Understanding Abusive Behaviour Between Online and Offline Group Discussions

Abdulwhab Alkharashi
University of Glasgow
Glasgow, Scotland, UK
a.alkharashi.1@research.gla.ac.uk

Tim Storer
University of Glasgow
Glasgow, Scotland, UK
Timothy.Storer@glasgow.ac.uk

Joemon Jose
University of Glasgow
Glasgow, Scotland, UK
Joemon.Jose@glasgow.ac.uk

Andrew Hoskins
University of Glasgow
Glasgow, Scotland, UK
Andrew.Hoskins@glasgow.ac.uk

Catherine Happer
University of Glasgow
Glasgow, Scotland, UK
Catherine.Happer@glasgow.ac.uk

ABSTRACT

Online discussion platforms can face multiple challenges of abusive behaviour. In order to understand the reasons for persisting such behaviour, we need to understand how users behave inside and outside a community. In this paper, we propose a novel methodology to generate a dataset from offline and online group discussion conversations. We advocate an empirical-based approach to explore the space

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland Uk

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5971-9/19/05…$15.00
https://doi.org/10.1145/3290607.3299042

of abusive behaviour. We conducted a user-study ($N = 15$) to understand what factors facilitate or amplify forms of behaviour in cases of online conversation that are less likely to be tolerated in face-to-face. The preliminary analysis validates our approach to analyse large-scale conversation dataset.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; *User studies*; • **Security and privacy** → *Social aspects of security and privacy*.

KEYWORDS

Abuse behaviour; Online communities; Group discussion

ACM Reference Format:

Abdulwhab Alkharashi, Tim Storer, Joemon Jose, Andrew Hoskins, and Catherine Happer. 2019. Understanding Abusive Behaviour Between Online and Offline Group Discussions. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts), May 4–9, 2019, Glasgow, Scotland Uk*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3290607.3299042.

INTRODUCTION

Online discussion platforms are today commonplace in facilitating remote communications via a range of collaborative and sharing services. For example, Reddit an online discussion platform is often conceptualised as social media. The average monthly active users in Reddit is more than 330 million users and 138 million active communities [13]. There are also other forms of online group discussion which limit the number of members per group and offer other features, e.g, Facebook group, Google group, WhatsApp group and Slack channels . The anecdotal evidence [3] suggests that standards of behaviour in online forums is different from in person norms and perceptions. There is a considerable body of work concerned with measuring and analysing a range of online interactions [4, 9]. Our work, by contrast, makes more explicit comparisons between behaviour on and off-line, in particular, community interaction in shaping user behaviour. This group or abusive behaviour influences the way users react and think. For example, an exploratory study [6] examined perception of abusive behaviours that differed across three countries to learn about the mistreatment of a child, and concluded that misunderstanding the perception of cross-cultural or socio-cultural differences can lead to abusive behaviours.

The key contribution of this paper is the development of a dataset of conversation transcripts for the comparison of behaviour between online and face-to-face small group discussions. Our overarching aim is to explore the aspects of technology mediated discussion that have an impact on user behaviour, using in-person discussions as a baseline. We also seek to investigate the influence of a range of characteristics (e.g, age, gender, nationality etc.) in shaping perceptions of behaviour in small group discussions. The research questions directing this research are listed in Sidebar 1. To address these questions, we have designed an observational and experimental research method described in this paper and begun to collect data. In summary, we first apply a pre-discussion survey to obtain some

RQ1 What differences arise in standards of behaviour between online and in-person discussions?

RQ2 What impact does participant background (age, gender, nationality, culture) have on behaviour and acceptability of behaviour in online discussions?

RQ3 How do the technological features different platforms afford different cultures of acceptability?

Sidebar 1: List of research questions

Factors of Abusive Behaviour

Abusive behaviour in human subjects has been examined by using the Affective Events Theory to measure the relationship between abusive supervision and workplace deviance [11]. The process in general, shows how people can influence or affect the overall task performance of individuals in one specific community. In particular, it leads to examining the well established the Five Factor Model of personality (FFM) that differentiates between differences among individuals' temperaments [7]. The model uses six measures to predict effects and consequences from particular events. Previous work has addressed several motivations for participating in abusive behaviour: form of aggression and moral disengagement among youth that caused stress and suicidal ideation [10] and examples of vandalism behaviours and motivations (e.g., boredom, attention seeking, and revenge) [15]. Nevertheless, these studies have been mostly qualitative and non-causal, and their generalisability is not clear. Another recent study [3] uses two trigger mechanisms of abusive behaviour; mood and discussion context to try to create their effects using both a controlled experiment and a large-scale longitudinal analysis.

Figure 1: Steps for building the dataset.

details of their background, context and perspectives. Then, participants were asked to join an online or in-person group discussion about a specific topic. We record a transcript of each discussion for later analysis. Finally, after the end of each discussion each participant is asked to identify examples of acceptable and unacceptable behaviour they may have just experienced. We envisage employing a range of techniques for quantifying the impact of technology on behaviour in online discussions, as well as the contributing factors, including lexical analysis and sentiment analysis.

BACKGROUND

This section provides an overview of existing work on factors abusive behaviour including: aggression and Affective Events Theory, user behaviour, and features for abusive message detection.

Online and Offline Abusive Behaviour

The abusive behaviour in online communities can be observed similarly as offline behaviour which is formed by aggression acts and cyber-bullying [2, 14]. These studies were looking for the similarities and the differences existing in particular groups and concluded that each behaviour is mostly influenced by direct personal and contextual factors. Huang et al. [8] studied how people use online event invitations and how it effects their online and offline engagement— showed that social perceptions about particular event or invitation, can affect their decision before they chose to join or not.

Detection Features of Abusers

Chen et al. [5] presented predictive techniques to detect abusive behaviour in three diffrent online discussion communities by measuring the edit activity of two groups of users: future banned and never banned users. This feature is useful since it has been shown that abusive comments tend to prompt a higher number of replies in an online community. Another approach [12] uses Markov chain to compute the average emission probabilities of the n-grams in a user-specific conversation before and after the targeted message. Out of this review, we select some motivational aspects for measuring the abusive behaviour of group discussions from online platform and face-to-face data using our proposed methodology for building the dataset described in the next section.

METHOD

In this work, we seek to understand the factors which influence behaviour in group discussions. To understand the impact of different factors on behaviour in discussions, we are building a dataset of conversation transcripts captured from both face-to-face and online-discussions. To this end, We have identified the experimental variables that may be configured in each case. To gather each discussion transcript for the experiment, we follow the process illustrated in In summary, we conducted a pre-discussion survey identifying background information for recruits; conducted discussion sessions,

Group	Participant	Gender	Age	Nationality
G1	P1	F	21-29	American
	P2	M	30-39	Saudi
	P3	M	21-29	British
G2	P4	F	21-29	Lithuanian
	P5	F	30-39	Italian
	P6	M	21-29	Saudi
	P7	M	21-29	British
G3	P8	F	21-29	Polish
	P9	M	30-39	Saudi
	P10	M	21-29	Indian
G4	P11	F	30-39	Egyptian
	P12	M	30-39	Saudi
	P13	M	30-39	Nigerian
	P14	M	30-39	Saudi
	P15	M	21-29	Canadian

Table 1: Demographics of the groups.

Q1: Should the death penalty be allowed?
Q2: Is college education worth it?
Q3: Can you really be obese yet healthy?
Q4: Is Manchester United F.C. better than Liverpool F.C.?
Q5: Are social networking sites good for society?
Q6: Should abortion be legal?

Table 2: Example topic questions from the Pro/Con and Google Trends websites.

Group	Offline	Online
G1	Q1	Q5
G2	Q5	Q1
G3	Q1	Q5
G4	Q5	Q1

Table 3: Assigned discussion topics.

comprising one or more discussions and then applied a post-discussion survey to assess their experience of the behaviour of participants during the discussion. Figure 1.

Recruitment

For our initial dataset, we advertised opportunistically among students at the University of Glasgow who are currently enrolled in an Engineering and Science courses. All students were therefore near to taking their final exams in April 2018. In our advertising we explained the nature of the study and offered participants the opportunity to win a prize of a £25 Amazon gift card. Respondents to our advertising were directed to an online survey that explained more details about the study, ascertained that they were aged 18 or older and obtained their consent and gathered basic background information, as well as an email address so that we could issue invitations to discussions at a later date. The background information included the participants' age, gender, nationality, education and experience of social media platform discussions.

Respondents

We received a total of 25 respondents (64% male), (52% aged 23-29) (40% PhD student); completed the pre-survey and recruited to participant in the study from the University of Glasgow. The 15 participants (60%) were recruited to discuss one topic (either Q1 or Q5) face-to-face and use the online discussion website to discuss the other topic that has not been discussed offline. The remaining 10 participants were not included in this study since they have not attend the face-to-face meeting in order to complete the tasks. The selected topics (Q1 and Q5) listed in Table 2, cover political views and social science and technology. Offline group discussion spent on average 12 minutes discussion and 5 minutes in online discussion. Table 1 summarizes the demographics of participants and Table 3 shows topic discussion assignment per group.

Study Procedure

Participants were asked to express their interest in a number of different topic categories (politics, sport, technology and so on) so that they could be matched to discussion groups of common interest. To reduce the risk of a group conflict, we only recruited students from one college. Following the recruitment phase, each participant received an email inviting them to participate in a small group discussion. The parameters for the discussion were configured to create a 'baseline' scenario for comparison with future data sets. We divided the participants equally into online and in-person discussion groups of four or five. Online discussions were held synchronously in our platform and all discussions lasted for at most 20 minutes. The topics selected for discussion were Q1 and Q5 listed in Table 2. Each group discussed one topic online and the other offline. All participants played the same role, with no moderator appointed. Advanced online discussion features, such as threading, content

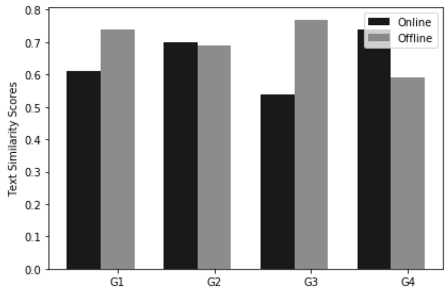


Figure 2: Text similarity between online and offline groups.

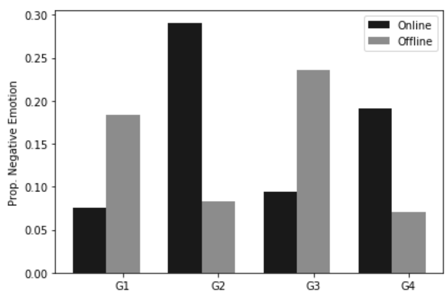


Figure 3: Polarity between online and offline groups.

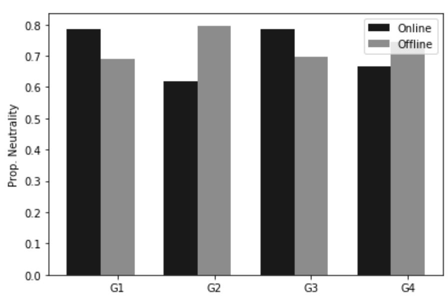


Figure 4: Subjectivity between online and offline groups.

rating and user-content editing were disabled. Online participants were given pseudo-anonymous identifiers and notified that the discussion platform would not be publicly visible by anyone other than the participants and the researchers. For each transcript obtained, we record the configuration of the discussion for future analysis as well as the transcript of contributions. Upon the completion of each group discussion, each participant is asked to complete a post-survey. The survey was used to understand how each participant evaluates abusive behaviour in each topic differently by asking participants to give examples of acceptable or none acceptable discussion points. We seek to see expectations of group discussions, e.g. background or gender perceptions, and how they engage discussions in different group settings and topics.

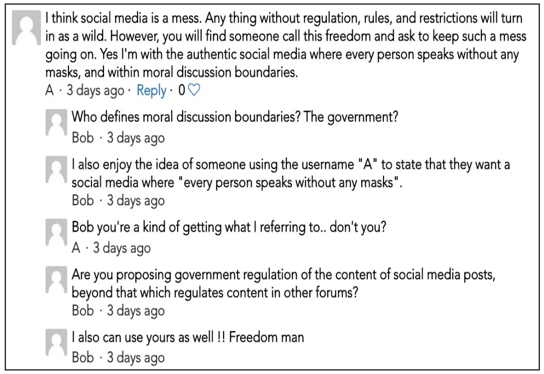
Ethics

The study design received ethical approval from the University of Glasgow (number 300170156). In all cases, conversations were monitored by the researcher. We ask participants to discuss potentially controversial topics, e.g. on politics or sport. We strongly emphasised in the information sheet that if at any point in the discussion participants feel uncomfortable they are encouraged to leave. Also in extreme circumstances the researcher would end a discussion prematurely if necessary.

PRELIMINARY FINDINGS

Differences between Online and Offline Behaviour (RQ1)

Finding the similarity scores of conversations between online and offline topic discussion is significant to see how users can remain on-topic or swerve off-topic. Previous work [1] has conducted research to examine the impacts and nature between online and offline behaviour on anti-Muslim hate crime; shown that abusers tend to adopt linguistic conventions. Here, we compare the average text similarity of a user's respond with the other group who contributed in a similar thread, obtained by computing the cosine similarity of words used in these offline conversations and online posts. We find that the text similarity (shown in Fig. 2) of posts written by online groups is significantly lower than that of offline groups on Q1. This might suggest that online political discussions make less of an effort to remain on-topic. We also find that online groups are likely to use negative words (shown in Fig. 3) in society and technology discussion (Q5) than offline groups and use more negative words in political discussion. The subjectivity analysis in Fig. 4 shows how each group feel neutral discussing a specific topic online or offline. In particular, we see that the society and technology topic was discussed more neutrally online, and the social network subject shows less interest in death plenty discussion. On average, the number of replies (depicted in Fig. 6) is higher in offline groups in general, but topic category and group structure may affect the number of replies, e.g., G1 in most cases.



(a) Online Group Discussion

P13: *My friend from Ghana told me oh do not buy phone from any shop, I will get you one from Facebook, they are cheap in Facebook.*
P15: *I guess largely the way things we use it is not necessarily technology inherently being negative itself, it is the way people we interact with one another.*
P11: *Everything has good side and bad side. But it is filling a gap too, because like millions of users use it down so it is actually good thing apparently.*

(b) Offline Group Discussion

Figure 5: Example of online/offline form Q5 topic discussion.

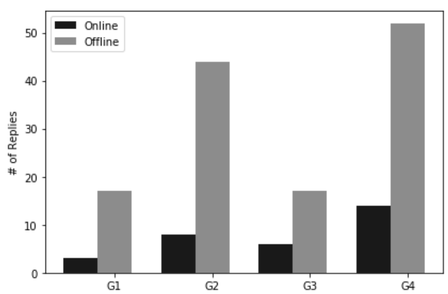


Figure 6: # Replies between online and offline groups.

Background Implications (RQ2)

The average gender and age ratio for each sampling is ≈ 2:1. Participants were asked to avoid interaction with fellow group members and unlikely to have met before joining a group discussion. We then compare both online and offline conversations (see an example in Fig. 5 from Q5 group discussion) with random groups and same topic. About (60%) of the participants are members of a topic focused group on a social networking communities and (76%) have met friends on social/discussion websites in person. It is also interesting to see that some users think that they might accept friend request from stringers about (36%) and about (92%) often sees unacceptable edits online. Several users (28%) have experienced online offensive comments.

Measuring Acceptability (RQ3)

To understand the group dynamics clearly, we use three measures to explore the differences between both settings. The measurement of acceptability includes the number of replies, negative emotions and neutrality. The number of replies measure has failed to show statistical significance *($P < 0.66$)*. The remaining three measures have rejected the null hypotheses of statistical significance as shown in Table 4. This can imply that our methodology of building the conversational dataset validated the initial results for the polarity, similarity and subjectivity measures, yet need further samples to validate the number of replies measure or observe other factors.

Measure	R^2	Adjusted R^2	p-value
Polarity	0.68	0.52	0.05*
Similarity	0.85	0.78	0.02**
Subjectivity	0.95	0.93	0.002***

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4: Results for the statistical significance of the groups discussions.

DISCUSSION

Factors of cohesiveness. We observed that background similarity among group members affect the interaction during each topic discussion. For example, G2 and G4 were active more in both topic discussions. This might indicate that the age factor may influence the number of continuation regardless of the size of group.

(Un) Acceptable behaviour. Forms of acceptability can vary from one individual to another based on multiple factors related to background differences and so on. Unacceptable behaviour in all forms, in general, may involve actions, words or gestures that can sensibly be recognized to be the cause irritation or anguish of an individual. In the post-discussion questionnaire, participants reports the following as an acceptable behaviour in group discussion: *people talking without violence* (P4), *using social media to build a successful business* (P8). As an unacceptable behaviour in group discussion, participants also share the following: *using social media to influence people opinions* (P8), *using other people names* (P4), *breaching my privacy* (P6).

Anonymity. Anonymity may impact the level of trust when it comes to communicating. Most online users chose to not disclose their identity, which decreased the number of continuation over time. It is also possible that when people might have privacy concerns to a particular unknown platform, or may not feel directly and personally involved in the threads.

Limitations. Although, the sample size is small on this study, it took substantial effort to recruit participants to join the group discussion in both settings. Our sample is not a representative of all discussion platforms users, yet provides an insight of exploring the causes of abusive behaviour. We also learned that recruiting groups to contribute in both online of offline settings can lead to sampling bias problems. Instead, we need to ask each group to participate in one particular discussion mode.

Future work. In the longer term, we anticipate mimicking the discussion formats of popular online social media platforms to collect large-scale of data. We will expand the dataset with a larger longitudinal study with a larger group and also considering the qualitative analysis to observe group interaction and behavioural changes over a period of time in a student group project.

CONCLUSION

In this study, we conducted a preliminary user-study to build a dataset of conversation transcripts to analyse the behaviour changes between online and offline group discussion. We built an online discussion platform to interrogate the meanings of abusive behaviour based on user perceptions. We presented a new method to collect conversation data to build a dataset and to assess behaviour in different settings of online discussion communities. Our preliminary analysis suggests the opposite of what we initially expected. Surprisingly, finds significant changes in behaviour between online and offline group discussion (i.e., online group discussions lead to swerve off-topic in political topics). We explored and expectations of acceptability behaviour from a user perspective and measure the actual behaviour of each group.

REFERENCES

[1] Imran Awan and Irene Zempi. 2016. The affinity between online and offline anti-Muslim hate crime: Dynamics and impacts. *Aggression and violent behavior* 27 (2016), 1–8.
[2] José A Casas, Rosario Del Rey, and Rosario Ortega-Ruiz. 2013. Bullying and cyberbullying: Convergent and divergent predictor variables. *Computers in Human Behavior* 29, 3 (2013), 580–587.
[3] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 1217–1230.
[4] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2014. How Community Feedback Shapes User Behavior. In *Eighth International AAAI Conference on Weblogs and Social Media*.
[5] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2015. Antisocial Behavior in Online Discussion Communities.. In *ICWSM*. 61–70.
[6] Sunday B Fakunmoju, Funmi O Bammeke, Thomas Antwi Bosiakoh, Rabiu Kwaku Boakye Asante, Nikki R Wooten, Anthony C Hill, and Hannah Karpman. 2013. Perception and determination of child maltreatment: Exploratory comparisons across three countries. *Children and Youth Services Review* 35, 9 (2013), 1418–1430.
[7] Lewis R Goldberg. 1993. The structure of phenotypic personality traits. *American psychologist* 48, 1 (1993), 26.
[8] Ai-Ju Huang, Hao-Chuan Wang, and Chien Wen Yuan. 2014. De-virtualizing social events: understanding the gap between online and offline participation for event invitations. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 436–448.
[9] Imrul Kayes, Nicolas Kourtellis, and Adriana Iamnitchi. 2017. The good, the bad and the deviant in community question answering. *Online Social Networks and Media* 2 (2017), 45–59.
[10] Robin M Kowalski, Gary W Giumetti, Amber N Schroeder, and Micah R Lattanner. 2014. Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological bulletin* 140, 4 (2014), 1073.
[11] Jesse S Michel, Kerry Newness, and Kris Duniewicz. 2016. How abusive supervision affects workplace deviance: A moderated-mediation examination of aggressiveness and work-related negative affect. *Journal of Business and Psychology* 31, 1 (2016), 1–22.
[12] Etienne Papegnies, Vincent Labatut, Richard Dufour, and Georges Linarès. 2017. Detection of abusive messages in an on-line community. In *14ème Conférence en Recherche d'Information et Applications (CORIA)*. 153–168.
[13] Reddit. 2018. Reddit by the Numbers. https://www.redditinc.com Accessed: 11 April 2018.
[14] Howard Rheingold. 2000. *The virtual community: Homesteading on the electronic frontier*. MIT press.
[15] Pnina Shachaf and Noriko Hara. 2010. Beyond vandalism: Wikipedia trolls. *Journal of Information Science* 36, 3 (2010), 357–370.