Modeling Emotion in Text: An analysis of Russian Twitter Trolls in U.S. Politics

Tara L. Valladares1 & Hudson Golino1

1 University of Virginia

“\clearpage”

Author note

Correspondence concerning this article should be addressed to Tara L. Valladares, 1023 Millmont St, Charlottesville, VA 22903. E-mail: [tls8vx@virginia.edu](mailto:tls8vx@virginia.edu)

Abstract

Emotion serves an important role in spreqding (dis)information online. Although there is extensive research on how Russian Twitter accounts that purposely spread misinformation (“trolls”) acted during the 2016 U.S. election, the use of emotion by these accounts is less understood. In this paper, we introduce a new method of anlayzing emotion in text called Emoxicon and detail its application to a set of tweets by Russian troll accounts.Emoxicon combines an emotion lexicon with Rasch modeling to describe how emotions are used in text. The results from the Emoxicon analysis suggest that left-wing and right-wing troll accounts used emotion in disparate ways. Different results were found when comparing emotions using the number of tweets or the number of accounts. Afraid and Annoyed were more common emotions within right-wing tweets, while angry and sad were more common among left-wing tweets. Under all circumstances, happy was the most uncommon expressed emotion.

Modeling Emotion in Text: An analysis of Russian Twitter Trolls in U.S. Politics

# Introduction

The study of disinformation intersects a myriad of different fields, including psychology, sociology, politics, computer science, and engineering. These fields often converge on the topic of state sponsored disinformation campaigns and their detection, influences, and impacts. Though organized efforts to spread disinformation are not new, the rise of the internet and social media as major players in shaping public opinion have allowed such campaigns to be conducted at a scale that was previously impossible. Thus, modern text analysis techniques are necessary to understand and quantify these largely text-based disinformation campaigns. As emotion plays a critical role in human communication, so is the quantification of emotion in text critical in understanding how Russian-linked Twitter accounts spread disinformation. In the current paper, we will first briefly summarize what is known about the Russian troll accounts, the role of Twitter (<https://twitter.com>) in politics, and the role of emotion in spreading (dis-)information. Following, we will introduce Emoxicon - a new method of extracting emotion from text using psychometric modeling, and its application to Russian-linked Twitter accounts around the time of the U.S. 2016 presidential election.

### Russian-linked Troll Accounts.

The Russian state has historically engaged in disinformation campaigns in an effort to influence foreign politics (Broniatowski et al., 2018; Mejias & Vokuev, 2017). These campaigns have utilized a variety of strategies to sow discord, including online media manipulation, cyber-espionage, and even falsely inciting protests in the real world (Linvill, Boatwright, Grant, & Warren, 2019; Richey, 2018). During the 2016 U.S. presidential election, Russian-based Twitter accounts posed as U.S. natives on Twitter and other social media sites to influence political discourse and public opinion (DiResta et al., 2019). Such accounts are often called “troll” or “sock puppet” accounts; they are social media accounts that claim to be a certain identity when the account is actually under the direct control of an obscured third party. These actions eventually culminated in the indictment of thirteen Russian individuals and three Russian entities by the United States Justice Department on charges relating to interference in the U.S. presidential election (Barrett, Horwitz, & Heiderman, 2018). The Internet Research Agency (IRA), a company based in St. Petersburg, Russia, was among the companies charged and has been a primary focus of investigations into state-sponsored Russian interference in U.S. politics (Permanent Select Committee on Intelligence, 2018). The IRA, in particular, has been described as “a sophisticated marketing agency” using a budget of at least $25 million U.S. dollars and one thousand employees to spread disinformation online in the United States (DiResta et al., 2019).

The IRA Twitter data was made accessible by FiveThirtyEight and the researchers Linvill and Warren (Linvill & Warren, 2018; Roeder, 2018). The topics tweeted about by the Russian troll accounts varies widely. Linvill and Warren (2018) consolidated the data and documented four major categories of English-language accounts that were active before the 2016 election: right-wing trolls, left-wing trolls, news feed accounts, and hashtag gamers. In the current analysis, we focused on only tweets from right-wing and left-wing troll accounts.

The right-wing and left-wing troll accounts were actively involved in the creation and dissemination of politicized content and propaganda (Zannettou et al., 2019). Political tweets from right-wing accounts generally followed popular conservative talking points of the time, supporting Donald Trump as the Republican presidential nominee and attacking Hillary Clinton, the Democratic party presidential candidate. Other popular right-wing topics included anti-Islam sentiment and terrorism, liberal media bias, gun rights, and supporting the police (Golino, Christensen, Moulder, Kim, & Boker, 2020). Political tweets from left-wing accounts largely attacked Trump, but they less frequently either attacked or supported Clinton (Linvill & Warren, 2018). A small number of accounts supported Bernie Sanders, an independent candidate. The left-wing accounts largely tweeted from a pro “Black Lives Matter” perspective, focusing on topics such as police brutality, racism in America, current events involving the deaths of black Americans (e.g., Trayvon Martin, Ferguson, Missouri, and Michael Brown), and popular black musicians and media (Golino et al., 2020). Approximately half of the left-wing troll tweets were from accounts that specifically posed as Black Americans (Freelon et al., 2020). Thus, the right-wing and left-wing trolls tweeted from opposite ends of the political spectrum and with few overlapping topics.

Notably, the majority of tweets from both the left- and right-wing accounts were deemed innocuous camouflage tweets by Linvill et al. (2019). These camouflage tweets were not overtly political but instead consisted of mundane interactions with followers or references to popular culture. There were apparent differences in how the account categories tweeted, though. For example, the right-wing trolls had less camouflage tweets than the left-wing trolls; over half of tweets from left-wing troll accounts were camouflage tweets, compared to less than 15% of tweets from right-wing troll accounts (Linvill et al., 2019). This suggests that left-wing and right-wing troll accounts were not only tweeting about different subjects but also employing different strategies in their attempts at political influence.

It is widely understood that social media disinformation campaigns pose a significant threat to representative democracies worldwide (DiResta et al., 2019; Sunstein, 2018), though the final impact of the IRA’s actions on the U.S. 2016 election has been debated (for dissenting opinions see Boyd et al. (2018) and Zannettou et al. (2019)). Nonetheless, extensive research has focused on how, why, where, and with whom the IRA-linked accounts spread disinformation, which will be explored further in the next section. Understanding the roles that both Twitter and troll accounts play is critical in identifying and understanding disinformation campaigns and limiting the impact of malicious actors in shaping public opinion.

### The Role of Social Media and Emotion in Politics.

Social media has cemented itself as a major player in the American political process in less than two decades. In 2019, 69% of U.S. adults used Facebook, with about 75% of users checking it at least once a day (Center, 2019). Political participation, protest activity, and voting have each been associated with social media behavior (Halpern, Valenzuela, & Katz, 2017; Min & Wohn, 2018; Skoric, Zhu, Goh, & Pang, 2016; Valenzuela, 2013; Valenzuela, Halpern, Katz, & Miranda, 2019). In conjunction, by 2018 every single U.S. Congressional Representative and Senator had a Twitter account and only a handful lacked Facebook and YouTube accounts (Straus, 2018). Because of social media’s role in politics and opinion formation, there is an extensive body of research describing how the websites are used and how users interact with each other.

Though Twitter can be a highly valuable source of information for its users, its unmoderated content is not read and spread in unbiased manner. The results from Valenzuela et al. (2019) suggest that highly active social media users also tend to share more misinformation regardless of how (mis)informed these users actually are. As a whole, social media users tend to follow and interact with accounts that hold similar political beliefs, forming “echo chambers” online (Del Vicario et al., 2016; Garrett, 2009; Törnberg, 2018). Highly active social media users tend to be more polarized and to follow fewer sources of information (Bessi et al., 2015; Schmidt et al., 2017). Politically active Twitter users rarely retweet content from accounts or news sources from different political parties. In Boutet, Kim, and Yoneki (2012) the majority of all interactions on Twitter occurred between users with homologous political views. While users largely retweet homologous opinions, user-to-user mentions (i.e., tagging another user in a tweet) frequently cross political lines (Boutet et al., 2012; Conover et al., 2011). This cross-over exposure may not reliably prevent polarization though, especially if users are interacting with opposing, polarizing content from either real or troll accounts (Karlsen, Steen-Johnsen, Wollebæk, & Enjolras, 2017; Yardi & Boyd, 2010). In fact, a handful of studies have found a “backfire” response where exposure to opposing views increases polarization (Bail et al., 2018; Nyhan & Reifler, 2010).

Emotion plays a significant role in driving public opinion and online political discourse (González-Bailón, Banchs, & Kaltenbrunner, 2012; MacKuen, Wolak, Keele, & Marcus, 2010; Peng, Sun, & Wu, 2017). González-Bailón et al. (2012) found that shifts in emotional arousal on a political discourse forum were linked to corresponding shifts in public opinion. On Twitter, influential political users are twice as likely to tweet negatively about politicans and political parties, and more emotional political tweets tend to spread further and faster (Brady, Wills, Jost, Tucker, & Van Bavel, 2017; Dang-Xuan, Stieglitz, Wladarsch, & Neuberger, 2013). In Brady et al. (2017), adding a single moral-emotional word (e.g., “hate” or “blame”) to a tweet increased the expected sharing rate by 17%-24%, depending on the topic. The use of moral-emotional language is also associated with message spread from the political leaders to their followers (Brady, Wills, Burkart, Jost, & Van Bavel, 2019). Similarly, politicians with extreme ideological positions also tend to have more Twitter followers (Hong & Kim, 2016).

Specific emotion categories have been identified as contributing to information spread. For example, the use of anger and disgust may be especially impactful among conservative political figures on Twitter (Brady et al., 2019). This may be driven by either reader characteristics or the topic. For example in Brady et al. (2017), anger and sadness words increased sharing for climate change tweets, while positive emotions increased sharing for same-sex marriage. The impact of emotion on the spread of ideas is not limited to Twitter and may be a more general phenomenon. One study using data from The New York Times found that articles which elicited stronger emotional arousal (i.e., awe, anger, and anxiety) were more likely to be shared; the odds that an article made the top shared list increased by 34% for every one standard deviation increase in anger, 30% for awe, and 21% for anxiety (Berger & Milkman, 2012). Guerini and Staiano (2015) found that similar emotions (inspired, angry, and annoyed) most strongly influenced the likelihood that an article on the news website Rappler.com was shared or commented on. Further, moral-emotional language may have more of an impact in spreading ideas within-groups than between-groups, contributing to polarization and the “echo chamber” effect (Brady et al., 2017). Anger, specifically, may also increase partisan beliefs and susceptibility to misinformation (MacKuen et al., 2010; Suhay & Erisen, 2018; Wollebæk, Karlsen, Steen-Johnsen, & Enjolras, 2019). In total, the available evidence supports the role of emotion in spreading political ideas and information, though the impact of specific emotions may be context dependent.

Current research on the activities of Russian IRA accounts suggests that they leveraged both the role of emotion and current echo chamber landscape to spread political disinformation. Zannettou et al. (2019) showed that Russian troll tweets are more negative and less positive than tweets from random subset of one thousand Twitter users with similar posting activity. In fact, some of the most popular IRA tweets were also the most racist and offensive (Freelon & Lokot, 2020).

Twitter estimates that by early 2018, approximately 1.4 million users (2% of active users) had interacted with IRA accounts (, Twitter Public Policy, 2018). Bail et al. (2020) surveyed 1,200 active Twitter users who identified as either a Republican or a Democrat and found that 9% had interacted with IRA accounts and 11.3% directly engaged with them. These higher percentages suggest that partisans were more likely to interact with IRA accounts. In the same study, stronger political interest and more homologous social networks were also associated with higher likelihood of IRA interactions. Other research also suggests that troll accounts largely interacted with users already within their political cluster (Freelon & Lokot, 2020; Stewart, Arif, & Starbird, 2018). In addition, conservatives may be more likely to retweet IRA content than liberals (Badawy, Ferrara, & Lerman, 2018; Bail et al., 2020), though it is unclear if this is due to either the IRA’s actions or user characteristics.

## Aims

Though emotion plays a critical role in shaping political thought online, we do not have a thorough understanding of how the IRA accounts’ used emotion in their disinformation campaign. The purpose of the current study is to introduce a novel method of emotion detection termed Emoxicon, and detail its application to the IRA Twitter accounts. Emoxicon combines lexicon scoring with Rasch modeling (Rasch, 1960) to understand the distribution and probability of emotion categories. Because so much of human communication now occurs on mediums characterized by brief segments of texts, there is a great need for flexible, low-computation methods of emotion detection that reliably work on relatively short documents. Using Emoxicon, we aim to map out the distribution of emotions and this differs between both right-wing and left-wing trolls and the accounts within these groups.

# Methods

## Emoxicon

Emoxicon is a novel method of analyzing the emotional content of text. By combining lexicon-based scoring with Rasch modeling, Emoticon provides a low-computation analysis of the probability that different emotions categories appear within text. Lexicon-based analyses are already well-established methods used to identify emotion within text (Brady et al., 2017; González-Bailón et al., 2012; Peng et al., 2017). Our contribution lies in the use of Rasch models to quantify the distribution and likelihood of the co-occurrence of emotions and the “emotional voice” of the author. Emoxicon is a two part method that starts by scoring data using an emotion lexicon and then models the distribution of the emotion categories using the Rasch model.

### Lexicon Scoring.

We used the DepecheMood++ lexicon created by Araque, Gatti, Staiano, and Guerini (2018). Lexicons form the basis of most methods of emotion detection. Lexicons are dictionaries that contain features (such as words or punctuation), the category they belong to, and their designated score. DepecheMood++ is a high-coverage emotion lexicon that contains roughly 37,000 terms and their associated probability weights for 8 emotion categories. The lexicon was created by scraping data from the news website Rappler which features a native “Mood Meter” interface on each article. Readers of the site are encouraged to select one or more of eight reactions (Afraid, Amused, Angry, Annoyed, Don’t Care, Happy, Inspired, or Sad) to the article. DepecheMood++ has several features that make it particularly useful. DepecheMood++ was naively crowdsourced from at least one million responses to over 53,000 documents; it did not rely on hand-annotation by knowledgeable participants (Araque et al., 2018). As a result, DepecheMood++ is much larger and covers more terms than other commonly used lexicons. Lexicon-based analyses are generally limited by the lexicon’s size and complexity. The larger the lexicon, the more robust the scoring can be. Smaller lexicons may be poor matches to a target text if there is little overlap in words, and significant bias can occur when a target text is poorly matched to the lexicon.Previous work has shown that the DepecheMood++ lexicon performs as well or better than other emotion lexicons in text analysis (Araque et al., 2018; Staiano & Guerini, 2014).

A simple bag-of-words method is applied to score the emotional content of a given document. When using bag-of-words, all grammar and word placement is discarded to create an orderless set of word counts. Though relatively crude, bag-of-words can produce compelling results on its own or in combination with machine learning techniques. Bag-of-words combined with standard machine learning algorithms can produce accuracy rates in tweet sentiment classification around 70-80%, which improves with the addition of sentiment lexicons (Da Silva, Hruschka, & Hruschka Jr, 2014). In the original DepecheMood++ lexicon, each word is assigned a weight for each of the eight dimensions. Here, the lexicon is condensed; each word in the lexicon becomes associated with the emotion category with the highest probability weight. Then, each tweet is assigned a sum score of how many words are most highly associated with each emotion category. This aids in the scoring of the text and reduces noise in the Rasch model.

### Rasch Model Analysis.

The Rasch model (Rasch, 1960) is used to evaluate the relative probability of an emotion weight appearing in a given tweet within the Emoxicon methodology. By using the Rasch model, we are able to analyze the fit of the emotion weights to the document. The Rasch model originated in the field of psychometrics to relate scales and questionnaires to the underlying latent trait. The Rasch model is a logistic model that places persons and items on the same interval-level scale to calculate the probability of a given person endorsing a given item. Fundamentally, the Rasch model can be used to generate latent trait scores for entities that produces a set of responses determined by an underlying latent trait. The basic form of the Rasch model for dichotomous items, as adapted from Rasch (1960), is:

Where represents the trait level of person and represents the difficulty or location of item . In Emoxicon, tweets take the place of persons and emotional scores are represented as items. Each tweet is scored based on how many words are present from each emotion category. A mean split was performed on the emotion word counts within each emotion category to dummy code each tweet as 1 (high) or 0 (low) for each category. For example, a tweet with a higher than average number of *Happy* words would receive a 1 for the *Happy* item. The dichotomized scores are run through the Rasch model to produce trait scores for the tweets and item locations for the emotion categories.

Under the Rasch model, items receive estimates of their relative frequency of occurrence on the latent dimension, also known as item difficulty or location. An item with a high value on the latent dimension is traditionally described as a difficult item, indicating that it is rarer for this item to receive a “1” (“be passed”). An item with a low value on the latent dimension is traditionally described as an easy item, indicating that it more common for this item to be passed. Like items, persons also recieve a latent score on the same dimension. A high person score indicates that the person is likely to pass more items.

An underlying assumption of the Rasch model is unidimensionality, which is the requirement that there is only one underlying latent trait driving all responses. In standard applications of the Rasch model, this is a trait such as math ability. Here, we propose that the dimension that the texts are scored on is *emotional voice* of the author. If an author writes in a particular style with particular opinions, we propose that the use of emotion categories in their writing will not be random. Rather, particular authors will tend to express certain emotions more regularly than others. There will be a pattern to the expression of emotions that is consistent across writings that goes further than which emotions are common and which are uncommon.

Under the Rasch model, item estimates have meaningful order conditional on the latent dimension. This meaning can be expressed with two priciples. First, an item with a lower estimate is more likely to be passed than an item with a higher estimate. Second, if a person passes a more difficult item, then it is also likely that they also passed the easier items. The item estimates are thus also estimates of conditional probability. Deviation from these assumptions will produce poor model, item, or person fit. This conditional fit underlies the interpretation of the latent dimension as the emotional voice of the author.

In summary for the Emoxicon application of the Rasch model, persons represent tweets and items represent the emotion categories from the DepecheMood++ lexicon. Item categories that have high trait estimations are less likely to appear in a given tweet. Tweets that recieve high trait estimates are more likely to contain many emotion words in many emotion categories. Poor model fit indicates that the emotion catgories are either unrelated to each other and there is no conditional order of the emotion catgories.

## Data Pre-processing

The purpose of this analysis is to examine differences in tweets written by the IRA among the left-wing and the right-wing troll accounts. We will refer to such tweets as “authored” tweets. By removing non-authored tweets, we ensure that all the tweets support the goals of the accounts and do not belong to accounts with other “voices.” To this end, we included only English-language troll accounts labeled as left-wing ("Left") or right-wing ("Right") by Linvill and Warren (2018). URLs, retweets, manual reposts (i.e., tweets that begin with “rt”, meaning “retweet”), and duplicate tweets were removed.

Because the DepecheMood++ lexicon was sourced from news articles, it contains emotion associations for a number of politically relevant words, such as "Hillary," "Clinton," and "Trump". We removed a small set of topically relevant terms that we anticipated had different affective associations between the left-wing and right-wing accounts (see the Appendix for a full list of removed words). That is, left-wing and right-wing trolls may use the word "Trump" just as frequently, but that word does not have the same relationship to the tweets true emotional content for each group. As these words were used often, they also would artificially inflate their respective category frequency. Similarly, we did not split up multiword hashtags (e.g., “#blacklivesmatter”) to prevent misleading category inflation; the emotional valence of the whole hashtag was likely not a sum of its parts.

## Analysis Plan

We ran two separate but complimentary applications of the *Emoxicon* method to the Russian Trolls dataset: an omnibus Rasch model for the 849 number of Twitter accounts, and 592 Rasch models for each individual account. The omnibus Rasch model included all cleaned tweets from both left-wing and right-wing trolls. In this strategy, we are able to identify differences between left-wing and right-wing trolls in the majority of the authored tweets. However, these broad group-level differences obscure the actions of less productive accounts and are largely dominated by a handful of mega-accounts with thousands of self-authored tweets. To compliment the omnibus model, we also ran individual Rasch models for Twitter accounts. This strategy allows us to examine differences between accounts regardless of their productivity. For the individual models, only accounts with at least 30 tweets were included to sufficiently fit each Rasch model (Linacre, 1994).

Each method used the same data scored with the DepecheMood++ lexicon (Araque et al., 2018). Using the eight emotion catgories, raw scores ranged from 0 to 8. All models and fit statistics were produced using the R package eRm (Mair, Hatzinger, & Maier, 2020).

The main variable of interest was the estimated item parameters produced by the Rasch model, specifically the item difficulties. In the omnibus model, we used Andersen’s likelihood ratio test in the eRm package to identify differences in item difficulties between left-wing and right-wing trolls (Andersen, 1973; Mair et al., 2020).

Because the mean of either the person or item parameters of the Rasch model must be set to zero before estimation to establish the scale and the individual models use separate data, item parameters are not directly comparable across individual Rasch models. Therefore to compare the relative difficulties between models, item parameters from the individual models were transformed into ranks from 1 (easiest) to 8 (hardest) within each model. Ties in rank were broken randomly. The distribution of the ranks could then be examined for consistency within and differences between Right and left-wing trolls. If an individual account did not use any words from a specific emotion category, that emotion category was given the rank of 8 (hardest to endorse).

To evaluate the ranks of the emotion categories in the individual models, we examined their distribution using Kolmogorov-Smirnov distribution and Wilcoxon rank-sum tests (Corder & Foreman, 2014). In addition, we evaluated the consistency of the item difficulties within the left-wing and right-wing troll groups using the multiinformation (total correlation) function from the R package infotheo (Meyer, 2008), and the diversity of the emotion labels using the Gini coefficient (Ong, Benson, Zautra, & Ram, 2018).

Our final sample consisted of 420,447 tweets from 622 right-wing accounts ( = 347,489) and 227 left-wing accounts ( = 72,958). Among Twitter handles with at least 30 tweets, there were 429 right-wing accounts ( = 345,489) and 163 left-wing accounts ( = 72,162). Tweets dated from 1/1/2013 to 9/9/2017. See Table 1 for more information on the distribution of the final sample of tweets.

Table 1:

*Distribution of Tweets in Accounts*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| x | Accounts, n | Total Tweets | Minimum | 2nd Quartile | Median | 3rd Quartile | Maximum |
| Left-wing, all accounts | 227 | 72958 | 1.00 | 26.50 | 106.00 | 302.50 | 7315 |
| Left-wing, 30+ tweets | 163 | 72162 | 30.00 | 92.50 | 199.00 | 376.00 | 7315 |
| Right-wing, all accounts | 622 | 347489 | 1.00 | 18.00 | 76.00 | 450.00 | 35082 |
| Right-wing, 30+ tweets | 429 | 345489 | 30.00 | 72.00 | 191.00 | 654.00 | 35082 |

*Note.* Tweets removed during pre-processing are not included.

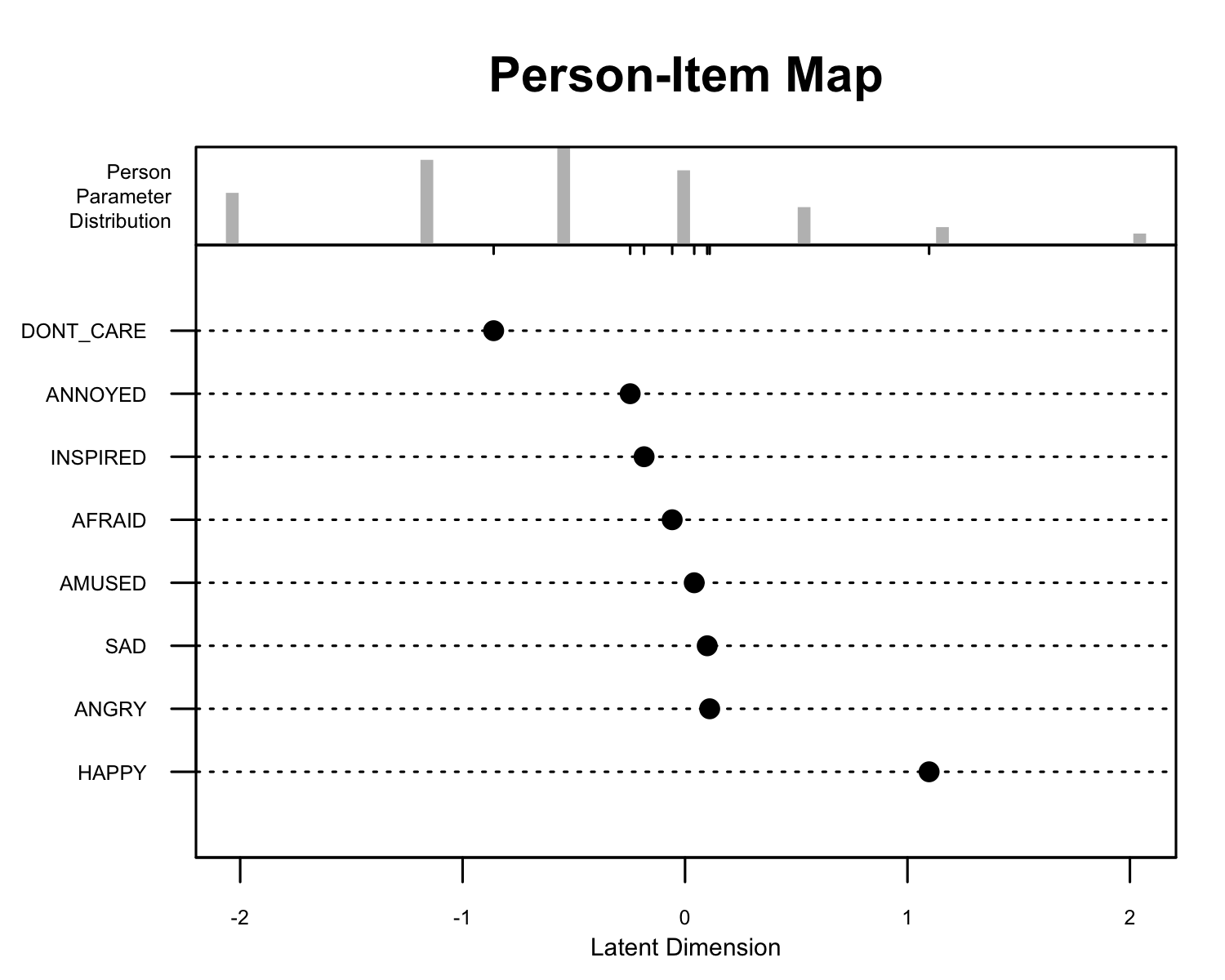
# Results

## Omnibus Rasch Model

### Model Fit.

Overall model fit for the full Rasch model was appropriate. Only 1.17% of tweets showed significant misfit using a Chi-square based Z-value cutoff of 1.96. The items *Amused* and *Inspired* showed misfit based on chi-squared item fit statistics ( (399,040) = 401,893.5, p = .001 and (399,040) = 418,125.2, p < .001 , respectively). Due to the large sample size, and the results in the following sections, the misfit in these items is not of great concern. To check for multidimensionality, we performed a principal components analysis on the Rasch model residuals. The largest eigenvalue was 1.28, indicating that there was not a meaningful impact of multidimensionality on the model fit (Linacre, 1998).

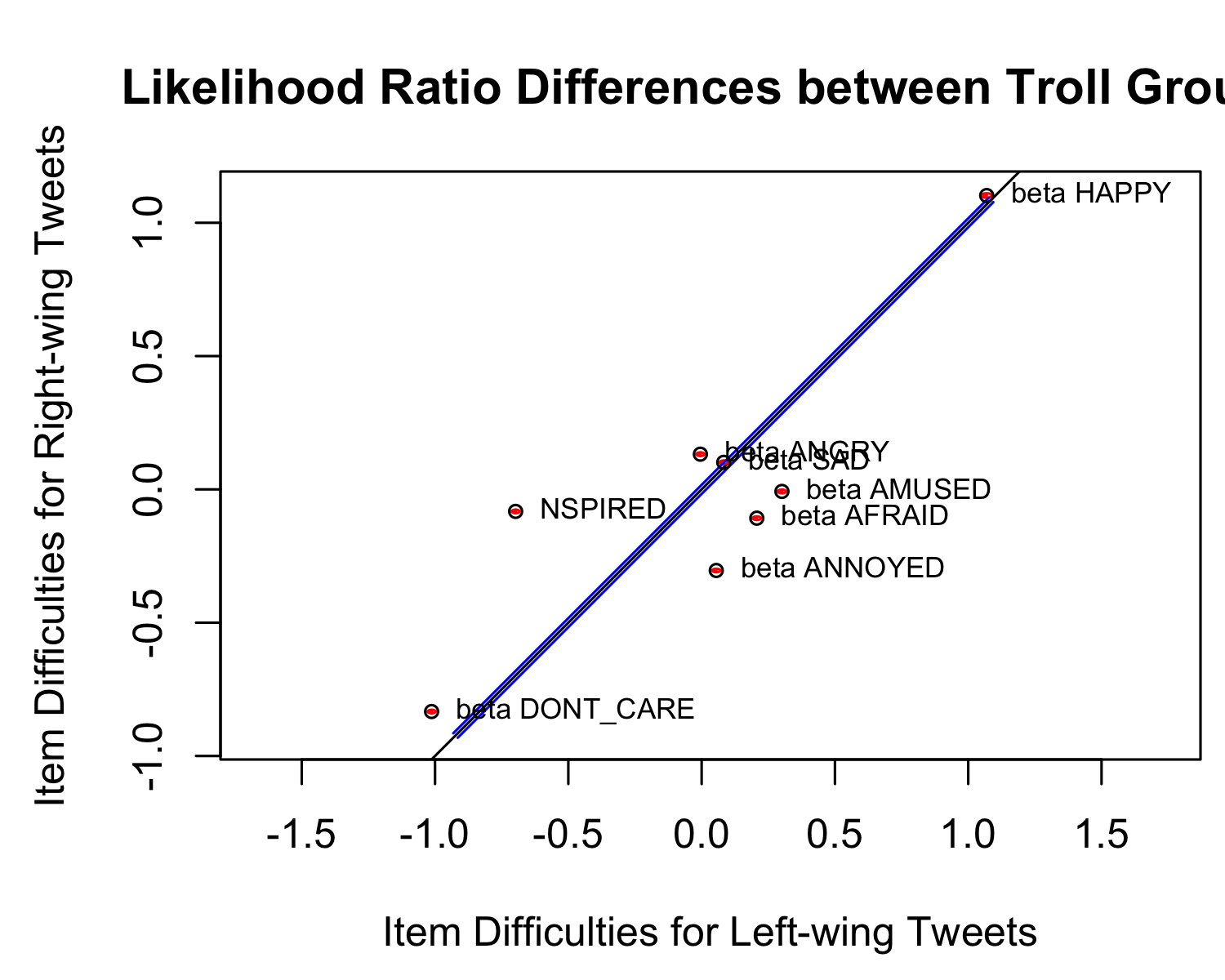
Right-wing tweets had slightly higher Rasch scores (M = -0.73, SD = 0.91) than left-wing tweets M = -0.84, SD = 1.08), *t*(96,172.72) = -26.91, *p* < .001, though the difference was empirically small. Under the model, estimated item difficulties were generally well-matched to estimated tweet scores (Figure 1). In the Rasch model, *Don’t Care* was the easiest item with the lowest trait estimates, while *Happy* was the item with the highest trait estimate. The locations of the other items were estimated to be relatively close to one another, indicating that the categories were used at relatively similar rates when both left-wing and right-wing trolls were combined.



*Figure* *1:*. There were 5.09% tweets with full zero/complete response patterns. They are not included in the pictured person distribution.

### Left-wing and Right-wing Differences.

The Andersen’s likelihood ratio test showed significant differences between the left-wing and right-wing troll tweets ((7) = 9,768.27, p < .001; Figure 2) (Mair et al., 2020). The items *Amused*, *Afraid*, and *Annoyed* were more difficult among left-wing Tweets. *Inspired* was more difficult for right-wing tweets. *Don’t Care* and *Angry* were also more difficult for right-wing tweets, but to a much smaller degree.The items *Happy* and *Sad* were similarly used between both left-wing and right-wing trolls.



*Figure* *2:*. Items closer to the black identity line were estimated more similarly between left-wing and right-wing tweets. The 95% confidence intervals are shown in blue for the identity line and in red for the item difficulties.

## Individual Rasch Models

### Model Fit.

We fit 592 Rasch models for each of the individual Twitter handles with at least 30 authored tweets. Among these models, there were 10 models with full zero responses for the category *Happy* and 1 model with full zero responses for the category *Afraid*. The average percentage of misfitting persons was 0.73% (SD = 0.82) for left-wing accounts and 0.99% (SD = 0.91) for right-wing accounts. Item misfit was also generally low (Table 2). The majority of items misfit in less than 6% of models, except for *Amused*, *Happy*, and *Inspired*. Overall, the measures of model fit were appropriate.

Table 2:

*Percent of Models with Misfitting Items*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Group | Amused | Angry | Annoyed | Don’t Care | Inspired | Sad | Afraid | Happy |
| Left-wing | 7% | 2% | 4% | 2% | 11% | 6% | 2% | 9% |
| Right-wing | 13% | 0% | 6% | 1% | 4% | 3% | 2% | 7% |

### Category Ordering.

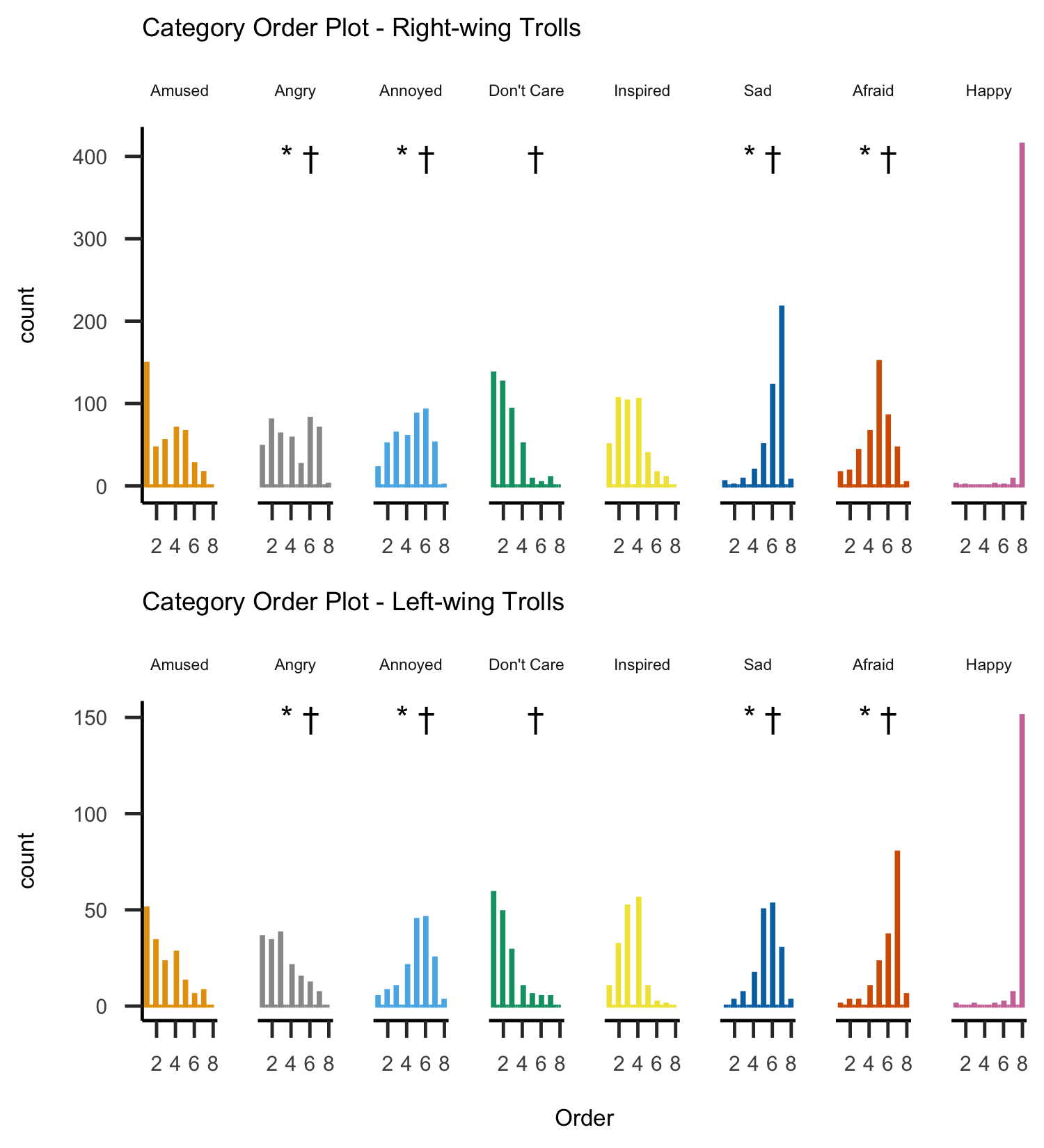
To compare the estimates between individual models, item parameters were transformed into ranks from 1 (easiest) to 8 (hardest) within each model. Figure 3 displays the distributions of the order of item difficulties for both left-wing and right-wing troll accounts. Mean rank differences were evaluated by using the Wilcoxon rank sum test. Distributional differences were evaluated with Kolmogorov-Smirnov Tests (Corder & Foreman, 2014). Differences in spread were evaluated using Levene’s test with the median as the center (Levene, 1960). Across both groups, the hardest category was *Happy* for 95.61% of accounts. There were significant median and distributional differences between left-wing and right-wing trolls for several categories (see Table 3 for descriptive statistics). The following emotions showed both median and distribution differences: *Angry* (*U*(*nR*=429, *nL*=163) = 44,924, *p* < .001; D = 0.24, *p* = < .001), *Annoyed* (*U*(*nR*=429, *nL*=163) =27,597, *p* < .001; D = 0.19, *p* < .001), *Sad* (*U*(*nR*=429, *nL*=163) = 48,227.5, *p* < .001; D = 0.32, *p* < .001), and *Afraid* (*U*(*nR*=429, *nL*=163) = 16,379.5, *p* < .001; D = 0.44, *p* < .001). The category *Don’t Care* showed distributional differences (D = 0.05, *p* = 0.049), but not median differences. The category *Happy* showed a difference in medians (*U*(*nR*=429, *nL*=163) = 36,391.5, *p* = .031) though the empirical mean difference was small. There were no median or distributional differences in the item difficulty orders for the categories *Amused* or *Inspired*. Among right-wing troll accounts, the categories *Annoyed* and *Afriad* were generally easier. Among left-wing troll accounts, the categories *Angry* and *Sad* were generally easier.

Differences in variance were found for the categories *Angry* (*F*(590) = 23.2, *p* < .001) , *Annoyed* (*F*(590) = 9.1, *p* = .003) , and *Inspired* (*F*(590) = 11.52, *p* < .001) , while *Amused* (*F*(590) = 3.65, *p* = .054) approached statistical significance. The variance was always larger for the right-wing trolls among these four categories, indicating higher spread and less interal cohesion around the median.

Table 3:

*Means and Standard Deviations of Emotion Category Orders*

|  |  |  |
| --- | --- | --- |
| Category | Median, Mean, (SD) - Right-Wing | Median, Mean, (SD) - Left-Wing |
| Amused | 3.00,  3.01,  (1.86) | 2.00,  2.80,  (1.75) |
| Angry | 4.00,  4.10,  (2.07) | 3.00,  3.06,  (1.72) |
| Annoyed | 5.00,  4.46,  (1.75) | 5.00,  5.12,  (1.54) |
| Don’t Care | 2.00,  2.34,  (1.34) | 2.00,  2.33,  (1.51) |
| Inspired | 3.00,  3.15,  (1.42) | 3.00,  3.21,  (1.10) |
| Sad | 7.00,  6.21,  (1.15) | 6.00,  5.50,  (1.19) |
| Afraid | 5.00,  4.81,  (1.49) | 7.00,  6.14,  (1.28) |
| Happy | 8.00,  7.92,  (0.61) | 8.00,  7.84,  (0.76) |



*Figure* *3:*. The distributions of emotion categories within accounts when ordered from 1 (easiest) to 8 (hardest). \*indicates significant differences between left-wing and right-wing accounts based on Wilcoxon rank-sum tests. † indicates significant differences based on Kolmogorov-Smirnov distribution tests.

### Gini Index (Emodiversity).

Following the emodiversity method used in Ong et al. (2018), we used the Gini coefficient to estimate the diversity in category usage between left-wing and right-wing trolls. High Gini coefficient values indicate more less diversity (more inequality or unevenness) among emotion categories. We used the raw emotion scores rather than the Rasch model dichotomized scores. Left-wing trolls showed more inequality among emotion categories than right-wing trolls when the Gini coefficient was calculated either across all tweets (left-wing = 0.64, right-wing = 0.6; (*t*(103,073.6) = 65.18, *p* < .001) or within authors (left-wing = 0.46, right-wing = 0.42; (*t*(413.76) = 7.02, *p* < .001).

### Multiinformation.

We calculated multiinformation (also known as total correlation (Meyer, 2008)) with bootstrapped confidence intervals for the category orders for the left-wing and right-wing accounts to understand the consistency in the orders within groups. Higher values of multiinformation signify higher entropy, that is, lower relationships between variables. For left-wing accounts, multiinformation was 7.05 (95% CI [6.34, 7.24], bias = 0.26); for right-wing accounts, multiinformation was 6.71 (95% CI [6.2, 6.64], bias = 0.29). We also bootstrapped the average multiinformation among a random ordering of the numbers 1 through 8, resulting in a maximum value of multiinformation value of 10.22 (95% CI [9.65,9.75]) for a data structure like the category orders in this study. Both left-wing and right-wing trolls showed more cohesion in their category orders than would be expected by chance, but they did not differ from one another.

### Summary.

# Discussion

Topics to hit:

* How did Emoxicon perform? Did this study support our claims that it is a good method?
* Which emotions were most common?
* Do these results link up to the previous research on virality and emotions?
* The differences in emotion expression may be linked to different methods of eliciting emotion. Different purposes, topics eliciting different emotions?

*Happy* was the hardest emotion category for both left-wing or right-wing trolls when examining the total number of tweets and the individual models.

Of note, the emotion categories with that showed misfit most often (*Amused*, *Happy*, and *Inspired*) were also the categories that generally did not show differences between left-wing and right-wing trolls.

The higher inequality among left-wing troll tweets measured via the Gini ceofficient corresponded to the category order results. The variance of the distrubtions of emotion category orders tended to be larger for the right-wing trolls as compared to the left-wing trolls. These larger variances imply more variety (evenness) in the difficulties of the emotion categories among the right-wing accounts. The results of both the Gini coefficient and the category orders suggest that right-wing troll accounts varied more in their emotional expression than the left-wing accounts.

## Limitations

It is critical to interpret the results of this study in context of its methods. Emotion is not a universal, context-free phenomenon; statements that are neutral in one context may elicit powerful reactions in another. The emotion categories delineated in the DepecheMood++ lexicon were sourced from a specific time and group membership. When a tweet is rated highly on *Happy*, it does not imply that all readers of the tweet would have the same reaction. This may be even more true when dealing with highly politically and emotionally charged content that was written to elicit strong emotion reactions.

We propose that differences in emotion expression in this study should rightly be interpreted as differences in how one, perhaps neutral, subset of readers may react. We do not believe that these differences are not true or present in reality - rather that it is an overstep to assume that every person who reads a left-wing troll tweet is more likely to react with anger than if they read a right-wing troll tweet.

# Words removed from Lexicon

hilary hillary russia russian trump donald bernie sanders clinton rt obama barack america president black white racist

, Twitter Public Policy. (2018). Update on twitter’s review of the 2016 US election. Retrieved from <https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html>

Andersen, E. B. (1973). A goodness of fit test for the rasch model. *Psychometrika*, *38*(1), 123–140.

Araque, O., Gatti, L., Staiano, J., & Guerini, M. (2018). DepecheMood++: A bilingual emotion lexicon built through simple yet powerful techniques. Retrieved from [https://arxiv.org/abs/1810.03660](%20https://arxiv.org/abs/1810.03660)

Badawy, A., Ferrara, E., & Lerman, K. (2018). Analyzing the digital traces of political manipulation: The 2016 russian interference Twitter campaign. In *2018 ieee/acm international conference on advances in social networks analysis and mining (asonam)* (pp. 258–265). IEEE.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., … Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221.

Bail, C. A., Guay, B., Maloney, E., Combs, A., Hillygus, D. S., Merhout, F., … Volfovsky, A. (2020). Assessing the russian internet research agency’s impact on the political attitudes and behaviors of american Twitter users in late 2017. *Proceedings of the National Academy of Sciences*, *117*(1), 243–250.

Barrett, D., Horwitz, S., & Heiderman, R. (2018). Russian troll farm, 13 suspects indicted in 2016 election interference. *Washington Post*.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, *49*(2), 192–205.

Bessi, A., Zollo, F., Del Vicario, M., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Trend of narratives in the age of misinformation. *PloS One*, *10*(8).

Boutet, A., Kim, H., & Yoneki, E. (2012). What’s in your tweets? I know who you supported in the [uk] 2010 general election. In *Sixth international aaai conference on weblogs and social media* (pp. 411–414).

Boyd, R. L., Spangher, A., Fourney, A., Nushi, B., Ranade, G., Pennebaker, J., & Horvitz, E. (2018). Characterizing the Internet Research Agency’s social media operations during the 2016 US presidential election using linguistic analyses. *PsyArXiv*. Retrieved from <https://psyarxiv.com/ajh2q/>

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

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, *114*(28), 7313–7318.

Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., … Dredze, M. (2018). Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *American Journal of Public Health*, *108*(10), 1378–1384.

Center, P. R. (2019). Share of u.s. Adults using social media, including Facebook, is mostly unchanged since 2018. Retrieved from <https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/>

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on twitter. In *Fifth international [aaai] conference on weblogs and social media* (pp. 89–96).

Corder, G. W., & Foreman, D. I. (2014). *Nonparametric statistics: A step-by-step approach*. Hoboken, New Jersey: John Wiley & Sons.

Dang-Xuan, L., Stieglitz, S., Wladarsch, J., & Neuberger, C. (2013). An investigation of influentials and the role of sentiment in political communication on Twitter during election periods. *Information, Communication & Society*, *16*(5), 795–825.

Da Silva, N. F., Hruschka, E. R., & Hruschka Jr, E. R. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, *66*, 170–179.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., … Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559.

DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., … Johnson, B. (2019). *The tactics & tropes of the Internet Research Agency. White Paper for the United States Senate Select Committee on Intelligence*. Austin, Texas: New Knowledge.

Freelon, D., Bossetta, M., Wells, C., Lukito, J., Xia, Y., & Adams, K. (2020). Black trolls matter: Racial and ideological asymmetries in social media disinformation. *Social Science Computer Review*, 1–19.

Freelon, D., & Lokot, T. (2020). Russian Twitter disinformation campaigns reach across the american political spectrum. *Harvard Kennedy School Misinformation Review*.

Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among internet news users. *Journal of Computer-Mediated Communication*, *14*(2), 265–285.

Golino, H., Christensen, A. P., Moulder, R. G., Kim, S., & Boker, S. M. (2020). Modeling latent topics in social media using dynamic exploratory graph analysis: The case of the right-wing and left-wing trolls in the 2016 US elections. Preprint at https://psyarxiv.com/tfs7c.

González-Bailón, S., Banchs, R. E., & Kaltenbrunner, A. (2012). Emotions, public opinion, and US presidential approval rates: A 5-year analysis of online political discussions. *Human Communication Research*, *38*(2), 121–143.

Guerini, M., & Staiano, J. (2015). Deep feelings: A massive cross-lingual study on the relation between emotions and virality. In *Proceedings of the 24th international conference on world wide web* (pp. 299–305).

Halpern, D., Valenzuela, S., & Katz, J. E. (2017). We face, I tweet: How different social media influence political participation through collective and internal efficacy. *Journal of Computer-Mediated Communication*, *22*(6), 320–336.

Hong, S., & Kim, S. H. (2016). Political polarization on twitter: Implications for the use of social media in digital governments. *Government Information Quarterly*, *33*(4), 777–782.

Karlsen, R., Steen-Johnsen, K., Wollebæk, D., & Enjolras, B. (2017). Echo chamber and trench warfare dynamics in online debates. *European Journal of Communication*, *32*(3), 257–273.

Levene, H. (1960). Robust tests for equality of variances, 279–292.

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, *7*, 328.

Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, *2*, 266–283.

Linvill, D. L., Boatwright, B. C., Grant, W. J., & Warren, P. L. (2019). “THE RUSSIANS ARE HACKING MY BRAIN!” Investigating Russia’s Internet Research Agency Twitter tactics during the 2016 United States presidential campaign. *Computers in Human Behavior*.

Linvill, D. L., & Warren, P. L. (2018). Troll factories: Manufacturing specialized disinformation on twitter. Retrieved from <http://pwarren.people.clemson.edu/Troll_Factories_v2_Linvill_Warren.pdf>

MacKuen, M., Wolak, J., Keele, L., & Marcus, G. E. (2010). Civic engagements: Resolute partisanship or reflective deliberation. *American Journal of Political Science*, *54*(2), 440–458.

Mair, P., Hatzinger, R., & Maier, M. J. (2020). *eRm: Extended Rasch Modeling*.

Mejias, U. A., & Vokuev, N. E. (2017). Disinformation and the media: The case of Russia and Ukraine. *Media, Culture & Society*, *39*(7), 1027–1042.

Meyer, P. E. (2008). *Information-theoretic variable selection and network inference from microarray data* (PhD thesis). Universite Libre de Bruxelles.

Min, S. J., & Wohn, D. Y. (2018). All the news that you don’t like: Cross-cutting exposure and political participation in the age of social media. *Computers in Human Behavior*, *83*, 24–31.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330.

Ong, A. D., Benson, L., Zautra, A. J., & Ram, N. (2018). Emodiversity and biomarkers of inflammation. *Emotion*, *18*(1), 3–14.

Peng, T.-Q., Sun, G., & Wu, Y. (2017). Interplay between public attention and public emotion toward multiple social issues on twitter. *PloS One*, *12*(1).

Permanent Select Committee on Intelligence. (2018). Schiff statement on release of Twitter ads, accounts and data. Retrieved from <https://democrats-intelligence.house.gov/news/documentsingle.aspx?DocumentID=396>

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Oxford, England: Nielsen & Lydiche.

Richey, M. (2018). Contemporary Russian revisionism: Understanding the Kremlin’s hybrid warfare and the strategic and tactical deployment of disinformation. *Asia Europe Journal*, *16*(1), 101–113.

Roeder, O. (2018). Why we’re sharing 3 million russian troll tweets. *FiveThirtyEight*. Retrieved from <https://fivethirtyeight.com/features/why-were-sharing-3-millionrussian-troll-tweets/>

Schmidt, A. L., Zollo, F., Del Vicario, M., Bessi, A., Scala, A., Caldarelli, G., … Quattrociocchi, W. (2017). Anatomy of news consumption on facebook. *Proceedings of the National Academy of Sciences*, *114*(12), 3035–3039.

Skoric, M. M., Zhu, Q., Goh, D., & Pang, N. (2016). Social media and citizen engagement: A meta-analytic review. *New Media & Society*, *18*(9), 1817–1839.

Staiano, J., & Guerini, M. (2014). Depechemood: A lexicon for emotion analysis from crowd-annotated news. *arXiv Preprint arXiv:1405.1605*.

Stewart, L. G., Arif, A., & Starbird, K. (2018). Examining trolls and polarization with a retweet network. In *Proceedings of wsdm workshop on misinformation and misbehavior mining on the web*.

Straus, J. R. (2018). *Social media adoption by members of congress: Trends and congressional considerations (CRS Report No. R45337)*. Congressional Research Service.

Suhay, E., & Erisen, C. (2018). The role of anger in the biased assimilation of political information. *Political Psychology*, *39*(4), 793–810.

Sunstein, C. R. (2018). *# Republic: Divided democracy in the age of social media*. Princeton University Press.

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS One*, *13*(9).

Valenzuela, S. (2013). Unpacking the use of social media for protest behavior: The roles of information, opinion expression, and activism. *American Behavioral Scientist*, *57*(7), 920–942.

Valenzuela, S., Halpern, D., Katz, J. E., & Miranda, J. P. (2019). The paradox of participation versus misinformation: Social media, political engagement, and the spread of misinformation. *Digital Journalism*, *7*(6), 802–823.

Wollebæk, D., Karlsen, R., Steen-Johnsen, K., & Enjolras, B. (2019). Anger, fear, and echo chambers: The emotional basis for online behavior. *Social Media+ Society*, *5*(2), 257–273.

Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on twitter. *Bulletin of Science, Technology & Society*, *30*(5), 316–327.

Zannettou, S., Caulfield, T., De Cristofaro, E., Sirivianos, M., Stringhini, G., & Blackburn, J. (2019). Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web. In *Companion proceedings of the 2019 world wide web conference* (pp. 218–226).