Supplementary Information (SI) for

Online engagement between opposing political protest groups via social media is linked to physical violence of offline encounters

1 - Sampling, Data Collection and Processing, and Events Selected for Analysis

2 - Violence Measurements and Item Response Theory

3 - Correlation between text analysis measures

4 - Best Linear Unbiased Predictor Aggregation

5 - Robustness Checks

6 - Neural Network Details

7 - Further exploratory analysis

1 - Sampling, data collection and processing, and events selected for analysis

A two-stage process was conducted to identify cases of events that could be used for analysis. For the United Kingdom sample, the two largest and most active street-protest groups from each side of the political spectrum were identified; these were English Defence League (EDL) and Britain First (BF) from the far-right, and Unite Against Fascism (UAF) and Anti-Fascist Network (AFN) from the left. For each of these groups an initial sample of events where the group participated was selected using a historical open-source search of past events hosted on the groups' Facebook page. Criteria for inclusion were that the events must have a corresponding counter-march or protest by the opposing political side, organized in tandem, on the same day and at the same location. This method resulted in the selection of 18 events. Following this, an additional search was conducted more broadly across reports of political violence in the United Kingdom to identify events organized by alternative groups (often localized events or events held under alternative banners). This resulted in an additional two events being identified. Once identified the same criteria for inclusion were used. Importantly. we have not sought to define groups ourselves into political ideology, but rather relied on existing formalizations. At the point of collection, this dataset represented the entirety of the publicly available data of this type from the United Kingdom.

For the United States sample five events that took place in the summer of 2017 were selected. All of these cases occurred in response to the 'Unite the Right' rally in Charlottesville (the data from the Charlottesville protest itself is not publicly available as Facebook removed the pages prior to the event. We contacted Facebook to see if they would be willing to share this data, but they did not respond.) These cases were selected as they bore a strong resemblance to the United Kingdom based protests, far-right marches taking place with an opposing left-wing counter protest.

Once the political events were identified, we collected the conversations taking place online using a Python (version 2.7) script connected to the Facebook Graph Application Program Interface (API). The process consisted of two stages. Once a Facebook event page had been identified then the unique identifier for this page was gathered. This identifier was then used to collect the content of each 'status' posted to the page along with the unique identifier for this status. This generated a first dataset. The second stage used this dataset as an input, and collected all the comments posted below each status. This produced a second dataset. Finally, the two datasets were combined to produce the full set. There was a single case of a comment being replicated 95 times in quick succession in the dataset. This was presumed to be due to a glitch in either posting to the social network or in data collection and so only a single version of this comment was included in subsequent analysis.

All comments and statuses were given equal weighting. In some cases, multiple event pages were identified for each side (left or right wing) from multiple organizing groups within the same protest. In these cases, the event page comments were collected separately and the data then compiled into a single set for either the left or the right groups. The final list of events included in the analysis is shown in Table S1. This data collection method gathered an average of 1473 comments per event page, giving a total of 2946 comments on average per event and 73,632 comments in total. Once collected, the data were cleaned to remove any blank entries (posts containing only images or video and no text) and to remove any conversation that occurred after the planned start time of the event. In doing this we can safely assume that any violence at the later event had no impact on earlier conversation online.

Events selected for analysis

Table S1 shows the events which were selected for inclusion in the current study, along with the groups associated with the right-wing and left-wing sides. Table S2 shows the types of data source used to generate violence metrics for each event.

Table S1. Real world protests, marches and rallies selected for analysis

Date	Location	Country	Right Wing Group(s)	Left Wing Group(s)
05/09/2015	Rotherham	UK	Britain First	Rotherham Unite Against Fascism
17/10/2015	Burton	UK	Britain First	East Staffordshire Trades Council
30/01/2016	Dewsbury	UK	Britain First	We Are Dewsbury
16/07/2016	London	UK	English Defense League	Unite Against Fascism (UAF)
06/08/2016	Nottingham	UK	English Defense League	Midlands Anti-Fascist Network
24/09/2016	Newcastle	UK	English Defense League	Newcastle Unites
05/11/2016	Telford	UK	English Defense League	ShropRad
25/02/2017	Rotherham	UK	English Defense League	Rotherham Unite Against Fascism
25/02/2017	Telford	UK	Britain First	Shropshire and Telford Trades Council & UAF
01/04/2017	London	UK	English Defense League & Britain First	Anti-Fascist Network (AFN) & Unite Against Fascism
08/04/2017	Birmingham	UK	English Defense League	Birmingham Unite Against Fascism
15/04/2017	Wishaw	UK	Scottish Defense League	Unite Against Fascism Scotland
03/06/2017	Liverpool	UK	English Defense League	Merseyside Unite Against Fascism
11/06/2017	Manchester	UK	UK Against Hate	Stand-Up to Racism
24/06/2017	Birmingham	UK	Britain First	Birmingham Unite Against Fascism
22/07/2017	Rochdale	UK	Britain First	Unite Against Fascism
29/07/2017	Rochdale	UK	English Defense League	Unite Against Fascism
13/08/2017	Seattle	U.S.	Patriot Prayer	Greater Seattle IWW General Defense Committee Local 24
19/08/2017	Boston	U.S.	Boston Free Speech Coalition	Black Lives Matter Network, Black Lives Matter Cambridge, and Black Lives Matter Boston
26/08/2017	San Francisco	U.S.	Patriot Prayer	Peace Love & Understanding (among others)
27/08/2017	Berkeley	U.S.	No Marxism in America	SAFEBay - Solidarity Against Fascism East Bay
02/09/2017	Keighley	UK	English Defense League	Unite Against Fascism
10/09/2017	Portland	U.S.	Patriot Prayer	Rose City Antifa, Queer Liberation Front
21/10/2017	Peterborough	UK	English Defense League	Peterborough Trades Union Council - PTUC
04/11/2017	Bromley	UK	Britain First	Stand Up To Racism - South East London and Unite Against Fascism

Table S2. Types of data source used for violence measurement at each real-world event.

			Data Sources used to generate offline violence metric					
Date	Location	Country	Police Reports	Mainstream National News	Local News	Citizen Journalism (Blogs)	Photo Journalism (e.g. PhotoBucket)	Crowdsourced Video (e.g. YouTube)
05/09/2015	Rotherham	UK	✓	✓	✓		✓	✓
17/10/2015	Burton	UK		✓	✓	✓		✓
30/01/2016	Dewsbury	UK		✓	✓			✓
16/07/2016	London	UK		✓	✓		✓	✓
06/08/2016	Nottingham	UK	✓	✓				✓
24/09/2016	Newcastle	UK	✓	✓	✓			✓
05/11/2016	Telford	UK	✓	✓	✓		✓	✓
25/02/2017	Rotherham	UK	✓	✓	✓			
25/02/2017	Telford	UK	✓	✓	✓			✓
01/04/2017	London	UK	✓	✓			✓	✓
08/04/2017	Birmingham	UK	✓	✓	✓		✓	✓
15/04/2017	Wishaw	UK		✓	✓			✓
03/06/2017	Liverpool	UK	✓	✓	✓			✓
11/06/2017	Manchester	UK	✓	✓		✓		✓
24/06/2017	Birmingham	UK		✓	✓			✓
22/07/2017	Rochdale	UK			✓			✓
29/07/2017	Rochdale	UK			✓		✓	✓
13/08/2017	Seattle	U.S.	✓	✓	✓	✓	✓	✓
19/08/2017	Boston	U.S.	✓	✓	✓		✓	✓
26/08/2017	San Francisco	U.S.		✓	✓			
27/08/2017	Berkeley	U.S.	✓	✓	✓		✓	✓
02/09/2017	Keighley	UK		✓	✓			✓
10/09/2017	Portland	U.S.	✓	✓	✓			✓
21/10/2017	Peterborough	UK		✓	✓		✓	✓
04/11/2017	Bromley	UK			✓	✓	✓	✓

2 - Violence measurements and item response theory

Identifying violent and non-violent events

Our initial measure demonstrates whether or not meaningful violence was sparked, but not the level or degree of this violence. The primary marker for this was whether the event was described afterward in reports as violent. For example: 'Violent clashes occurred at event x' would be classified as violent, while 'a peaceful protest took place', would be classified as peaceful.

Degree of violence

The second measure takes into account the same evidence but uses a range of indicators (Table S3) to classify more specifically how violent an event was depending on how many of these indicators were present. The following indicators were used: heavy police presence (indicating an expectation of violence), physical contact between groups, throwing of projectiles, bloodshed/actual bodily harm, arrest figures (1 indicator for minor arrests, up to 5, and an additional indicator for major arrests of 5+), and violence leading to a fatality. These indicators were selected following advice from analysts at an industry leading security firm familiar with the subject matter. This organization was not involved with analysis subsequently to providing this guidance.

We do not distinguish whether each indicator of violence occurred specifically between members of opposing groups, in contact with the police, or even within a group itself. While delving into more detail on the type and nature of violence at events may be interesting, there are limitations that prevent this, and the information is not always available. In many cases reports after the event did not specify between whom the violence took place and attempts to trace violence observed in photos or videos back to a Facebook profile to determine group membership go far beyond the scope of this study. Instead as violence between protesters and law enforcement, and violence between groups, is likely to be correlated (with one type of violence encouraging the other) we no do distinguish violence based on the recipient of the action, purely based on the presence of the indicator. The number of occurrences of each type of violence are shown in table S3 and the distribution of indicators across events in shown in table S4.

These indicators were weighted using a latent trait model using the 'ltm' package in R (Rizopoulos, 2017), which his creates a continuous scale of violence on the z scale (M = 0, SD = 1). The model is based on the relationship between performance on single item measures compared to overall performance. In the current case, this compares individual indicators of violence to the overall violence score measured with all indicators, without assuming that each item represents an equal level of violence. This results in an estimate of the 'severity' ('difficulty', in latent trait parlance) of each violence indicators on the latent violence scale, with the indicators indicative of more severe violence placed higher on the violence scale.

From the results, we can extract the severity parameter for each indicator, and calculate a per-event violence score on a continuous scale. Fig S1 shows the item characteristic curves for the indicators of violence with coefficients shown in Table S3. The occurrence column gives the total number of times this indicator was reached across all events, with a maximum possible score of 25 if an indicator was present at all events. Clearly, the LTM orders the violence indicators in a sensible way, with heavy police presence being triggered at the lowest level of violence, and bloodshed/bodily harm indicating the most violence. Moreover, the items covered a wide range of latent violence scores (*z*s from -1.28 to 0.74) and all indicators discriminate very well (discrimination parameters >= 1.77, (Baker, Boston, & Rudner, 2001)), suggesting they yield reliable information about violence. We calculated the latent event-level violence scores using these indicators, which were normally distributed according to a visual inspection of the density distribution. This violence measure was then used as the dependent variable in the linear model with the conversation quality metrics outlined in the main text.

Table S3. Violence Indicators, Violence Coefficients

Reference	Indicator of Violence	Violence *z*-score	Discrimination	Occurrence (max = 25)
g	Heavy Police Presence	-1.28	11.37	22
d	Minor arrests (1-5 Arrests)	-1.05	2.47	20
c	Projectiles / Smoke Bombs	-0.11	2.49	13
a	Physical contact between opposing groups	0.22	3.43	10
e	Major arrests (5+)	0.35	3.09	9
b	Bloodshed / Bodily Harm	0.74	3.42	6
f	Violence leading to a Fatality	n/a	n/a	0

(Note: Indicator 'f', Violence leading to a fatality, did not occur at any sampled events, and so did not receive an associated violence score)

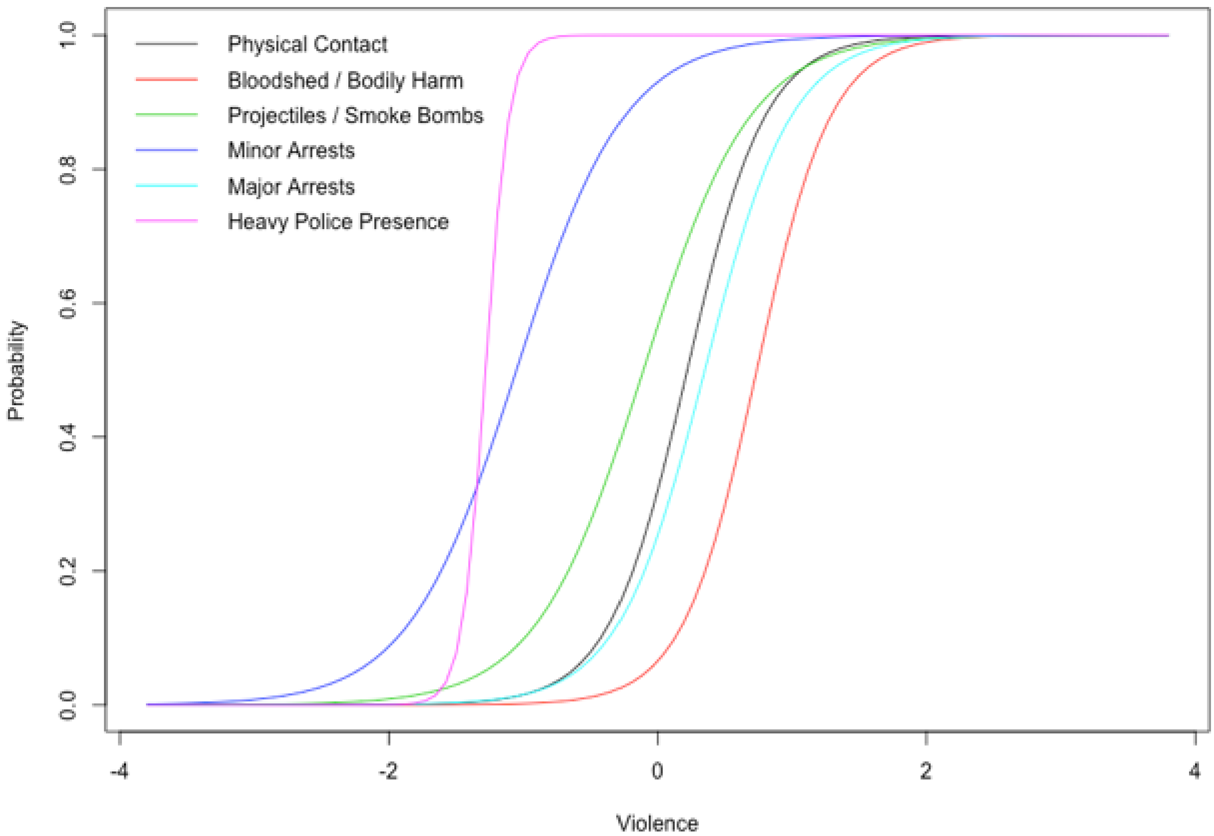

Fig S1. Item Characteristic Curves for the six occurring indicators of violence.
Higher inflection points indicate greater violence is needed for the indicators to occur

Table S4. Indicators of violence observed at each real-world event.

			Indicator of Violence						
Date	**Location**	**Country**	**a**	**b**	**c**	**d**	**e**	**f**	**g**
05/09/2015	Rotherham	UK	✓	✓	✓	✓	✓		✓
17/10/2015	Burton	UK				✓	✓		✓
30/01/2016	Dewsbury	UK				✓			✓
16/07/2016	London	UK							
06/08/2016	Nottingham	UK			✓	✓			✓
24/09/2016	Newcastle	UK				✓			✓
05/11/2016	Telford	UK				✓			
25/02/2017	Rotherham	UK				✓	✓		✓
25/02/2017	Telford	UK		✓	✓	✓			✓
01/04/2017	London	UK	✓	✓	✓	✓	✓		✓
08/04/2017	Birmingham	UK	✓			✓			✓
15/04/2017	Wishaw	UK			✓				✓
03/06/2017	Liverpool	UK	✓		✓	✓	✓		✓
11/06/2017	Manchester	UK	✓	✓	✓	✓	✓		✓
24/06/2017	Birmingham	UK			✓	✓			✓
22/07/2017	Rochdale	UK							✓
29/07/2017	Rochdale	UK							✓
13/08/2017	Seattle	U.S.	✓		✓	✓			✓
19/08/2017	Boston	U.S.	✓		✓	✓	✓		✓
26/08/2017	San Francisco	U.S.				✓			✓
27/08/2017	Berkeley	U.S.	✓	✓	✓	✓	✓		✓
02/09/2017	Keighley	UK			✓	✓			✓
10/09/2017	Portland	U.S.	✓	✓	✓	✓	✓		✓
21/10/2017	Peterborough	UK	✓			✓			✓
04/11/2017	Bromley	UK							

3 – Correlations between text analysis measures

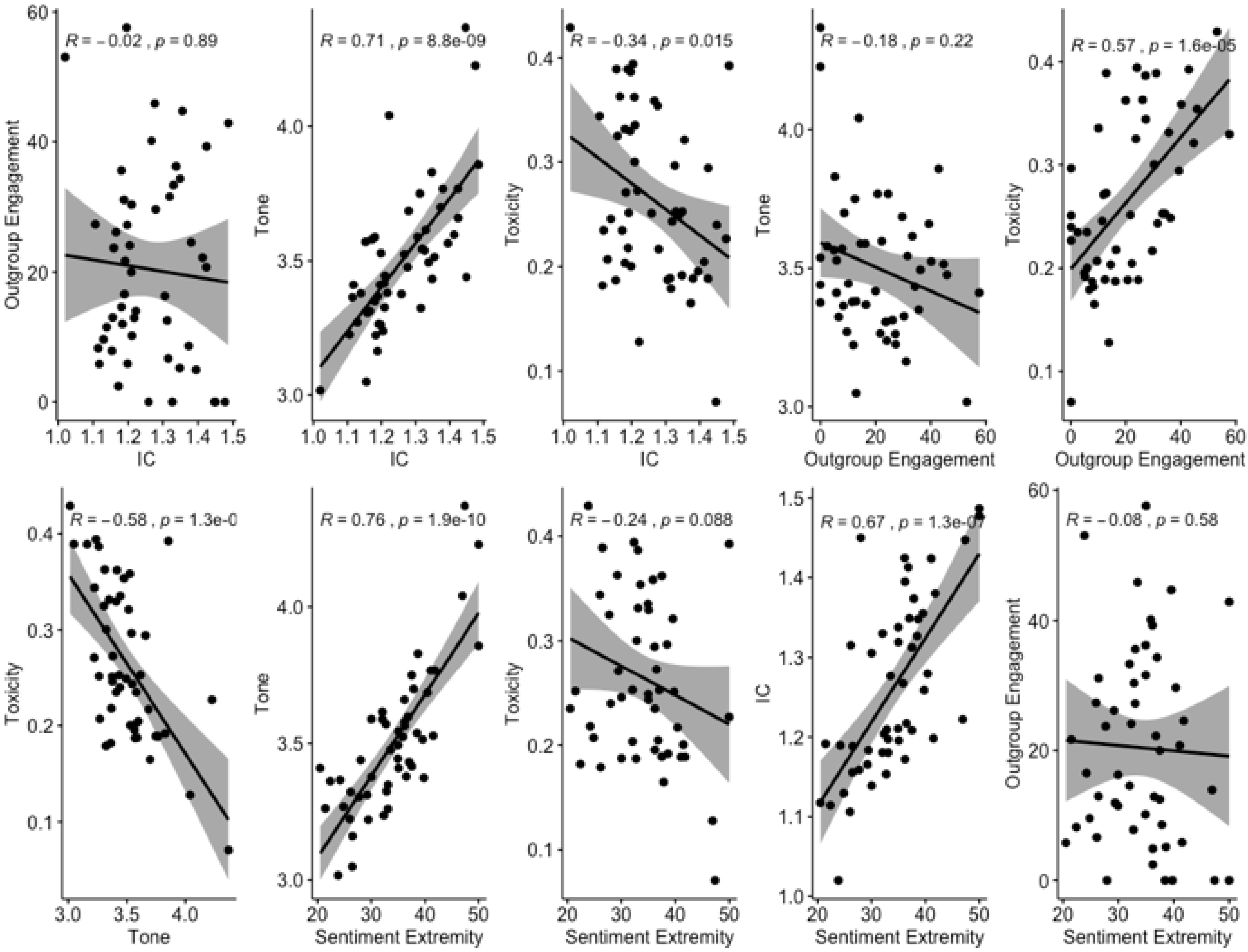

Fig S2. Correlation plots between text analysis measures used in the study.
Significant correlations were found between a number of the measures and therefore multicollinearity in models was checked using variance inflation factor (VIF) . Shaded areas represent 95% confidence interval.

Table S5. Correlation matrix among text analysis measures used in the study

Measure	1	2	3	4	5	6
1 – Sentiment – Tone						
2 – Sentiment – Extremity	0.76***					
3 – Toxicity	-0.58***	-0.24				
4 – Integrative Complexity	0.71***	0.67***	-0.34*			
5 – Outgroup Engagement	-0.18	-0.08	0.57***	-0.02		
6 – Degree of Violence	0.22	0.33	0.15	0.12	0.52**	
Minimum	3.02	20.51	0.07	1.02	0.0	
Maximum	4.37	50.00	0.43	1.49	57.59	
Mean	3.50	34.03	0.26	1.26	20.44	
SE	0.04	0.99	0.01	0.02	2.12	

4 - Best linear unbiased predictor aggregation

In order to account for the fact that the predictors (outgroup engagement, tone, sentiment extremity, integrative complexity and toxicity) were measured at the conversation level but violence was measured at the event level, we aggregated scores for all group variables to the event level. To do this we calculated adjusted means for each variable using the best linear unbiased predictor (BLUP) as in prior research (Croon & Van Veldhoven, 2007).This is required due to the effect that when predicting outcomes at the higher level from predictors nested within a lower level it is statistically biased to directly regress the higher level outcome variable on the unadjusted means of lower level predictors. The correction is shown to yield unbiased estimates of the parameters (Becker, Breustedt, & Zuber, 2017).

This analysis was done using the 'MicroMacroMultilevel' package in R (Becker et al., 2017; Lu, Page-gould, & Xu, 2017). The BLUP aggregations for the event level variables were then used as predictors of the presence/absence of violence within a GLM with a binomial distribution and a logit link function, and as predictors of the degree of violence with a linear model.

5 - Robustness checks

We checked the robustness of models by performing multicollinearity checks though the calculation of variance inflation factors (VIF), checking for the absence of influential data points, and inspecting the linearity, normality of residuals, and homoscedasticity through a visual inspection of residual plots (histogram, Q-Q and rotated fitted values vs residuals).

Sensitivity analysis

To test the robustness of the degree of violence measure, we performed a sensitivity analysis. In this test each indicator of violence (a-f) was removed from analysis in turn, the degree of violence for each event re-calculated using IRT, and the main effect of outgroup engagement on degree of violence tested. In all cases the effect was still found to be significant. These results are shown in Table S6. We conclude from this that the measure is robust and not reliant on any single indicator of violence.

Table S6. Sensitivity Analysis of degree of violence

Indicator Removed	t Value	*p* Value
None	3.20	0.0031 **
Physical Contact	2.99	0.0066 **
Bloodshed / Bodily Harm	3.69	0.0012 **
Projectiles / Smoke Bombs	3.29	0.0032 **
1-5 Arrests	2.99	0.0066 **
5+ Arrests	3.23	0.0037 **
Heavy Police Presence	3.15	0.0045 **

6 - Neural network text classifier for outgroup engagement

In order to classify comments into ingroup directed and outgroup directed, we trained a neural network machine learning classifier using a supervised learning method. The training set for this network consisted of 1,000 randomly selected comments from the entire dataset of available comments. Each comment was classified at either ‘within-group’ for comments that were directed towards other ingroup members, or ‘between-group’ for comments that were directed towards a member of the outgroup. Therefore a ‘between-group’ comment could either be a member injecting a comment into the event page of the opposing group, or a reply to this injection from a member of the incumbent page.

Group membership was identified from self-disclosed information in the content of the comments that a user posted to the Facebook page linked to an explicitly right-wing or left-wing political group. From the content of these comments, it is possible to infer either support for the event or opposition. These comments were manually coded by the first author, who is familiar with the right-wing/ left-wing online environment. Each comment was coded in isolation; however, common themes and patterns were identified and applied throughout the dataset. The default coding option was within-group communication, and so if a decision could not be made then this was the option selected. Additionally, for comments that did not contain enough information (such as photos which would be represented textually as [[PHOTO]]), the default coding option was selected.

This set was evenly balanced such that 500 comments were identified as ingroup and 500 comments were identified as outgroup. The entire dataset contains a higher proportion of ingroup compared to outgroup comments, and so in order to gather this balanced training set, over 1000 comments were classified until at least 500 ingroup and 500 outgroup comments were identified, and then a random selection of these taken.

The test dataset contained a second random sample of 1000 comments. This set was not manually balanced between ingroup and outgroup communication and so reflects the true balance with the entire sample. These comments were manually coded in the same way.

To ensure accuracy of the human coding, all comments from the training and test set were coded by a second coder who was blinded to the hypotheses, and inter-coder reliability (ICR) scores calculated. For the training set the ICR was 97.8% with a Scott's PI of 0.956. For the test set the ICR was 95.9%% with a Scott's PI of 0.895. These values were deemed high enough to be confident in the original classifications.

Ethical considerations prevent further profiling of individual users beyond the self-disclosed content of each message posted to the event page and all conversation data was anonymized at the point of collection. We made no attempts to link comments to individual Facebook profiles. Additionally, ethical constraints restrict accessing any information that is not made available via the public Facebook API. This information contains only the content of the post, the author, the date/time it was posted and any reactions to it. This means that even if conversations were not anonymized, we would not be able to gather any additional information about the participants beyond this. It is therefore possible, due to this anonymization, that some ingroup critique was coded as outgroup critique as the judgement was made solely on the basis of what was said in the comment. However, in most cases this coding was straightforward for the human coders to complete, and certain words and phrases used only by outgroup members clearly highlighted outgroup engagement (calling someone offensive or derogatory terms for example is unlikely to come from an ingroup member). The high level of agreement between both human coders (97.8% for the training set, 95.9% for the test set) reflects this.

The neural network itself was created using the natural language toolkit (NTLK) library with the Python coding environment and using a 'bag of words' approach (Pinto, 2017). For the entire training set each word was tokenized, stemmed, converted to lower-case, and then added to a dictionary or 'bag of words' with all non-standard characters removed. The model then learns over time the associated word patterns that occurred during ingroup and outgroup

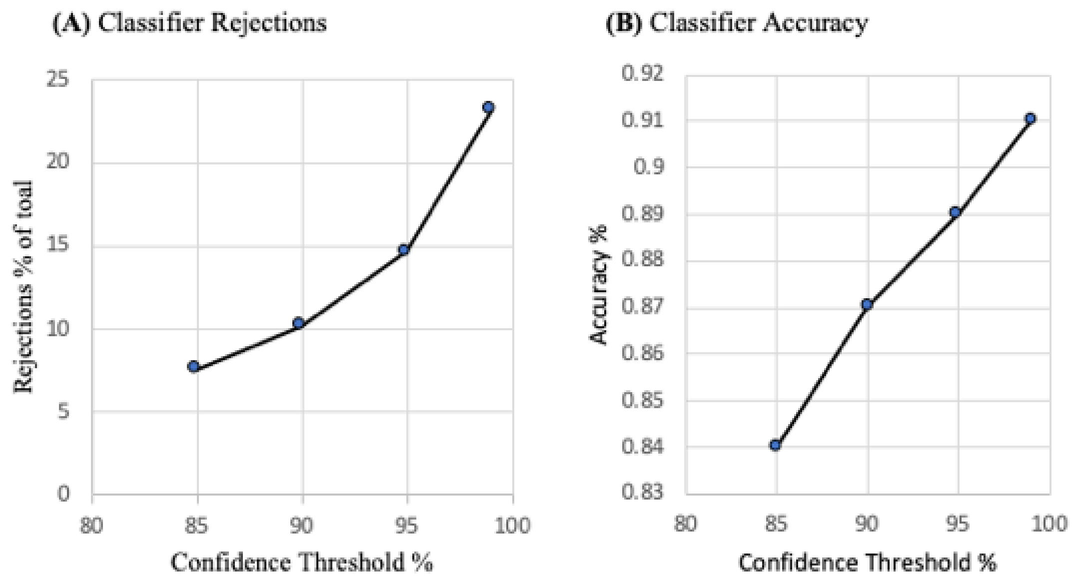

Fig S3 – (A) The proportion of comments within the test set which were rejected at each confidence threshold. (B) The overall classifier accuracy at each confidence threshold

communications. The neural network contained two layers of neurons, with one hidden layer of 20 neurons and was run over 100,000 iterations and an alpha reinforcement learning rate of 0.1.

When classifying comments the network gives a classification along with a confidence judgement for each entry. This confidence value is used to create a threshold below which comments would not be included in further analysis. A confidence threshold of 95% was selected as the best trade-off between minimizing false positives whist avoiding too many unclassified comments (Fig S3). At this threshold, 14.7% of comments were excluded due to low confidence.

Applying this confidence threshold led to an overall accuracy for the classifier of 89.0%, with a sensitivity of 85.9% and a specificity of 89.9%. (Therefore this is a conservative judgement classifier with regard to outgroup classification, reflecting the conservative nature of setting ingroup contact as the default classification setting). The baseline accuracy figures without a confidence threshold applied are 84%, with a sensitivity of 84% and a specificity of 88%. By way of comparison the second human coder in the inter-coder reliability test achieved 97% accuracy, with a specificity and sensitivity both of 97% when compared to the initial coder across both the training and test set.

It is important to note that the network classifies each comment as if it were occurring in isolation within the dataset – this is then adjusted using two subsequent rules to increase accuracy of the network. Firstly, a proximity rule is applied whereby if a comment is both preceded and followed by a case of outgroup engagement, then its classification is changed to engagement. This was designed to make the classification system more aware of the overall conversation. Secondly, the classification was manually corrected by a human coder (the same human coder from the original classifications). The aim of this was not to reclassify all cases, but rather to check for obvious errors, long strings of replicated comments or cases whereby the initial proximity rule was applied incorrectly. These two rules accounted for only small changes in the contact figures, with only 2.7% of the comments of each event page adjusted on average at stage 1 and 3.0% at stage 2. Running the main analysis (outgroup engagement on degree of subsequent violence) without these two corrections did not alter the conclusions.

7 - Further exploratory analysis

Group comparisons, left vs. right

In order to investigate differences in conversations occurring between left-wing and right-wing group pages we performed comparisons using linear mixed models (LMMs) with the lme4 package. We investigated association of sentiment (tone and sentiment extremity),

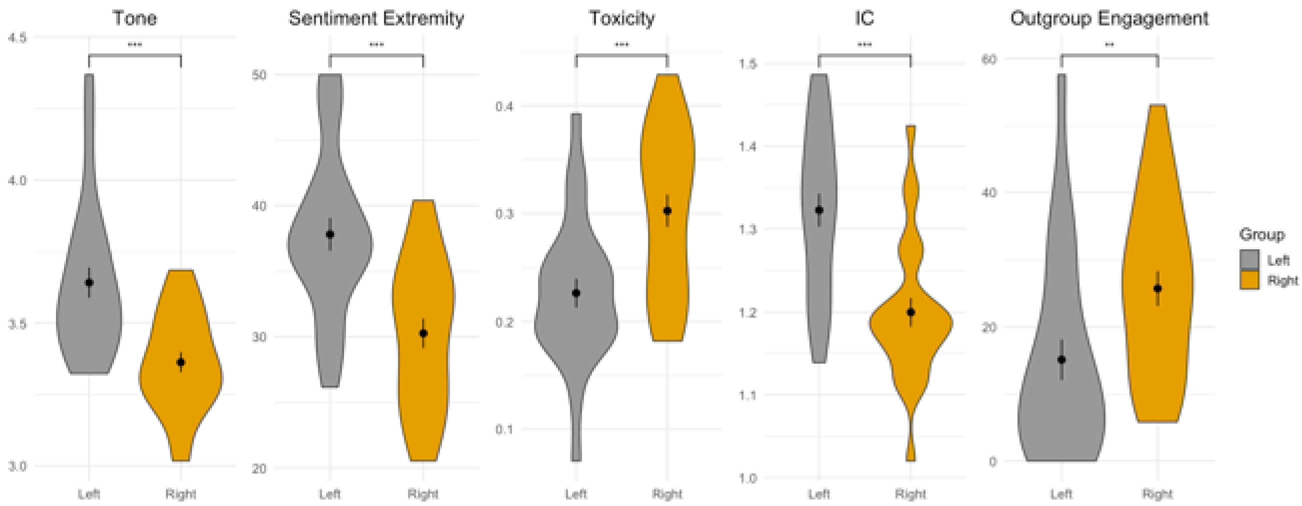

Fig S4. Comparison of right-wing vs left-wing conversation qualities.
This figure demonstrates the differences in Tone, Sentiment Extremity, Toxicity, Integrative Complexity and outgroup engagement occurring on right-wing and left-wing Facebook event pages. Mean and standard error is given by dots and lines, respectively. *** denotes $p < 0.001$, ** $p < 0.01$.

outgroup engagement, toxicity, and integrative complexity across the two groups, and included event ID as a random effect. Significance levels of fixed effects (and interactions) were obtained by comparing the full model to the null model with a χ^2 test.

Results showed that right-wing pages (n=25) had higher toxicity scores (indicating a more toxic conversation) than left-wing pages (Right-wing: M=0.30±0.02, Left-wing: 0.23±0.01, LMM, $\chi_1^2 = 12.93$, p <0.001) and a lower average value of sentiment extremity (Right-wing: M=30.27±1.13, Left-wing: 37.80±1.25, LMM, $\chi_1^2 = 17.32$, p <0.001). Conversations within right-wing pages displayed lower level of Integrative Complexity (Right-wing: M=1.20±0.02, Left-wing: M=1.32±0.002, LMM, $\chi_1^2 = 19.19$, p <0.001) and lower tone scores (Right-wing: M=3.36±0.03, Left-wing: M=3.64±0.05, LMM, $\chi_1^2 = 16.94$, p <0.001). Finally, right-wing pages showed a larger degree of outgroup engagement than left-wing pages (Right-wing: M=25.74%±2.61, Left-wing: M=15.13%±3.02, LMM, $\chi_1^2 = 7.39$, p = 0.007). These results are shown in Fig S4 and Table S7.

This echoes previous evidence showing that left-leaning politicians tend to display higher levels of IC than right-leaning politicians (Tetlock, 1983). Additionally, integrative complexity in online communication can provide information about the extent to which individuals hold radical or extremist views (Smith, Suedfeld, Conway, & Winter, 2008), therefore suggesting that it may the content on the right-wing pages in the current study which displays more extreme views.

The result that outgroup engagement is more likely to occur in the conversations on the event page hosted by a right-wing group suggests that it is individuals from the left-wing groups who are more likely to ‘seek out’ this contact by visiting the opposing page and engaging in discussion.

Table S7. Differences in conversation metrics between right-wing and left-wing event pages (Mean ± SE)

Group	**Tone**	**Sentiment Extremity**	**Toxicity**	**IC**	**Outgroup Engagement**
Left	3.64 ± 0.05	37.8 ± 1.25	0.23 ± 0.01	1.32 ± 0.02	15.1 ± 3.02
Right	3.36 ± 0.04	30.3 ± 1.13	0.30 ± 0.02	1.20 ± 0.02	25.7 ± 2.61

SI References

Baker, F., Boston, C., & Rudner, L. (2001). *The Basics of Item Response Theory* (2nd ed.). ERIC Clearning House on Assesment and Evaluation.

Becker, D., Breustedt, W., & Zuber, C. I. (2017). Surpassing simple aggregation : Advanced strategies for analyzing contextual-level outcomes in multilevel models. *Methods, Data, Analyses*, 1–31. https://doi.org/10.12758/mda.2017.05

Croon, M. A., & Van Veldhoven, M. J. P. M. (2007). Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model. *Psychological Methods, 12*(1), 45–57. https://doi.org/10.1037/1082-989X.12.1.45

Lu, J. G., Page-gould, E., & Xu, N. R. (2017). Package ' MicroMacroMultilevel .' Retrieved from https://cran.r-project.org/web/packages/MicroMacroMultilevel/MicroMacroMultilevel.pdf

Pinto, A. (2017). Text Classification using Neural Networks , Machine Learnings. Retrieved April 12, 2018, from https://machinelearnings.co/text-classification-using-neural-networks-f5cd7b8765c6

Rizopoulos, D. (2017). Package ' ltm .' Retrieved from https://cran.r-project.org/web/packages/ltm/ltm.pdf

Smith, A., Suedfeld, P., Conway, L., & Winter, D. (2008). The language of violence: distinguishing terrorist from nonterrorist groups by thematic content analysis. *Dynamics of Asymmetric Conflict, 1*(2), 142–163. https://doi.org/10.1080/17467580802590449

Tetlock, P. E. (1983). Cognitive style and political ideology. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.45.1.118