at the beginning of February, then dipped and subsequently increased. The pattern observed for the United States is similar to that seen for early members of NATO countries, but the values are generally lower. Notably, post-Cold War NATO nations and early members of NATO countries have inverse patterns. Ukraine's scores are

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

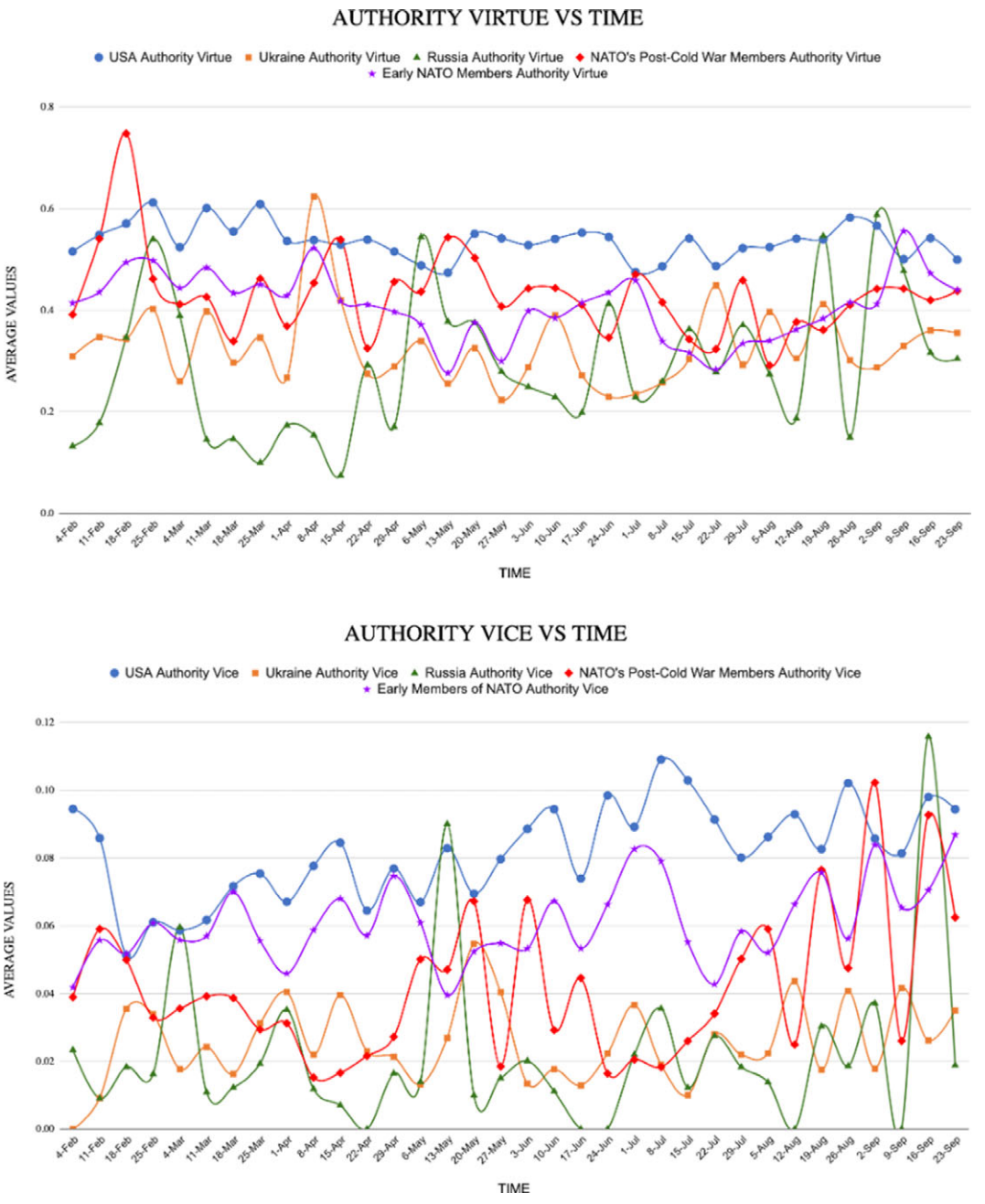

Figure 9. *Authority virtue (top) and authority vice (bottom) by region over 36 weeks.*

similar to those of NATO's post-Cold War members. Meanwhile, Russia's scores are generally lower, with a dip observed at the end of March. However, two peaks were observed in Russia, one on February 25th and another on May 6th.

For loyalty vice (Figure 7, bottom), Russia's scores remained relatively flat at the end of March. Overall, the data display similar patterns and peaks throughout. However, there is a reciprocal relationship in the scores between Russia and Ukraine. Ukraine peaks on August 5th, while Russia is at its lowest

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

point, and when Ukraine's scores decline, Russia's scores tend to increase. There is an inverse correlation between the scores of Russia and Ukraine for the loyalty vice.

Figure 8 (top) shows that the United States had a consistent pattern regarding the sanctity virtue. The United States experienced a peak in mid-February. Meanwhile, Russia had a significant dip near March 18th, followed by a peak on April 29th and another on August 26th. The overall graph for sanctity virtue displays an oscillating pattern that is consistent across.

Figure 8 shows that the United States peaked in mid-February for sanctity vice. On the other hand, NATO post-Cold War nations peak in mid-July. The plot for the sanctity vice displays a generally oscillating, clean, consistent pattern across all. Notably, Russia's scores are consistently low throughout the data set.

For authority virtue, Figure 9 shows that NATO's post-Cold War members had the highest peak on February 18th. The United States remained consistently high throughout the data set. Ukraine had its peak on April 8th. Russia had several low points on March 25th, April 15th, August 12th, and 26th. However, the overall graph (Figure 9) for the authority virtue shows an oscillating pattern that is consistent across.

The authority vice graph (Figure 9) shows that the United States dipped on February 18th. However, it remained relatively high throughout the data set. The United States also had a peak on July 8th. Conversely, Russia had a generally low score but with high peaks on March 4th, May 13th, and September 16th. NATO's post-Cold War members demonstrated a peak on September 2nd. The overall values for the authority vice and virtue were relatively minimal across all 36 weeks.

6. Discussion

These results suggest great consistency in the moral components of English-language social media content across regions and time. The conflict was conceptualized in terms of loyalty and secondarily in terms of care/harm moral foundations. There was little expression of fairness concerns in discourse about the conflict for any of the nations studied. During conflict and war, individuals may express 'care' as feelings of empathy and compassion for victims of injustice.

Conflict characterized by suffering can infuse feelings of compassion for others but can also foster feelings of patriotism and concern for the well-being of one's group. Expressions of loyalty can be high because wars often create a dynamic among individuals to support their fellow citizens as well as develop a sense of ingroup identity. Loyalty can also be due to cultural differences, as some cultures may see their own group's well-being as more important than that of a group outside their region.

The U.S. and Cold War NATO allies utilized moral foundations content more than post-Cold War NATO allies, Ukraine or Russia. Most moral content was positively framed (virtue) rather than negatively framed (vice). One notable exception was for care/harm, which was more often framed negatively as harm than positively as care.

Our study offers a novel approach to understanding social media users' thoughts and feelings about an armed conflict. This study extends research on self-reported beliefs and attitudes about international conflicts to examine the moral components people use to conceptualize the conflict.

The implications of these findings for policy are significant. Tailoring policy to resonate with the moral values corresponding to the public discourse can enhance public support for war-related policies. In some cases, policies can be framed to resonate with the current moral framework of the nation and potentially amplify public support. Policies that align with a moral foundation and are not part of the public discourse may need to be carefully introduced to allow public understanding of moral terms that are not part of the current discourse. Policy communications should be sensitive to the cultural and moral context within each region.

Our findings indicate that different regions had different reactions to the progression of events during the first 36 weeks of the conflict. Results reveal distinct, systematic patterns over time, suggesting that different moral foundations are more relevant as events unfold. Additionally, our findings advocate for strategic evaluation of how news is disseminated in a region appropriate to the cultural context it is reaching. By doing so, the receptivity and effectiveness of messages across diverse landscapes can be

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

improved, and they can impact the relevance of policies by region. Policymakers could use these insights to tailor their messaging to resonate more effectively with diverse populations, potentially aiding conflict de-escalation or attracting international support.

We acknowledge a potential dark side of moral discourse in international conflict (Graham and Haidt, 2012). Nations wanting to build public support and patriotism for the war effort would be expected to use the binding moral foundations (loyalty, authority, and sanctity), emphasizing ingroup versus outgroup differences. In addition, pointing out the harms (morality vices) carried out by the opposing nation would also naturally fit the public discourse about the war. Policy communications in terms of fairness/injustice or sanctity would not have a natural fit in how the public frames the conflict.

7. Limitations and future research

We emphasize that our use of only English-language Twitter data limits our conclusions. We cannot generalize these findings to entire populations of any nation or group of nations. While Russian Twitter activity decreased significantly after April, we cannot rule out that some Russian users may have continued posting using VPNs or other social media platforms that are more popular in the region (e.g., VK, Telegram). Future research is needed to explore the moral content of social media posts about the war in languages commonly spoken in the target countries and commonly used platforms in that region.

Our results are also limited to the first 36 weeks of the conflict. It is certainly possible that different patterns will evolve as the conflict unfolds. Future research should investigate the moral language used on social media at the end of the conflict and compare it to that used at the beginning.

Our results are also limited to the particular approach of defining seed words and pattern-based matching to identify synonyms of seed words and words with prefix differences. The extent to which these methods produce results similar to transformer-based methods is a topic for future research. Transformer-based models could potentially offer advantages through their potential ability to capture contextual nuances and handle semantic variations more effectively that might be missed by pattern-based approaches.

Data availability statement. Restrictions apply to the availability of these data, which were used under license for this study.

Author contribution. Conceptualization: R.J. Methodology: R.J; E.A. Data curation: E.A. Data visualisation: E.A; R.J. Writing original draft: E.A; R.J. All authors approved the final submitted draft.

Funding statement. This research was supported by a grant awarded to the first author from the USC Dornsife College of Letters, Arts, and Sciences.

Competing interest. The authors declare no competing interests exist.

References

Alizadeh M, **Weber I**, **Cioffi-Revilla C**, **Fortunato S and Macy M** (2019) Psychology and morality of political extremists: Evidence from Twitter language analysis of alt-right and Antifa. *EPJ Data Science 8*(1), 1–35.

Amin AB, **Bednarczyk RA**, **Ray CE**, **Melchiori KJ**, **Graham J**, **Huntsinger JR and Omer SB** (2017) Association of moral values with vaccine hesitancy. *Nature Human Behaviour 1*(12), 873–880.

Atari M, **Haidt J**, **Graham J**, **Koleva S**, **Stevens ST and Dehghani M** (2023) Morality beyond the WEIRD: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology 125*(5), 1157–1188. https://doi-org.libproxy2.usc.edu/10.1037/pspp0000470.

Bae Y and Lee H (2012) Sentiment analysis of Twitter audiences: Measuring the positive or negative influence of popular twitterers. *Journal of the American Society for Information Science and Technology 63*(12), 2521–2535.

Bahgat M, **Wilson SR and Magdy W** (2022) LIWC-UD: Classifying online slang terms into LIWC categories. In *Proceedings of the 14th ACM Web Science Conference*, pp. 422–432. Association for Computing Machinery (ACM). 14th ACM Web Science Conference 2022, Barcelona, Spain, 26/06/22. https://doi.org/10.1145/3501247.3531572.

Biddle N, **Edwards B**, **Gray M**, **Hiscox M**, **McEachern S and Sollis K** (2022) Data trust and data privacy in the COVID-19 period. *Data & Policy 4*, e1.

Borghouts J, **Huang Y**, **Gibbs S**, **Hopfer S**, **Li C and Mark G** (2023) Understanding underlying moral values and language use of COVID-19 vaccine attitudes on Twitter. *PNAS Nexus 2*(3), pgad013.

Bruns A and Liang YE (2012) Tools and methods for capturing Twitter data during natural disasters. *First Monday 17*(4), 1–8.

Chen E and Ferrara E (2023, June) Tweets in time of conflict: A public dataset tracking the Twitter discourse on the war between Ukraine and Russia. In *Proceedings of the International AAAI Conference on Web and Social Media*, vol. *17*, pp. 1006–1013.

Chen K, **Duan Z and Yang S** (2022) Twitter as research data: Tools, costs, skill sets, and lessons learned. *Politics and the Life Sciences 41*(1), 114–130.

Christie NC, **Hsu E**, **Iskiwitch C**, **Iyer R**, **Graham J**, **Schwartz B and Monterosso JR** (2019) The moral foundations of needle exchange attitudes. *Social Cognition 37*(3), 229–246.

Chung CK and Pennebaker JW (2018) What do we know when we LIWC a person? Text analysis as an assessment tool for traits, personal concerns, and life stories. In Zeigler-Hill V and Shackelford T (eds), *The Sage Handbook of Personality and Individual Differences*. Thousand Oaks, CA: Sage Publications, pp. 341–360.

Clifford S and Jerit J (2013) How words do the work of politics: Moral foundations theory and the debate over stem cell research. *The Journal of Politics 75*(3), 659–671. https://doi.org/10.1017/S0022381613000492.

Clifford S, **Jerit J**, **Rainey C and Motyl M** (2015) Moral concerns and policy attitudes: Investigating the influence of elite rhetoric. *Political Communication 32*(2), 229–248.

Frimer JA, **Boghrati R**, **Haidt J**, **Graham J and Dehgani M** (2019) Moral foundations dictionary for linguistic analyses 2.0. Unpublished manuscript.

Graham J and Haidt J (2012) Sacred values and evil adversaries: A moral foundations approach. In Mikulincer M and Shaver PR (eds), *The Social Psychology of Morality: Exploring the Causes of Good and Evil*. Washington, DC, USA: American Psychological Association, pp. 11–31. https://doi.org/10.1037/13091-001.

Graham J, **Haidt J**, **Koleva S**, **Motyl M**, **Iyer R**, **Wojcik SP** & **Ditto PH** (2013) Moral foundations theory: The pragmatic validity of moral pluralism. In Patricia Devine and Ashby Plant (eds), *Advances in Experimental Social Psychology*, vol. *47*. San Diego, CA: Academic Press, pp. 55–130.

Graham J, **Haidt J**, **Motyl M**, **Meindl P**, **Iskiwitch C and Mooijman M** (2018) Moral foundations theory: On the advantages of moral pluralism over moral monism. *Atlas of Moral Psychology 211*, 222.

Graham J, **Haidt J and Nosek BA** (2009) Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology 96*(5), 1029–1046.

Graham J, **Nosek BA**, **Haidt J**, **Iyer R**, **Koleva S and Ditto PH** (2011) Mapping the moral domain. *Journal of Personality and Social Psychology 101*(2), 366–385.

Grover T, **Bayraktaroglu E**, **Mark G and Rho EHR** (2019) Moral and affective differences in U.S. immigration policy debate on Twitter. *Computer Supported Cooperative Work (CSCW) 28*, 317–355.

Haidt J, **Graham J and Joseph C** (2009) Above and below left–right: Ideological narratives and moral foundations. *Psychological Inquiry 20*, 110–119.

Haidt J and Joseph C (2004) Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus 133*(4), 55–66.

Haq E-U, **Tyson G**, **Lee L-H**, **Braud T and Hui P** (2022) Twitter dataset for 2022 Russo-Ukrainian crisis. arXiv e-prints, arXiv-2203

Hopp FR, **Fisher JT**, **Cornell D**, *et al.* (2021) The extended moral foundations dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods 53*, 232–246. https://doi.org/10.3758/s13428-020-01433-0.

Kaur R and Sasahara K (2016, December) Quantifying moral foundations from various topics on Twitter conversations. In *2016 IEEE International Conference on Big Data (Big Data)*, pp. 2505–2512. IEEE.

Kertzer JD, **Powers KE**, **Rathbun BC and Iyer R** (2014) Moral support: How moral values shape foreign policy attitudes. *The Journal of Politics 76*(3), 825–840.

Kim B, **Cooks E and Kim SK** (2022) Exploring incivility and moral foundations toward Asians in English-speaking tweets in hate crime-reporting cities during the COVID-19 pandemic. *Internet Research 32*(1), 362–378.

Koleva SP, **Graham J**, **Iyer R**, **Ditto PH and Haidt J** (2012) Tracing the threads: How five moral concerns (especially purity) help explain culture war attitudes. *Journal of Research in Personality 46*, 184–194.

La Gatta V, **Wei C**, **Luceri L**, **Pierri F and Ferrara E** (2023, April) Retrieving false claims on Twitter during the Russia-Ukraine conflict. In *Companion Proceedings of the ACM Web Conference* 2023, pp. 1317–1323.

Lämmerhirt D, **Micheli M and Schade S** (2024) Exploring the practices of "data-driven innovation" in the European public sector. *Data & Policy 6*, e24.

Liu X and Dijk M (2022) The role of data in sustainability assessment of urban mobility policies. *Data & Policy 4*, e2. https://doi.org/10.1017/dap.2021.32.

LIWC.app. (n.d.). How it works. Retrieved from https://www.liwc.app/help/howitworks

Mueller H, **Rauh C and Seimon B** (2024) Introducing a global dataset on conflict forecasts and news topics. *Data & Policy 6*, e17. https://doi.org/10.1017/dap.2024.10.

Murphy M, **Sharpe E and Huang K** (2024) The promise of machine learning in violent conflict forecasting. *Data & Policy 6*, e35. https://doi.org/10.1017/dap.2024.27.

https://doi.org/10.1017/dap.2024.98 Published online by Cambridge University Press

Nasim M, **Sharif N**, **Bhandari P**, **Weber D**, **Wood M**, **Falzon L and Kashima Y** (2022, December) Investigating language use by polarised groups on Twitter: A case study of the Bushfires. In *Proceedings of the 26th Australasian Document Computing Symposium*, pp. 1–7.

Parmelee JH, **Roman N and Beasley B** (2024) Moral framing in Ukraine war coverage. *Media, War & Conflict*, 17506352241264197.

Reimer NK, **Atari M**, **Karimi-Malekabadi F**, **Trager J**, **Kennedy B**, **Graham J and Dehghani M** (2022) Moral values predict county-level COVID-19 vaccination rates in the United States. *The American Psychologist 77*(6), 743–759.

Ronzhyn A and Wimmer MA (2021) Research directions in policy modeling: Insights from comparative analysis of recent projects. *Data & Policy 3*, e13. https://doi.org/10.1017/dap.2021.8.

Rowe F, **Mahony M**, **Graells-Garrido E**, **Rango M and Sievers N** (2021) Using Twitter to track immigration sentiment during early stages of the COVID-19 pandemic. *Data & Policy 3*, e36.

Sagi E and Dehghani M (2014) Moral rhetoric in Twitter: A case study of the U.S. federal shutdown of 2013. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. *36*, no. 36.

Schmitz M, **Muric G and Burghardt K** (2023, June) Detecting anti-vaccine users on Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, vol. *17*, pp. 787–795.

Smetana M and Vranka M (2021) How moral foundations shape public approval of nuclear, chemical, and conventional strikes: New evidence from experimental surveys. *International Interactions 47*(2), 374–390.

Sylwester K and Purver M (2015) Twitter language use reflects psychological differences between democrats and republicans. *PLoS One 10*(9), e0137422.

Tatalovich R and Wendell DG (2018) Expanding the scope and content of morality policy research: Lessons from moral foundations theory. *Policy Sciences 51*(4), 565–579.

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

Wendell DG and Tatalovich R (2021) Classifying public policies with moral foundations theory. *Policy Sciences 54*, 155–182.

Cite this article: Amjadi E and John RS (2025). Count on Me: Moral Language in Social Media and Policy Discourse during the Ukraine-Russia Conflict. *Data & Policy*, 7: e22. doi:10.1017/dap.2024.98