How People Talk about Armed Conflicts

Jeremy Cole, Ying Xu, and David Reitter

The Pennsylvania State University
College of Information Science and Technology
University Park, Pennsylvania, USA
{jrcole,ying.xu,reitter}@psu.edu

Abstract. Armed conflicts around the world produce displacement, injury, and death. This study examines how anonymous and pseudonymous Internet commenters discuss such conflicts. Specifically, we ask how permissible it is to express positive or negative sentiments about these conflicts as a function of variables including region, conflict nature, and severity. Data from the Armed Conflicts Database is aggregated to identify a number of potential factors that may influence views on acceptable sentiments. A large-scale sample of the Reddit comment forum was coded for positive and negative sentiments using sentiment analysis techniques. Permissibility is judged using the Reddit voting features. Regressions reveal that positive sentiments are found not permissible for higher numbers of fatalities, and that negative sentiments are found permissible for certain regions, more permissible for older conflicts, and less permissible for territorial conflicts. A number of alternative, non-correlated predictors are presented.

Keywords: Behavioral and Social Sciences; Corpus Linguistics; GLMM; Armed Conflicts; Public Opinion

1   Introduction

According to the Armed Conflict Database, there are 42 active conflicts around the world causing 180,000 fatalities and more than 12 million refugees. For people who live in relatively peaceful areas, such as in North America, their perception, opinion and action towards these international crises are important. Collectively, these perceptions are influential in shaping their countries' foreign policy [7].

Traditional media and social media interact to help form these perceptions. On the other hand, consumers of media implicitly or explicitly rate the quality of the content produced. Contributors and evaluators naturally have biases that can build on each other, if contributions that follow these biases are rated more positively than those that do not. The question is what are those biases to begin with?

In particular, we seek to answer this question in the context of armed conflicts. For instance, how do people talk, evaluate, and contribute to the discussion of various armed conflicts? Specifically, we want to understand when and why

people accept negative discussions and positive discussions. We examine the effect of features such as the number of fatalities, refugees and Internally Displaced People (IDP). Further, we investigate a variety of characteristics that may influence objectivity, such as the location of the conflict. We hope to discover the relationship between these features and the view of specific armed conflicts.

To accomplish this, we turn to the Reddit Internet forum. Reddit's relative homogeneity in both demographics and opinion give it certain advantages over other social media. In a significant way, Reddit represents young Western people. Further, their opinions on these armed conflicts are also fairly homogeneous, as will be demonstrated later. This gives us an ideal dataset to reflect on cultural judgments of acceptability.

2   Background

2.1   Public Perception of Armed Conflicts

We investigated past research on the perception of armed conflicts from the perspectives of the general public. Public opinion on armed conflicts from World War II to Vietnam, the Gulf, Afganistan and Iraq war have also been investigated to find patterns of public responses to international conflicts [2]. Particularly, with the pervasive access of the Internet and use of social media, public opinions are becoming more influential on political decision-making; they are even changing the political cycle [16]. However, systematic studies of public opinion from social media on armed conflicts is limited. Studies focusing on social media in armed conflicts often regard social media tools as agents or platforms to express views and coordinate actions: for example, arranging protests and organizing uprisings [10].

Public perception of armed conflicts through both traditional and social media suffers from several limitations. Advanced communication technologies make it possible to disseminate information about various conflicts around the world [14]. However, government and news organizations cover and frame conflicts in a certain way that adds their own bias. For example, when covering foreign events, reporters largely focus on war, terrorism and political violence [13]. Even still, most of the world's conflicts are largely unreported by the media [8]. Even if it is unintentional, this causes a significant limitation for the public to be able to obtain the broad-spectrum of information they need to evaluate conflicts objectively [11]. News agencies in general focus on news in their own country or nearby countries, shrinking the amount of time focused on international coverage [15].

Despite these biases, growing studies are covering the relationship between the characteristics of armed conflicts and the public perception of them. Berinsky has found that public reactions to conflicts have been shaped less by their defining characteristics, such as fatalities and resources costs, than by political affiliations [2]. In other words, if your political party supports the war, you most likely do as well. Furthermore, Gartner et al found that marginal fatalities, which are the number of fatalities that occurred that year, are more important in ex-

plaining opinion than the cumulative number of fatalities [5]. They also studied the relationship between race and opinion towards the Vietnam war, finding that people were likely to view the conflict more negatively based on the number of people who died that were in the same locale, regardless of race [6]. Surprisingly, they found that Asian Americans had greater support for the Vietnam War; this could be because some of them were fleeing communism themselves [6].

2.2   Social Computing

The usage of large-scale datasets from social media has produced a flood of research and discourse in areas like sociology, political science, and psychology. Researchers have extensively studied the role of social media as a platform in coordinating activities before, during, and after conflicts. For example, time-series analysis of social media datasets has offered an ever-evolving account of public opinion and attention on a variety of issues, such as economic and social welfare, foreign affairs, and environmental issues [12]. Twitter, in the realm of social media, is one of the most popular corpora for studying armed conflicts and other political events. While many studies have looked at public perception of individual armed conflicts, we are not aware of any studies that have leveraged the mass of social media data to examine what causes perceptual differences among them.

2.3   Reddit Corpus

Reddit is an ideal corpus for asking such a question for a variety of reasons. First, Reddit is, at least partially, driven by news articles. Most Reddit submissions start with a user submitting a hyperlink. This has the added benefit of giving discussants a clear shared context.

After the submission, people can reply with their thoughts in a *comment*. Users can also comment on these comments. Users can also provide feedback with Reddit's curation system. Reddit users can *upvote* a comment or submission to express approval or *downvote* a comment or submission to express disapproval.

Both submissions and comments are sorted by *score*[1], the number of upvotes minus the number of downvotes. If a comment is sufficiently downvoted, it will be hidden. Hidden comments can still be read, replied to, upvoted, and down-voted, but a user has to manually display hidden comments. When a comment receives both upvotes and downvotes, it is considered *controversial*. In the case of armed conflicts, this is vanishingly rare: less than 1% of comments are considered controversial.

Reddit is organized into a highly expandable number of smaller forums, normally referred to as subreddits. Each of these subreddits can be about any given topic, general or specific. For instance, it is possible to have subreddits about science, biology, and genetics. These three subreddits can be completely independent, as they are not organized hierarchically.

---

[1] While comments are hierarchical, at each level of the hierarchy, they are sorted by score.

Besides the organizational ways that Reddit is different than Twitter, it is also more culturally homogeneous. The vast majority of the discussions are in English. Over half of the users are in the United States, and the majority are fairly young [2]. Among those not in the United States, the next most contributing countries are primarily East Asian, Australasian, Western European, and Northern European[3].

Researchers normally investigate phenomena in a small selection of subreddits. For instance, Hurricane Sandy has a dedicated subreddit. It was used to investigate Reddit's networked gatekeeping and how it impacts the framing of a crisis situation, as the voting system forms a non-traditional mechanism for negotiating what information is relevant [9]. Another study examined the sharing and seeking of mental health information to examine factors that drive social support [3].

2.4   Armed Conflict Database

The Armed Conflicts Database is developed by the International Institute for Strategic Study, containing various indexes of armed conflicts around the world [1]. They sort conflicts into several regions: Caribbean and the Americas, East Asia and Australasia, Europe, Middle East and North Africa, Russia and Eurasia, South Asia, and lastly, Sub-Saharan Africa. The Armed Conflict Database contains data on fatalities, IDP, and refugees from a conflict: both by year and in total. It additionally lists the year the conflict started and a variety of factors that relate to the conflict's origin, such as ethnic violence or terrorism. Lastly, they rate the current Intensity of the conflict, which can be Archived, Low, Medium, or High.

3   Research Questions

Our research questions aim to discover how people perceive and discuss armed conflicts. We suspect some of the same biases that affect traditional media will also affect social media, especially Reddit due to its news-driven process. However, it is unlikely there will be complete overlap.

Our data set consists of approximately 426GB of Reddit data, ranging from the year 2012 to the year 2014. We cross-referenced this with data from the Armed Conflicts Database, collecting a list of 48 conflicts that are considered by the Database to have been active in at least one of those years. This was to ensure none of the conflicts were seen by commenters as purely historical. See Figure 1 for a summary of where the conflicts occurred.

We gathered Reddit comments that are relevant to each of the 48 conflicts by searching for comments in every single subreddit. We compiled sets of keywords for every conflict then collected comments which matched them. For instance,

---

[2] Source–http://www.pewinternet.org/2013/07/03/6-of-online-adults-are-reddit-users/

[3] Source–http://www.redditblog.com/2013/12/top-posts-of-2013-stats-and-snoo-years.html

if a comment contained the phrase "Syrian Civil War" (not case-sensitive), we would mark that comment as relevant to the conflict in Syria. While there are certainly some false positives, we sought to counter that by having a sufficiently large data set. The biggest source of comments was the "worldnews" subreddit, with many of the others coming from similar subreddits. We took this as positive evidence that the majority of our comments were on topic.

Fig. 1. A map of the 48 conflicts. Black, yellow, orange and red circles indicate the current level of intensity as rated by the Armed Conflict Database: archived, low, medium and high, respectively.

We then analyze comments based on two main features: *sentiment* and *acceptability*. We rated over 25,000 comments using the StanfordNLP Sentiment Analyzer [17]. On the other hand, the acceptability refers to how the community perceives the comment, and thus the sentiment of the comment. In this case, an acceptable comment is one that the community would give upvotes, and an unacceptable comment is one the community would give downvotes.

We wanted to examine several characterizing features of conflicts that affect how they are discussed by traditional news agencies.

1. **Severity.** This includes the total number of people who were killed, made refugees, or internally displaced as a result of the conflict. While common sense would suggest that severity should play a large role in perception, prior work suggests it plays a very modest role in public perception of conflicts.
2. **Region.** There are six major regions that the Armed Conflict Database groups conflicts into. We hypothesize that regions that share more cultural similarity would be viewed differently than regions which share less. More clearly, regions where Reddit users are common would be different than regions where they are not.
3. **Marginal Severity.** This includes the number of people who were killed, made refugees, or internally displaced the same year as the comment was made. From previous research, we suspected that marginal severity would play a larger role than total severity.

4. **Age.** The number of years ago the conflict started. We hypothesized age could be significant due to waning interest. After decades of conflict, it is perhaps difficult for some users to still empathize with ongoing tragedy.

5. **Nature.** This includes a set of attributes: Separatism, Terrorism, Foreign Antagonism, Territorial Disputes, Criminal Violence, and Ethnic Violence. We hypothesized that attributes that are easier for Westerners to empathize with due to history, such as Separatism or Foreign Antagonism, might be treated differently. Further, attributes that seem very foreign or 'uncivilized', such as Ethnic Violence or Territorial Disputes, may also be seen differently. Importantly, a conflict can have more than one of these attributes.

6. **Expert Perception.** The Armed Conflict Database rates conflicts for their current level of intensity. While this is obviously correlated with deaths, refugees, and IDP ($p < 0.0001$), the average person may be more likely to get their ideas of severity from experts, rather than numbers. This could also influence the selection process and tone of traditional media articles.

4   Methods

4.1   Data Preparation

Interestingly, the sentiment analysis provided nearly twenty times more negative comments than positive comments. It is possible that this is due to the subject matter; in general, pity and sadness might be more common responses to tragedy than optimism and hope. It is also possible that some positive comments actually reflected misanthropic views about violence. Due to the difficulty of interpreting neutral sentiment comments, they were excluded.

Due to Reddit's manipulation of the probability of any given user seeing a post, and because controversy is so rare in our data set, we did not consider the actual number of upvotes or downvotes a comment received to be useful information. We thus coded anything upvoted at all as acceptable, and anything downvoted at all as unacceptable. We excluded comments that received no votes.

Our two filters resulted in 781 positive comments and 14289 negative comments. If the Positive Sentiment Model is substantially different than the Negative Sentiment Model, that suggests sentiment, perhaps through interaction with other variables, does influence acceptability. It is possible negative comments can be interpreted as the norm, instead of as a special case. If this is true, then that model, instead of reflecting when negative sentiment would be unacceptable, would reflect in general when posters are more likely to post something that is unacceptable.

4.2   Model

We use a Logistic Linear Mixed Effects Model (GLMM) to attempt to explain which sentiments were acceptable as determined by the conflict. Due to missing data for certain features (such as the marginal Refugees for any given year) and a high number of predictors, we decided to fit individual models for each predictor. This allows us to also see precisely how well each variable explains the variance.

	Estimate	Std. Error	z value	Pr(< |z|)
Intensity	0.3561	0.6668	0.5340	0.59300
Fatalities	0.1091	< 0.001	118.0000	< 0.00001*
IDP	0.3545	0.2568	1.3800	0.16700
Refugees	0.1361	0.3371	0.4040	0.68600
Separatism	-0.5505	< 0.001	-632.0000	< 0.00001*
Criminal Violence	0.5866	0.5738	1.0220	0.30700
Ethnic Violence	-2.4295	3.2196	-0.7550	0.45000
Terrorism	-0.8033	< 0.001	-861.0000	< 0.00001*
Territorial Dispute	-1.1330	1.2790	-0.8850	0.37600
Foreign Antagonism	1.0048	0.5995	1.6760	0.09380
Marginal Fatalities	0.3844	0.2611	1.4720	0.14100
Marginal Refugees	0.2577	0.3425	0.7520	0.45200
Marginal IDP	0.3223	0.2868	1.1240	0.261000
Age	-0.2801	0.3285	-0.8530	0.39400
Region-EastAsia/Australasia	-2.4860	2.9980	-0.8290	0.40689
Region-Europe	-0.4437	0.8270	-0.5360	0.59162
Region-MiddleEast/NorthAfrica	-0.2767	0.8544	-0.3240	0.74601
Region-Russia/Eurasia	-1.1910	1.6430	-0.7250	0.46847
Region-SouthAsia	-2173.0000	> 99999.0000	0.0000	0.99996
Region-SubsaharanAfrica	2.0970	2.3000	0.9120	0.36199

Table 1. The **positive** sentiment models. Region was fitted as a factor. Intensity was fitted as an ordinal variable where only the linear effect is reported. Significant variables are marked with an asterisk.

However, as we are making more than one comparison, we have to adjust our test of significance to avoid false positives. We use the Bonferoni correction [4], resulting in a significance threshold of $p = 0.0025$.

As every comment is coded as either unacceptable (1) or acceptable (0), positive β relates to downvotes and negative β relates to upvotes, which correspond with disapproval and approval effectively.

We grouped intercepts by Author and by Subreddit. As Reddit is pseudononymous, some authors may have a reputation for making consistently good or bad posts, and different communities may have different standards for acceptability.

5 Results

5.1 Positive Comments

For positive comments, there are reliable effects due to Fatalities, Separatism, and Terrorism. A conflict having more Fatalities increases the likelihood that a comment is considered unacceptable, while the presence of Terrorism and Separatism decrease that likelihood. See Table 1.

5.2 Negative Comments

For negative comments, there are reliable effects for Territorial Disputes, Age, and for the regions of Europe and East Asia / Australasia. Conflicts in both Regions make the comments less likely to be disapproved of, as does an older age.

The conflict being a Territorial Dispute, on the other hand makes the comment more likely to be disapproved of. See Table 2.

	Estimate	Std. Error	z value	Pr($<|z|$)
Intensity	0.1525	0.0697	2.1870	0.028700
Fatalities	0.0105	0.0247	0.4240	0.67200
IDP	0.0190	0.0332	0.5740	0.56600
Refugees	-0.0679	0.0397	-1.7100	0.08730
Separatism	-0.1036	0.0566	-1.8310	0.06710
Criminal Violence	0.0988	0.0559	1.7680	0.07700
Ethnic Violence	-0.2560	0.1605	-1.5950	0.11100
Terrorism	0.1249	0.0563	2.2160	0.02670
Territorial Dispute	0.4590	0.0798	5.7530	$< 0.00001*$
Foreign Antagonism	0.0958	0.0606	1.5820	0.11400
Marginal Fatalities	0.0097	0.0309	0.3140	0.75400
Marginal Refugees	-0.0615	0.0413	-1.4880	0.13700
Marginal IDP	-0.0501	0.0385	-1.3030	0.19200
Age	-0.1146	0.0277	-4.1400	$< 0.00010*$
Region-EastAsia/Australasia	-0.5248	0.1515	-3.2500	0.00116*
Region-Europe	-0.4437	0.8270	-0.5360	$< 0.00001*$
Region-MiddleEast/NorthAfrica	0.1383	0.0750	1.8420	0.06542
Region-Russia/Eurasia	-0.1895	0.0891	-2.1250	0.03355
Region-SouthAsia	-0.6498	0.2974	-2.1850	0.02891
Region-SubsaharanAfrica	0.1725	0.1578	1.0930	0.27448

Table 2. The **negative** sentiment models. Region was fitted as a factor. Intensity was fitted as an ordinal variable where only the linear effect is reported. Significant variables are marked with an asterisk

6   Discussion

There are several ways to interpret the rate of unacceptable posts for any given conflict. As the majority of Reddit users are fairly homogeneous in opinion, it is possible that people vary how likely they are to make posts about certain topics when they know that many people will disapprove. This variance could stem from *passion* about a topic: someone feels it is more important to speak their mind than to have a popular post. Alternatively, it could reflect *dissent*; while there might not be enough dissenters to affect the gatekeeping system, the willingness to express something others disapprove of at all suggests there are some who disagree with the majority opinion.

6.1   Negative Sentiment

Due to the vast majority of comments being of negative sentiment, we assume it is the default way to respond to a conflict. Thus, we will consider the Negative Sentiment model the same as the Default model.

The Regions of East Asia/Australasia and Europe both have reliable negative predictors. This implies the rate of disapproval when discussing these topics is very low. This is somewhat unsurprising given the general demographics of Reddit. Users from those regions, due to comparable socioeconomic status or

military alliances, may see themselves as similar, prompting a homophily effect. The lack of significant effects for the other regions could be because commenters conflate areas with which they are less likely to empathize, such as the Middle East/North Africa and SubSaharan Africa.

Age having a negative effect makes sense from the perspective of passion. While many people may have an opinion on older conflicts, these feelings may be less immediate due to the numbness and weariness of prolonged violence.

Territorial disputes on the other hand, are very logically controversial. To those whose country plays a role or are immediately affected by them, they may seem very important. However, to those farther away, they could seem like petty bickering.

Interestingly, the objective variables, such as Fatalities and IDP, played almost no role in the model. Absence of evidence is not evidence of absence; however, this integrates well with previous work [2].

6.2 Positive Sentiment

We can assume positive sentiment corresponds to hope, optimism, or perhaps even sarcastically phrased misanthropic sentiments. For instance, in the case of comments about conflicts with higher Fatalities being more likely to be disapproved of, it might be the latter.

On the other hand, positive sentiment surrounding conflicts containing notes of Terrorism or Separatism may correspond to hope. For instance, they could be expressions of hope for those attempting to separate from a regime where they do not feel represented, or wishes for those who are suffering terrorism to remain steadfast. These should both be uncontroversial ideas, so it is unsurprising they are more likely to be considered acceptable.

7 Conclusion

This paper sought to add to a growing body of work about media perception of armed conflicts by systematically investigating a large sample of Reddit data and cross-referencing the Armed Conflict Database. The vast majority of discussions are negative in tone, which is logical given the somber nature of violence. Among these, comments were less likely to be disapproved of if they were from the same demographic as the majority; they were more likely to be disapproved of if they concerned a Territorial Dispute or were older. Positive sentiment comments, on the other hand, were characterized differently. While Separatism and Terrorism likely evoked inspirational messages that were widely accepted, a high number of fatalities seemed to bring out sarcastic misanthropy.

8 Acknowledgements

References

1. Armed Conflict Database: Monitoring Conflicts Worldwide (2016), `https://acd.iiss.org/en`
2. Berinsky, A.J.: In Time of War: Understanding American Public Opinion from World War II to Iraq. University of Chicago Press, Chicago, IL (2009)
3. De Choudhury, M., De, S.: Mental Health Discourse on reddit: Self-disclosure, Social Support, and Anonymity. International AAAI Conference On Weblogs And Social Media (ICWSM 2014) (2014)
4. Dunn, O.J.: Estimation of the medians for dependent variables. The Annals of Mathematical Statistics pp. 192–197 (1959)
5. Gartner, S.S., Segura, G.M.: War, Casualties, and Public Opinion. Journal of Conflict Resolution 42(3), 278–300 (1998)
6. Gartner, S.S., Segura, G.M.: Race, Casualties, and Opinion in the Vietnam War. The Journal of Politics 62(1), 115–146 (2000)
7. Gelpi, C., Feaver, P.D., Reifler, J.: Paying the Human costs of War: American Public Opinion and Casualties in Military Conflicts. Princeton University Press, Princeton, NJ (2009)
8. Hawkins, V.: Media selectivity and the other side of the cnn effect: the consequences of not paying attention to conflict. Media, War and Conflict 4(1), 55–68 (2011)
9. Leavitt, A., Clark, J.A.: Upvoting hurricane sandy: Event-based news production processes on a social news site. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1495–1504. CHI '14, ACM, New York, NY, USA (2014)
10. Lim, M.: Clicks, Cabs, and Coffee Houses: Social Media and Oppositional Movements in Egypt, 2004-2011. Journal of Communication 62(2), 231–248 (2012)
11. Nelson, T.E., Clawson, R.A., Oxley, Z.M.: Media Framing of a Civil Liberties Conflict and Its Effect on Tolerance. American Political Science Review 91(3), 567–583 (1997)
12. Neuman, W.R., Guggenheim, L., Jang, S.M., Bae, S.Y.: The Dynamics of Public Attention: Agenda-Setting Theory Meets Big Data. Journal of Communication 64(2), 193–214 (2014)
13. Nossek, H.: Our News and their News: The Role of National Identity in the Coverage of Foreign News. Journalism 5(3), 343–368 (2004)
14. Sacco, V., Bossio, D.: Using social media in the news reportage of War & Conflict: Opportunities and Challenges. The Journal of Media Innovations 2(1), 59–76 (2015)
15. Seib, P.: Beyond the Front Lines: How the News Media Cover a World Shaped by War. Palgrave Macmilan (2004)
16. Shirky, C.: The political power of social media. Foreign affairs 90(1), 28–41 (2011)
17. Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the conference on empirical methods in natural language processing (EMNLP). vol. 1631, p. 1642