Characterizing Susceptible Users on Reddit's ChangeMyView

Humphrey Mensah
Department of EECS
Syracuse University
Syracuse, NY
hamensah@syr.edu

Lu Xiao
School of Information Studies
Syracuse University
Syracuse, NY
lxiao04@syr.edu

Sucheta Soundarajan
Department of EECS
Syracuse University
Syracuse, NY
susounda@syr.edu

ABSTRACT

In the study of persuasion, little attention is paid to understanding features that indicate one's level of susceptibility. In this work, we examine features that are indicative of an individual's susceptibility on Reddit's changemyview. Specifically, we explore attributes about the author of the post, the interactions between an author and other users, and the author's language style. We first categorize authors of posts on changemyview into two groups: susceptible and non-susceptible. We perform a test of significance on different features between susceptible and non-susceptible authors. Experiments showed that an individual's language style can be indicative of one's susceptibility to a change of opinion. Also, an author's prior position on a subject and their way of interacting with other users can indicate the likelihood of an author having an opinion change.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in collaborative and social computing*;

KEYWORDS

Persuasion, susceptibility.

ACM Reference format:
Humphrey Mensah, Lu Xiao, and Sucheta Soundarajan. 2019. Characterizing Susceptible Users on Reddit's ChangeMyView. In *Proceedings of , 2019, SMSociety '19, July 19–21, 2019, Toronto, ON, Canada (SMSociety '19, July 19–21, 2019, Toronto, ON, Canada),* 6 pages.
https://doi.org/https://doi.org/10.1145/3328529.3328550

1 INTRODUCTION

There is a growing interest in understanding persuasion processes in various social media platforms, e.g., the influential users in an online community [17, 18], the types of persuasion attempts [2] and the indicators of a social media comment's persuasion power [10, 20, 27]. The majority of these research activities have focused on the side of pursuing persuasion, with only a few studies that examine the other side - those who are being persuaded [20, 23, 24].

SMSociety '19, July 19–21, 2019, Toronto, ON, Canada, 2019

ACM ISBN 978-1-4503-6651-9/19/07…$15.00
https://doi.org/https://doi.org/10.1145/3328529.3328550

With the goal of identifying the properties of susceptible and non-susceptible individuals, we analyzed the Reddit changemyview subreddit. In this subreddit, the author of a post makes a submission on an opinion and seeks comments from other users to change her opinion. If a user is successful in changing the author's initial opinion, the user is awarded a point referred to as a delta point. As an author of a post on changemyview, you are required not only to issue a delta point when your opinion changes but also required to explain the reasons for the change in opinion.

In the context of Reddit changemyview discussions, what then could be possible sources of information in characterizing the original posts? We consider three broader sources of information in segregating the submissions: (1) the prior position of the author regarding the topic. (2) the interactions between the author of a post and their challengers (those individuals that interact with the original author in an attempt to influence his or her views). (3) the language use in the post. According to [29], authors of articles in web-based communication channels are more likely to live their own "writeprints" because web-based channels are relatively casual in comparison to formal publications. We therefore believe an author's contains "writeprints" that could characterize how susceptible an author might be. With the interactions between an author of a post and their challengers, work suggested that an author's interactions with other users could be a useful means of segregating the authors.

Understanding the features that are indicative of an individual's susceptibility is useful in many regards. In the study of influence, identifying traits that characterize susceptible and non-susceptible users provides useful insights to understanding how different people can be influenced. For example, suppose user A is attempting to persuade two other users, B and C. If, for instance, user A finds out that user B is not somebody who typically changes their mind on a particular subject, but user C is one who is very susceptible, this means that the amount and style of persuasion as far as users B and C are concerned should be different, as a lot of effort will be required to persuade user C. In cases where user A does not have any information on the susceptibility of the users, then they are likely to be treated equally.

In this work, we explore features which can significantly separate users that change their mind all the time and users that never changed their mind on the Reddit subreddit, changemyview. Experiments showed that various authors have unique features that can aid in identifying how susceptible an author is to an opinion change. With respect to language use, susceptible users use more punctuation in their writing than non-susceptible users. They also demonstrate more uncertainty in their writing than non-susceptible users. In the interaction of authors and other users, we observed that users that changed their mind most of the time are interactivity at the early part of a conversation in comparison to users that never change their mind.

2 RELATED WORK

Researchers have been interested in studying the behavior of social media users in different contexts. Michael and Christoph [19] studied the interactions among users in online social media. Specifically, they studied reciprocity of communications among different users and observe that there are always some features characteristics that can in aid in the inference of reciprocity. With a similar goal of inferring the reciprocity of a communication between two individuals, the authors in [4] suggested that different features such as in-degree, the number of incoming and outgoing messages, etc. have high predictive power in relation to reciprocity prediction. In identifying spammers, Tan et. al[21] posit that user generated content spammers are characterized by some unique features. One notable behavior was that most user generated content spammers makes posts that contains links to other websites. The authors in [9] provide a survey of some of the past works on understanding user behavior for various tasks. Some of the tasks discussed include the study of behavior of users in Online Social Networks (OSNs) and its relations to traffic activities, the study of user's behavior and their reaction to spam.

An area of research that is closely related to susceptibility is the concept of persuasion. A lot of research has been conducted to understand the factor behind a message's persuasiveness. Various theories and models have been proposed to explain the role of contextual factors, such as social judgement [15], elaboration likelihood model [3], inoculation theory [12], cognitive dissonance [5], and narrative paradigm [25]. The aspects of a message's content that indicate its persuasive power have also been explored, such as its structure, comprehensibility, and credibility [13].

Different works have been done in studying persuasion in different forms. Jaech et. al [8] investigated how languages affects the reaction of members of a community. A support vector machine (SVM) model was trained using different features to predict the rank order of a list of comments. Some of the features used included the similarity of a comment to the original post, word count and usage of urls. It was observed that the usage of language features can improve the comment ranking task in different subreddits. Authors in [10, 20, 27] found that certain linguistic properties of comments are indicative of the persuasion power of a text.

Some of the identified features from these studies overlap. For example, all three studies suggest that the sentiment level of persuasive comments (i.e., emotional tone) is lower than that of non-persuasive comments. [10] and [27] found that persuasive comments tend to use more punctuation marks including periods, commas, colons, dashes, and apostrophes, but less on question marks.

There are also features that show contradicting indications across the studies. [27] found that non-persuasive comments tend to be longer and use words that have six letters or very slightly more, contradicting the results from [10]. While persuasive comments used fewer parentheses in [27], they used more in [10]. Also, while persuasive comments had less cognitive processing in [10, 27] showed the opposite. [27] offered explanations of the observed discrepancies and speculated that these are due to the two different discussion contexts in the two studies, namely, the Reddit "changemyview" discussions vs. Wikipedia's Article for Deletion discussions.

Besides these surface level linguistic features, prior studies also discovered that the structure of the comment helps characterize the persuasion power of the text. For example, the authors in [30] showed that argumentation based features such as the number of connectives in a comment are indicative of persuasiveness at early part of a conversation. And Tan et. al [20] observed that there are different features that can characterize persuasive argument. For instance, it was observed that users that enter a conversation very early are more likely to succeed in a persuasive argument.

There are few works that studied susceptibility of users. [23] investigated various features that are indicative of a Twitter user's susceptibility to tweets from social bots, such as network features, linguistic features and behavioral features. The authors observed that susceptible users interact more with other users, they tend to be more open and demonstrates more affection than non-susceptible users. With a similar goal of identifying users that are susceptible to social bots on Twitter, [24] examined features that are indicative of how susceptible users are to social bots. Authors observed that a Twitter user's Klout score, friends count, and follower count were the top predictors of the susceptibility of a Twitter user to social bots. Klout score is a metric that determines an individual's overall social influence computed using multiple social networking profiles. A fairly recent work by Williams et. al [26] provides a review on the individual differences and contextual factors that are capable of affecting susceptibility. Authors discuss how different features could have varying impact on different users. For instance, individuals high in self-awareness consider their personal knowledge to a higher degree than others making them less susceptible in some instances. However, self-awareness can make people more susceptible in cases where authors make persuasive charity messages and a user considers herself similar to the author of the post.

Even though our work has the same goal as works done in understanding indicative attributes of susceptibility, our work is unique in that it explores the susceptibility of authors to other users in a conversation.

3 METHODOLOGY

Our main goal is to explore features that are indicative of the susceptibility of users. We use the `changemyview` subreddit for this study.

3.1 Data and Preprocessing

The `changemyview` subreddit provides a means for individual users to make posts in order to be persuaded into an opinion change by other users on the forum. The author of a post on `changemyview` is referred to as an OP (original poster). When a user makes a comment that successfully changes an OP's initial opinion, the OP replies to the user with an explanation on why the view changed, and grants that user a so-called `delta point`. There are three main ways of indicating a delta point: Δ, `!delta`, `∆`. The `changemyview` subreddit allows users other than the OP to grant delta points if their opinions are changed, but this is rare. The forum specifies rules governing the issuing of delta and how users interact on the platform. In particular, an automated checking bot called *deltabot* ensures that a delta issued for a comment meets the following specifications [1]:

- the delta is not issued from users to themselves;
- an issued delta is accompanied with an explanation with at least 50 characters of text;
- the delta is not in response to the OP or the deltabot;
- the delta is not in a quote; and
- a delta has not been issued by that same user to the comment already;

The data used for this project was extracted from conversations made between January 2014 and December 2016. After excluding submissions with no text and/or no comments, a total of 212,404 submissions remained from 13812 unique OPs. We only considered submissions by those OPs that made at least two submissions, leaving 2,821 OPs that made a total of 10,549. We assume that OPs that made exactly one submission may not yet have a full grasp of how the forum works, and so may not understand the delta point system.

We categorize a submission as one on which the OP had a change of opinion only when a delta is issued by the OP of that submission and has been confirmed by the deltabot. Considering the deltabot's confirmation is necessary in that it prevents issuing a delta point without any justification. We ignore delta points issued by users other than the OP because they are not authors of the submission and we could not establish their position before their mind was changed. We consider two groups of OPs: `susceptible` and `non-susceptible` OPs. A susceptible OP is one that changed her mind on all submissions that she made, and a non-susceptible OP is one that never changed her mind on any of the submissions made. Even though the majority of OPs fall in the middle group of sometimes changing and sometimes not changing their minds, we choose to exclude such OPs. We make this decision because we believe that by studying the extreme groups, we will gain better insight into the factors behind susceptibility. 220 OPs were categorized as susceptible and 1,222 OPs as non-susceptible. A total of 474 submissions were made by susceptible OPs while 2,917 submissions were made by non-susceptible OPs.

3.2 Feature Identification

After identifying appropriate data, the next task is to identify the features indicative of how likely it is for an OP to have a change of mind. As discussed earlier, we consider three possible sources of features: The prior position of the author, interactions between OPs and their challenger, and the language use in an OP's post.

3.2.1 Language Usage by an author. **Procedure:** To study the language usage of an OP, we perform LIWC analysis on the submissions. LIWC has 93 features corresponding to different language dimensions. Some of these dimensions are pronouns, authenticity, verbs, positive emotions, negative emotions, etc. Previous work[23] investigated how susceptible users were to social bots. This work suggested that there are some linguistic properties that can characterize how susceptible humans were to social bots. We believe in studying how susceptible humans are to other humans, there will also be some linguistic properties that can characterize their susceptibility. We use the Linguistic Inquiry and Word Count (LIWC) tool to identify linguistic features that characterizes susceptible users. LIWC uses a word counting strategy to assign a score to a submission in different dimensions. We apply LIWC to submissions from susceptible and non-susceptible OPs. For comparison of submissions from the two groups on the different LIWC categories, we perform a non-parametric test of significance (Mann-Whitney Test). We selected a non-parametric test based on a kurtosis test.
Results and Discussion: The LIWC categories that showed significant differences ($\alpha = 0.0005$) are shown in Table 1. We use an initial α level of 0.05, which corresponds to a Bonferroni α of 0.005 after correction. The LIWC results indicate that users that changed their opinion generally use more punctuation relative to users that never changed their mind. We believe the use of punctuation in an individual's piece of writing makes that piece easier and clearer to understand. For example, consider the two sentences below:

S1: Let's eat John.

S2: Let's eat, John.

Even though the two sentences have the same words, the one with punctuation (S1) is clearer to understand than that of S2. Using more punctuation is therefore likely to make an OP's opinion more clearer to understand. If users clearly understand the opinion of an OP, then one of these might succeed in changing the OP's opinion and hence a possible reason why susceptible OPs use a lot of punctuation.

The authors in [16] posited that a user that is excited about a concept is likely to attract other users to that concept, and used exclamation mark usage as a means of measuring enthusiasm. It is therefore reasonable to argue that OPs that use more exclamation marks have a higher tendency to attract more users to their conversation. By attracting more commenters, the OP is likely to get diverse opinions from different users within which one might be successful in changing the OP's opinion.

According to [14], the analytic category in LIWC measures the degree to which one uses words suggesting a higher level of formal, logical, and hierarchical thinking. In [14], the authors used analytical thinking as a feature in characterizing suicidal Twitter posts. Submissions from an OP with higher analytical thinking might be difficult for other users to actually understand the OP's opinion to even attempt to change that. Also, even if lots of users attempt to change the opinion of an OP, only few of them might really understand what the OP really means, and a lot of effort will be required for OPs with such level of thinking to give in during a discussion. This could be a possible reason why users that displayed higher forms of thinking never changed their mind on any submission.

The "I" category of LIWC captures one's usage of first-person singular pronouns. The usage of many first-person singular pronouns indicates the drawing of attention to one's self. Our results indicate that users that used more first-person singular pronouns were more likely to change their opinion. This corroborates a previous finding in [20] that suggested that that people who use a lot of such pronouns are likely to be influenced during a discussion.

3.2.2 Prior Position of an Author. **Procedure:** We estimate an OP's prior position on submission by examining an OP's confidence on a subject. We explore the confidence of an author by examining

Table 1: LIWC categories that showed significant differences between the two groups of users(users that changed their mind all the time and users that never changed their mind)

	susc	non-susc
Allpunc	++	–
Exclam	++	–
Colon	++	–
Comma	++	–
SemiC	++	–
WC	++	–
Insight	++	–
Verb	++	–
Pronoun	++	–
I	++	–
Personal pronoun	–	++
Nonfluency	–	++
Analytic	–	++

the usage of hedge words (hedges) and booster words (boosters). Hedges refer to words that make issues difficult to understand [6, 11]. According to [22], people that are uncertain tend to use lots of such words. Boosters on the other hand refer to words used to express conviction and an indication of confidence in an asserted proposition. With hedges and boosters, we count the number of hedges and boosters used in an OP's submission. The hedges and boosters used in the experiment was provided by [7].

Results and Discussion: Figure 2 shows a box plot of the usage of hedge words and booster words by OPs. We observe that, on average, OPs that changed their mind use hedges more than those that never changed their mind. This observation is reasonable in that if an OP is uncertain with her opinion, then compared with an OP who is certain, the one with less certainty is more likely to change her opinion. This corroborates findings in [22]. For boosters, the expectation was that the confidence expressed by an OP could possibly deter other users from attempting to change the OP's opinion and hence succumbing to the view of the OP. However, that was not observed in the experiment. The insignificance in the usage of boosters among the two groups could be that OPs generally do not reveal how confident they are on a subject matter in their submission.

3.2.3 *OPs interactions with other users.*

Procedure: The authors in [20] showed that the interactions among users during a conversation in online social media is a significant source of information in identifying users that can succeed in successfully persuading other users. In [28], the authors find that debaters who follow up on points brought up by their opponents have higher chance of winning. These results suggest that the interaction dynamics of a conversation between the author of a post and other challengers could be a useful source of information in segregating susceptible and non-susceptible OPs. We considered three features as a way of capturing the interaction dynamics between an OP and other users: the number of unique users an OP engages in a back and forth with, the frequency of an OP's response, and when

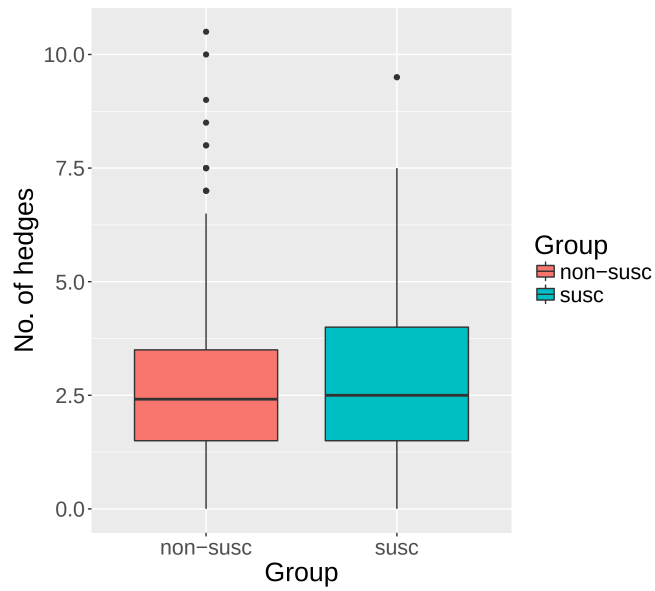

Figure 1: Box plot of the usage of hedge words. OPs that changed their mind all the time (susc) used more hedge words that never change their mind (non-susc) from the significance testing.

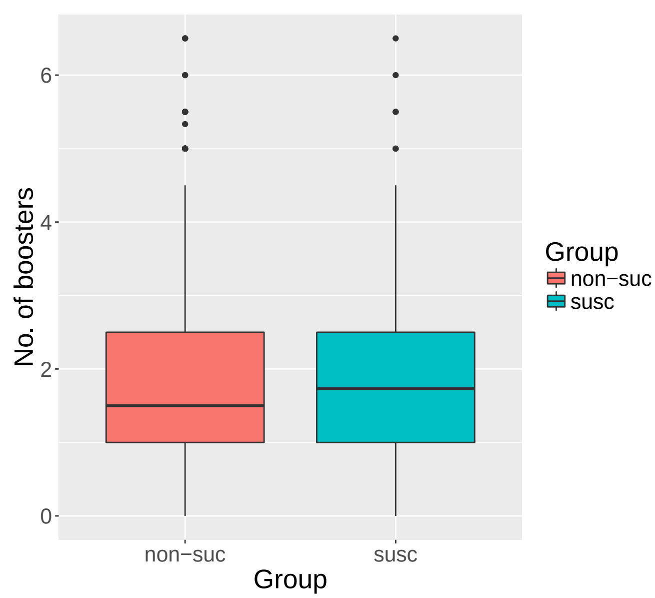

Figure 2: Box plot of the usage of booster words. There is no observed significant difference in the usage of booster words between OPs that changed their mind all the time (susc) and OPs that never changed (non-susc)

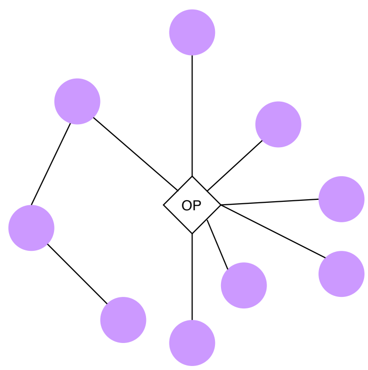

Figure 3: Illustration of the interaction network of an OP and users attempting to change OP's opinion.

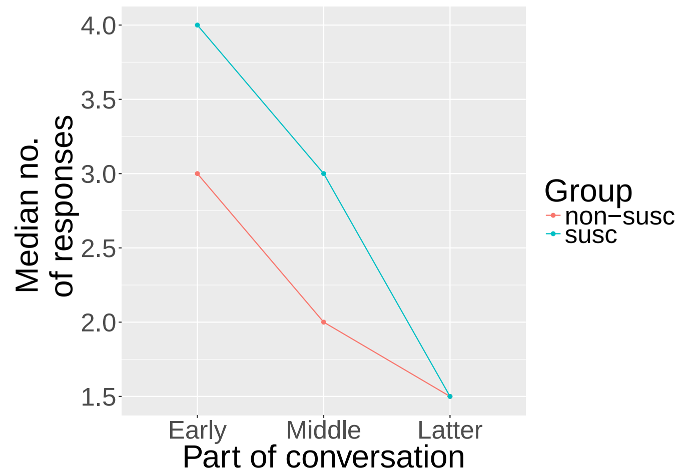

Figure 4: The number of interactions OPs made with other users at different parts of the conversation. Susceptible OPs engage with challengers more at the early part of the conversation

during a conversation are OPs active. Figure 3 shows a toy graph with the OP represented with a square and challengers as circles. Back and forth is defined as the OP replying back to a user that made a comment on the OP's submission, and the frequency of an OP's response is defined as the number of times the OP commented on another user's comment excluding delta replies. OPs are said to be active when they respond or comment on a user's post. The duration for conversation considered is the period between the first and last comment received by an OP after a submission was made. The duration for each conversation is partition into three parts: the early part of the conversation, the middle part of the conversation and the latter part of the conversation. For each part of the conversation, number of times an OP responded to other users is computed.

Results and Discussion: From our experiments, we observed that users that never changed their mind engage in more back and forth with their audience than users that changed their mind all the time ($p = 0.0004$). Among all the users that have made a comment on a submission made by an OP, if the OP engages in back and forth with just one of these users, then it could be argued that the back and forth can provide clarity to the opinion of the OP and hence the likelihood of OPs changing their initial opinion. However, if an OP engages in back and forth with one user and never changes the mind but instead engages several other users in such back and forth, such OPs are then less likely to change their opinion because they might be so firm in their opinion hence the reason why even when people try explaining their points, they never give in.

Also, we observed that OPs that changed their mind most of the time frequently interacted with others more than OP's that never changed their mind ($p = 0.004$). If an OP responds or comments on another user's post, then either the OP is seeking some clarification on the opinion of the user or simply disagrees with that opinion and is attempting to explain the reasons for her disagreement. For either case, it is reasonable to say that the OP is somewhat paying attention to the user's opinion. An OP who is indifferent to many users in a conversation is therefore less likely to change the view in comparison to one that is paying attention to the views of others. This is because if an OP pays attention to several other users, there is a higher chance that one of the users might make a point which could change the OP's initial opinion.

Figure 4 shows the number of responses made by OPs at different parts of the conversation. We observed that even though all OPs generally decrease their interactions with other users towards the end of a conversation, users that changed their mind all the time interact more with other users at the early and middle part of the conversation. Previous work [20] had suggested that users that enter a conversation late after the submission has been posted is less likely to succeed in changing an OP's opinion in comparison to users that enter early. This suggests that OPs that are susceptible are likely to be active at the early part of the conversation. If an OP is susceptible, then after having an opinion change, the OP might not be as active as she was before since there is an opinion change.

4 CONCLUSION

In this work, we investigated features that are useful in segregating susceptible and non-susceptible OPs on reddit's `changemyview`. We explored three main sources of information in characterizing users on this forum (1) the OP of a post (2) prior stance of the OP before seeking an opinion change and (3) the interactions between OPs and their challengers. For the prior stance of an OP, we explored how much confidence is expressed by an OP in a submission. In measuring confidence, we used an OP's hedge/booster words usage as a way of characterizing the confidence. For interactions between an OP and their challengers, we explored the number of unique users the OP engages with back and forth, the number of responses made by an OPs and their challengers and which part of a conversation are OPs active. We performed LIWC analysis as a means of understanding an OP's language usage in a post.

Experimental results showed that OPs who never changed their mind are more analytical in thinking when writing than susceptible OPs. Also, susceptible OPs use more hedge words than non-susceptible users. This means OPs who changed their mind most of time have more uncertainty in their submissions than non-susceptible users. On an OP's interaction with other users, susceptible users tend to interact with their challengers more at the early part and middle part of the conversation.

Our goal is to discover the differences between susceptible and non-susceptible users in their digital traces. Subsequently, our comparison of their submissions is intended to discover the differences in language use between these two groups of users. On the other hand, the grouping of these submissions can also be interpreted as merely by whether or not OPs changed the original view. This implies that the differences we observed could be interpreted as merely the differences between the two types of submissions, not the two types of users. We will collect the users' other Reddit comments in the next comparison to address this confounding issue.

Our observations are also likely specific to Reddit's Change-MyView context. In our future work, to be able to make a general conclusion about an OP on Reddit as whole, we will collect the users' other Reddit post/comments. If similar observation is made with a user's comments on Reddit, then we could generalize the susceptibility of the users to their presence on Reddit as a whole. Also, for us conclude that an OP that is susceptible on Reddit based on these observed differences is susceptible in general, we will be investigating the users on different platforms other than Reddit. Presently, we do not explore the relationships between different topics and the susceptibility of users on `changemyview`. As a future work, we will also be investigating the roles that topics play in changing one's opinion. Additionally, we will investigate the factors that causes some OPs to sometimes change or not to change their minds.

REFERENCES

[1] [n. d.]. The delta system. ([n. d.]). https://www.reddit.com/r/changemyview/wiki/deltasystem

[2] Pranav Anand, Joseph King, Jordan Boyd-Graber, Earl Wagner, Craig Martell, Doug Oard, and Philip Resnik. 2011. Believe me - we can do this! Annotating persuasive acts in blog text. In *Computational Models of Natural Argument.*

[3] John Cacioppo and Richard Petty. 1986. The Elaboration Likelihood Model of Persuasion. *Communication and Persuasion* (1986), 1–24.

[4] Justin Cheng, Daniel Romero, Brendan Meeder, and Jon Kleinberg. 2011. Predicting reciprocity in social networks. In *Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. IEEE, Boston, MA, USA, 49–56. https://doi.org/10.1109/PASSAT/SocialCom.2011.110

[5] Leon Festinger. 1957. *A theory of cognitive dissonance.*

[6] Ken Hyland. 1998. *Hedging in scientific research articles.* John Benjamins Publishing Company.

[7] Jumayel Islam. [n. d.]. Tension analysis. ([n. d.]). https://www.github.com/jumayel06/Tension-Analysis/tree/master/resources

[8] Aaron Jaech, Vicky Zayats, Hao Fang, Mari Ostendorf, , and Hannaneh Hajishirzi. 2015. Talking to the crowd: What do people react to in online discussions?. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2026–203.

[9] Long Jin, Yang Chen, Tianyi Wang, Pan Hui, and Athanasios Vasilakos. 2013. Understanding user behavior in online social networks: a survey. *IEEE Communications Magazine* 51, 9 (2013), 144–150. https://doi.org/10.1109/MCOM.2013.6588663

[10] Taraneh Khazaei, Xiao Lu, and Robert Mercer. 2017. Writing to persuade: analysis and detection of persuasive discourse. *iConference 2017 Proceedings* (2017), 203–215.

[11] George Lakoff. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of philosophical logic* 2, 4 (1973), 458–508.

[12] William McGuire. 1964. Inducing resistance to persuasion: Some contemporary approaches. *Advances in experimental social psychology* (1964), 191–229.

[13] P.Karen Murphy. 2001. What makes a text persuasive? Comparing studentsâĂŹ and expertsâĂŹ conceptions of persuasiveness. *International Journal of Educational Research* 35, 7 (2001), 675 – 698. https://doi.org/10.1016/S0883-0355(02)00009-5

[14] Bridianne O'Dea, Mark Larsen, Philip Batterham, Alison L. Calear, and Helen Christensen. 2017. A linguistic analysis of suicide-related Twitter posts. *The Journal of Crisis Intervention and Suicide Prevention* 38, 5 (2017), 319–329. https://doi.org/10.1027/0227-5910/a000443

[15] Daniel O'Keefe. 1990. *Persuasion: Theory and research.* Vol. 2.

[16] Shrimai Prabhumoye, Samridhi Choudhary, Evangelica Spiliopoulou, Christopher Bogart, Carolyb Rose, and Alan Black. 2017. Linguistic markers of influence in informal Interactions. In *Proceedings of the Second Workshop on NLP and Computational Social Science.* 53–62.

[17] Daniele Quercia, Jonathan Ellis, Licia Capra, and Jon Crowcroft. 2011. In the mood for being influential on Twitter. In *IEEE Third International Conference on Privacy, Security, Risk and Trust and IEEE Third International Conference on Social Computing*. 307–314. https://doi.org/10.1109/PASSAT/SocialCom.2011.27

[18] Fabián Riquelme and Pablo González-Cantergiani. 2016. Measuring user influence on Twitter. *Inf. Process. Manage.* 52, 5 (Sept. 2016), 949–975. https://doi.org/10.1016/j.ipm.2016.04.003

[19] Michael Steurer and Christoph Trattner. 2013. Who will Interact with whom? A case-study in second life using online social network and location-based social network features to predict interactions between Users. In *Ubiquitous Social Media Analysis.* 108–127.

[20] Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International Conference on World Wide Web.* 613–624.

[21] Enhua Tan, Lei Guo, Songqing Chen, Xiaodong Zhang, and Yihong Zhao. 2012. Spammer behavior Analysis and detection in user generated content on social networks. In *2012 IEEE 32nd International Conference on Distributed Computing Systems.* 305–314. https://doi.org/10.1109/ICDCS.2012.40

[22] Yla Tausczik and James Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29, 1 (2010), 24–54. https://doi.org/10.1177/0261927X09351676

[23] Claudia Wagner, Silvia Mitter, Christian Körner, and Markus Strohmaier. 2012. When social bots attack: modeling susceptibility of users in online social networks. In *In Proceedings of the 2nd Workshop on Making Sense of Microposts held in conjunction with the 21st World Wide Web Conference 2012.* 41–48.

[24] Randall Wald, Taghi Khoshgoftaar, Amri Napolitano, and Chris Sumner. 2013. Predicting susceptibility to social bots on Twitter. In *IEEE 14th International Conference on Information Reuse Integration (IRI).* 6–13. https://doi.org/10.1109/IRI.2013.6642447

[25] Fisher Walter. 1984. Narration as a human communication paradigm: The case of public moral argument. *Communication Monographs* 51, 1 (1984), 1–22.

[26] Emma J. Williams, Amy Beardmore, and Adam N. Joinson. 2017. Individual differences in susceptibility to online influence: a theoretical review. *Computers in Human Behavior* 72 (2017), 412 – 421. https://doi.org/10.1016/j.chb.2017.03.002

[27] Lu Xiao. 2018. A message's persuasive features in Wikipedia's article for deletion discussions. In *Proceedings of the 9th International Conference on Social Media and Society.* ACM, New York, NY, USA, 345–349. https://doi.org/10.1145/3217804.3217942

[28] Justine Zhang, Sujith Kumar, Ravi, and Cristian Danescu-Niculescu-Mizil. 2016. Conversational flow in Oxford-style debates. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* 136–141. https://doi.org/10.18653/v1/N16-1017

[29] Rong Zheng, Jiexun Li, Hsinchun Chen, and Huang Zan. 2006. A framework for authorship identification of online messages: writingâĂŘstyle features and classification techniques. *Journal of the American Society for Information Science and Technology* 57, 3 (2006), 378–393. https://doi.org/10.1002/asi.20316

[30] Wei Zhongyu, Liu Yang, and Li Yi. 2016. Is this post persuasive? ranking argumentative comments in online forum. In *Association of Computational Linguistics.*