(Dis-)Respectful Public Discussions Online: Insights on Audience Polarization and Formation of Radical Hate or Support Groups

Inesa BIRBILAITĖ
i.birbilaite@pmdf.vdu.lt
PhD, Lecturer
Department of Public Communications
Vytautas Magnus University
Kaunas, Lithuania

Abstract: Debates related to the quality of online discussions are quickly engaging scholars from different fields, but still, there is no single answer if we can expect online discussions to be deliberative enough for the establishment and maintenance of a well-functioning public sphere online. In this paper, I presume that respect is an essential category determining overall quality of online public discussions. Therefore, I assess if discussions from a preselected Facebook page on climate change are respectful. I find that in comparison to similar face-to-face settings, members of preselected discussions are quite respectful to each other. On the other hand, foul language is often used to address outside actors or groups. In turn, I link these results to the phenomenon of audience polarization online and conclude that in my sample, discussions cannot be of good quality because they are dominated by similarly thinking members, which possibly leads to polarization and generation of support and/or hate groups.

Keywords: audience polarization, hate and support groups online, respectful discourse, deliberation
INTRODUCTION

Respect is a crucial category in determining the quality of public discussions as well as an important indicator revealing the existence of a well-functioning public sphere\(^1\) online. If the discussions between citizens and policy makers are based on disrespectful and offensive language, it cannot be expected that rational deliberation will be reached and a well-functioning public sphere will be established. Therefore, in this paper I measure the level of respect in pre-selected Web 2.0 based online communication environments in order to assess and define the quality of public discussions online and evaluate potentials for a well-functioning public sphere to emerge and develop in Web 2.0 based online communication environments. But first, why it is important to discuss the quality of virtual public discussions and the formation of a well-functioning public sphere online? In this paper, I consider a number of reasons, which are closely interconnected and condition each other; and which explain the possible importance of public online discussions.

First of all, emerging new global risk-related problems, such as climate change, virtual wildfires, and health crises, among others, require crucial changes in global and local political and social systems, including changing relations between citizens and policy makers, which in turn raises the question of shifting democratic traditions (i.e. deliberative turn). In this sense, discussions online could possibly constitute a core of deliberations where major issues could be discussed between citizens and preliminary decisions reached.

Secondly, a deliberative turn in policy making is determined also by the recognition of limitations and uncertainties of traditional political and scientific practices (especially related to the emergence of phenomenon of post-normal science or the politics of uncertainty) indicating that any scientific or political decision can be questionable and is not absolute. This means that there is no single and correct answer; therefore, we ourselves (in consultations with experts) have to decide which way to take it.

Such a situation requires reassessment of traditional forms of policy and science making, preferably by increasing the level of democratization (e.g., democratization of science and deliberative democracy). Hence, the third point deals with democratization, which is

\(^1\) In this paper, I perceive the concept of a well-functioning public sphere following Habermasian tradition. Specifically, I assume that a public sphere is well-functioning when it corresponds to the main criteria listed by Habermas – participation, respect, listening, justification, force of better argument, and truthfulness; however, here I limit my analysis to the category of respect, as I believe it is the most essential in defining level of audience polarization, which is the major obstacle for quality deliberation to occur.
perceived as a stronger and closer connection between citizens and policy makers. It is important because if the citizens’ positions are not properly considered by policy makers, the gap between the citizens’ aspirations and satisfaction with democratic politics increases, in turn causing democratic deficit and leading to declining public trust and support for political actors (Norris, 2011). The public’s trust and support for political decisions are indeed important when dealing with new global risks, because they require global solutions, which are based on local or even personal initiatives and actions. In other words, there is no global effect if actions are not taken at a personal level. To foster and maintain the quality of recent democracies, communication between society and policy-makers are crucial.

Fortunately, new information and communication technologies provide current democracies with unprecedented virtual environments, where two-way communication between citizens and their representatives becomes possible. New discursive spaces are being established, enabling new and possibly more democratic relations between citizens and policy makers. Some scholars express huge hopes for the Internet as a new type of public sphere, (Kenix, 2008; Dutton, 2009; Armstrong and Zúñiga, 2006; Reynolds and Ball, 2006; Benkler, 2006; Xenos and Bennett, 2007; Dahlgren, 2005) arguing that online communication environments help to improve communication practices (e.g., enable multi-directional communicational modes) and serve in democratizing the ways in which news can be generated and disseminated (Cox, 2013). Few go even further, suggesting that online communication environments also provide citizens with unprecedented communicative power. Meanwhile, others remain skeptical and point to different obstacles, including increasing individualization, audience fragmentation, and polarization (Sunstein, 2001; Gentzkow & Shapiro, 2010; Putnam, 2000; Bennett, 1998, 2012; Habermas, 2006) that prevent the formation of a well-functioning public-sphere online. Hence, the major goal of this paper is to bring some clarity into these theoretical discussions by generating some empirical evidence.

(Dis-)respectful public discussions online: insights on audience polarization and formation of radical hate or support groups

Inesa BIRBILAÏTE
METHODOLOGICAL FRAMEWORK AND INTRODUCTION TO THE CASE STUDY

Data for the empirical research was collected from the Facebook page UN Climate Change Conference, 2009 created for the 15th Conference of Parties also known as COP15 meeting. This annual global conference was organized in Copenhagen in 2009 with the major hope to establish a new climate change document, which would replace the Kyoto Protocol. However, the outcomes of the meeting were not as expected and led to public disappointment and dissatisfaction.

This conference, in general, is a good example, demonstrating that traditional democratic ways of policy-making are no longer sufficient in effectively dealing with recent global risks and uncertainties (Birbilaitė, 2013). In particular, no global-binding agreements were reached at the COP15 and in turn, while some countries took independent actions to fight climate change on their own, others still wait. Besides, during the conference, another important trend was noticed – surprisingly intense global public discussions occurring online on climate change.

I assessed the level of respect of each comment posted on the Facebook page UN Climate Change Conference, 2009. The final sample constituted 156 wall-posts published by the page moderator(s) and 2788 comments made by 1424 active participants. Data analysis was performed in two stages. Following them, I provide the results in this paper. First, I discuss descriptive analysis of my data. In addition, I compare results to other similar investigations where the same instrument was used to measure discourse quality in face-to-face discussions. Namely, I selected two PhD studies performed by Ugarriza (2011) and Caluwaerts (2012). While the discussions between ex-combatants of Columbia analysed in the first dissertation by Ugarriza demonstrated a low-level of deliberation, the second study reported rather high-quality discussions between linguistically divided citizens of Belgium. Hence, I aimed to identify where, in comparison with these different studies, my discussions, in terms of respect, stands. Further, I proceed to a more in-depth analysis of empirical data, aiming to identify significant internal and external factors, which might influence variations of quality level.

\[\text{Both studies were performed as a part of a larger research project on deliberation in deeply divided societies, coordinated by Jürg Steiner, professor at the University of Bern (Switzerland) and University of North Carolina (the USA). The project had an explicit comparative aim of determining the favorable conditions for deliberation in deeply divided societies.}\]
HOW TO MEASURE LEVEL OF RESPECT? – A BRIEF INTRODUCTION TO THE DISCOURSE QUALITY INDEX

Discourse quality index (DQI) was the main instrument used to collect and analyse empirical data. Although, this instrument was created to primarily analyse the quality of official face-to-face discussions (e.g., parliamentary debates) my research proves that with minor changes it can easily be adapted to study online content. The core of this index lies in the Habermasian theory of Communication action. In particular, it closely follows six normative discourse conditions discussed by Habermas: participation, respect, justification, common good, force of better argument, and truthfulness. Although some scholars criticize the DQI for being too focused on discourse itself while ignoring the broader context in which discourse takes place (O’Brien, 2009) or for distorting Habermasian ideals because it reduces them to observable phenomena and fails to measure discourse accurately and objectively (King, 2009), Habermas himself applauds the instrument and notices that the DQI captures ‘essential features of proper deliberation’ (Habermas in Bächtiger et al., 2010). Inventors of the DQI agree that coding following the DQI can be subjective and, therefore, requires broader interpretations (Steiner, 2012).

For the purpose of the present paper, I applied the DQI to measure the category of respect in chosen online discussions. Category of respect in the updated version of the DQI is measured in two stages. First, the researcher examines if foul language is used towards participants of the discourse (at a personal level) or towards their arguments. If yes, such contributions are coded under the group of foul language type I (FL I). Contributions that contain foul language at a personal level are considered to be least deliberative. Second, all contributions are assessed for the existence of respectful language towards other participants (at a personal level) and/or their arguments – respectful language type I (RL I).

As noticed by Talpin (2011), for ordinary citizens public expression of disagreement is a difficult move, agreement (arising as respectful language) might be a more favourable way to express positions and also oppositions. Hence, comments containing respectful language towards discourse participants (at a personal level) or towards their comments are perceived as most deliberative, because the participant
expresses his/her position (positive, negative or neutral) in a respectful way without any offensiveness.

In addition to the traditional DQI measurements of the category of respect, in my analysis I also considered respect towards participants who did not directly participate (i.e. politicians, experts, scientists, and others) in the discussions but were important actors of the discourse. I called these indicators foul language type II (FL II) and respectful language type II (RL II). I assumed that when dealing with the quality of online discussions these indicators might be significant, especially considering scholarly literature stressing that online discussions encourage public polarization, formation of hate groups, and might greatly determine low quality of the discourse.

Before turning to the analysis of the empirical data, it is important to ascertain that the coding process was reliable. Since the entire empirical data of this research project was collected and coded by only one coder, I also performed an inter-coder reliability test. Hence, after the main researcher coded the data, the DQI was introduced and explained to four other coders, who were third year bachelor students. One discussion was coded together with students and afterwards they were asked to analyse four separate, randomly selected discussions following the DQI. In total, 192 comments were re-coded. In general, results indicated that the coding process was reliable, as the overall ratio of coding agreement (RCA) was 0.935, which means that coders agreed on 93.5 per cent of the cases. Category of respect demonstrated very high level of inter-coder reliability. Coders agreed on all cases of category RL II, there was a perfect inter-coder agreement (1.000). RCA for RL I was 0.990 and for FL I – 0.974. Yet, a quite low level of reliability was achieved while coding category of FL II (0.828). This can be explained by the fact that sometimes, for additional coders, it was difficult to distinguish between slight foul languages pointed to outside actors or their thoughts, i.e. what was considered to be slight foul language by the main coder to others seemed to be simply strong contra-argument. In other words, the main coder was, in general, stricter. However, Cohen’s kappa indicated that the level of agreement was moderate but acceptable (Landis and Koch, 1977).
RESULTS AND DISCUSSION

Foul and respectful language type I: One can expect that in Web 2.0 based online communication environments, foul language is used more frequently and more freely (Steiner, 2012); however, the level of respect in my sample did not actually differ a lot compared to previous face-to-face studies. Namely, comparing frequencies of foul language in three research works, it was noticed that discussions under consideration were somewhere in between. Specifically, foul language (FL I) in my sample was less frequent when compared to the case of linguistically divided Belgians, but more frequent than reported by Ugarriza (2011) in his research on discussions by ex-combatants of Columbia (see Table 1). On the other hand, comments from my sample more often contained foul language towards individuals at a personal level than towards arguments (1.4 versus 0.3 respectively), which was similar to the Columbian discourse between ex-combatants. This implied a lower level of respect, because in qualitative discussions disagreements based on a personal level cannot be prioritized; instead personal matters should be silent in the name of common good.

In contrast to previous researches, where severe foul language was not recorded at all, I did find some examples of strong foul expressions, especially in those cases where offensive language was pointed towards discourse participants at a personal level. Participants were called idiots, freaks, fools, etc. For instance, in one comment a participant criticized the performance of the UN and named those who supported the UN – “dupes”: “Seriously, does anyone with any intel-

Table 1. Absolute and relative frequency distributions of foul and respectful language towards participants of the discourse.

<table>
<thead>
<tr>
<th>Language Type</th>
<th>Study of discussions between Colombian ex-combatants</th>
<th>Our study</th>
<th>Study of discourse of linguistically divided Belgians</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foul language I</td>
<td>6 (0.6)</td>
<td>40 (1.4)</td>
<td>13 (0.8)</td>
</tr>
<tr>
<td></td>
<td>3 (0.3)</td>
<td>7 (0.3)</td>
<td>55 (3.3)</td>
</tr>
<tr>
<td></td>
<td>1018 (99.1)</td>
<td>2741 (98.3)</td>
<td>1936 (95.5)</td>
</tr>
<tr>
<td></td>
<td>1027 (100)</td>
<td>2788 (100)</td>
<td>1664 (100)</td>
</tr>
<tr>
<td>Respectful language I</td>
<td>8 (0.8)</td>
<td>102 (3.7)</td>
<td>106 (10.2)</td>
</tr>
<tr>
<td></td>
<td>1019 (99.2)</td>
<td>2688 (98.3)</td>
<td>1495 (89.8)</td>
</tr>
<tr>
<td></td>
<td>1027 (100)</td>
<td>2788 (100)</td>
<td>1664 (100)</td>
</tr>
</tbody>
</table>
Foul language was also used towards arguments. Participants were attacked for their positions or opinions. In the following example, the participant was assaulted for his/her comment about the magnitude of US pollution, claiming that the US was responsible for only five per cent of the world’s pollution. The argument was demolished in a disrespectful way: “WTF??!?!!?”. Capital letters, exclamation and question marks emphasized the level of outrage and made this comment even more offensive.

Respectful expressions in the sample in general were more frequent than offensive ones. This might be a good sign, suggesting a higher quality of public discussions. Respectful language in many cases was used in order to support previous arguments and included such phrases as very well said, I fully agree, happy to hear that, brilliant, etc. For instance, to the comment criticizing US president Obama, a participant replied: “Thanks, XXX. We’re on the same page. Obama is doing lots of talking, but not taking any action”.

Results are somehow close to the findings reported by Caluwaerts (2012). The author elaborated that probably participants were looking for more respectful ways to express their disagreements: instead of using foul language they more frequently tended to explicitly and respectfully agree with one group in such a way, demonstrating their opposition and disagreement with other groups and avoiding conflict. On the other hand, the discussions between ex-combatants demonstrated much lower levels of respect. Although participants did not offend each other very often (0.9%), surprisingly they were even less engaged in a respectful manner (0.8%). It is likely that painful events of the past determined a closed manner of discussions between past enemies, where participants did not want to either start a fight nor make friends with the other side. Meanwhile, my example also contributes to the idea that people in general are more likely to agree than disagree. However, in my case, this data might be explained in different terms of online cultures and instead of indicating a higher level of respect, might be an important sign of audience polarization and hate/support group formation online.

Overall, the results in this section were rather surprising. I expected the discourse to be much more offensive because of well known char-
acteristics of online discursive spaces (e.g., anonymity), global origin of the discourse (leading to more diverse attitudes, experiences, and positions), and contradicting origin of the topics under discussion. However, relying on our data, it seems that, probably, direct offensiveness towards participants is not a major problem for a well-functioning public sphere to exist, instead, there are other concerns.

Foul and respectful language type II: Already, in the initial phase of study, I observed that strong, angry, or offensive language was more often pointed towards outside individuals than towards direct participants of the discussions. Later on, this remark was supported by empirical evidence. I counted comments expressing foul language towards outside actors and their ideas and found that in 118 comments (4.2%) participants spoke about outside individuals using foul language and in 118 comments (4.2%) foul language was used towards their ideas or thoughts (see Table 2). Caluwaerts (2012) and Ugarriza (2011) did not measured FL II and RL II.

Politicians, governments, local and global institutions, experts, media, and scientists were the most frequently offended actors. For instance, during the conference, one participant expressed his/her dissatisfaction about the performance of the governments of the world in the Summit. S/he posted: “Stupid governments of the World: <…> Use the money you are wasting on talks and begin building shelter cities, detention camps and food reserves <…>.”

On the other hand, respectful language towards outside individuals or groups was used similarly, compared to respectful language towards discourse participants. In those cases, politicians, media, investors and others were supported in a respectful way. In many examples the president of the US, Barack Obama, was respectfully supported and encouraged for major steps. For instance, one participant replied to the wall-post under the headline “Obama putting 3.4 billion US dollars toward a ‘smart’ power grid”. The participant

| Table 2. Absolute and relative frequency distributions of foul language towards outsiders. |
|-----------------------------------|-----------------|-----------------|-----------------|-----------------|
|                                   | As a personal level | Towards the comments | N (% of the entire sample) |
| Foul language II                  | 118 (4.2)         | 118 (4.2)        | 2552 (41.5)      | 2739 (100)      |
| Respectful language II            | 93 (3.3)          | 2165 (66.7)      | 2738 (100)       |

(Dis-)respectful public discussions online: insights on audience polarization and formation of radical hate or support groups
expressed his/her excitement and support for Obama – “Nice Job Obama!” The exclamation mark indicates that the participant was highly excited. Although rare, I found some examples where investors were also applauded. In the following comment, a participant expressed his/her support for the US investor George Soros who announced that he would invest one billion US dollars in clean energy technology:

“It’s a great investment from Soros for longevity of life on earth. Hope and wish Like XXX, there are stringent watchdogs to oversee the framing of right policies and subsequently their powerful implementation with the strong will of top politicians in every nation to aggressively drive this drive to its logical conclusion”.

Following findings in this section, my assumptions about audience polarization and hate/support group formation might be supported: participants tended to treat each other in a more respectful way which is characteristic for polarized audiences, and at the same time fostered each other’s (positive and negative) emotions about outside actors (by referring to them using foul or respectful language) in turn forming and maintaining radical support or hate groups.

**EXPLAINING VARIATIONS OF THE DELIBERATIVE QUALITY: TIMING AS AN EXPLANATORY FACTOR**

Now that we know the basic characteristics of the sample, we may proceed to more comprehensive explanations of the data and explore what effects different factors might have on the category of respect. In particular, let me explain how the level of respect was influenced by the external factor of time. In light of the general context of the conference, I explain the variation of respect in three time periods: pre-conference (20 April, 2009 – 7 December, 2009 or 223 days), during-conference (8 December, 2009 – 18 December, 2009 or 11 days), and post-conference (19 December, 2009 – 27 September, 2010 or 283 days) (see Table 3).

Data indicated that the majority of the wall-posts were generated before the Summit. That was not surprising because the pre-conference period was long (yet not the longest). It started on 20th April, 2009 and lasted for 226 days. On average, every second day the modera-
tor(s) introduced new wall-posts, which were followed by a total of 1553 comments. Moderator(s) actively stimulated discussions with new wall-posts likely aiming to inform, engage publics, and retain the pace of discussions. The rather low level of the public’s participation might be explained by the fact that it took some time to attract the public’s attention, gather Facebook users around the discussions, and engage them into the discourse before the event started.

The data indicated that foul language towards discourse participants or their arguments (FL I) was used similarly across the three time groups and no significant differences were found. Meanwhile, frequency of foul language towards outside actors and their positions (FL II) was significantly different before, during, and after the conference (p<0.001). Specifically, offensive expressions towards outside-discourse individuals or their positions were significantly more frequent after the event compared to time periods before (p<0.001) and during (p<0.001) the Summit (Table 4). Probably, this can be explained by the fact that after the event, moods changed to the negative side: participants were not satisfied with the Copenhagen Accord, and in turn, disappointment was openly demonstrated and outside actors and their ideas were addressed with foul expressions aiming to express disappointment, distrust, and resentment.

Similarly, respectful language towards discourse participants (RL I) or their arguments did not demonstrate any significant frequency differences in the three time groups. Hence, after having looked at foul and respectful language towards discourse actors, we can be brief - there were indeed no significant differences over the three time periods and, in general, participants tended to be more respectful than disrespectful towards each other.

Table 3. Wall-posts and comments within three time-periods.

<table>
<thead>
<tr>
<th></th>
<th>Pre-conference</th>
<th>Conference</th>
<th>Post-conference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total wall-posts</td>
<td>1283(82.3)</td>
<td>104(6.7)</td>
<td>12(6.6)</td>
</tr>
<tr>
<td>Wall-posts per day (mean)</td>
<td>0.55</td>
<td>1.8</td>
<td>0.04</td>
</tr>
<tr>
<td>Total comments</td>
<td>1553(55.7)</td>
<td>677(24.3)</td>
<td>558(20)</td>
</tr>
<tr>
<td>Comments per day (mean)</td>
<td>6.9</td>
<td>61.5</td>
<td>1.9</td>
</tr>
<tr>
<td>Comments per wall-posts (range)</td>
<td>1-51</td>
<td>21-57</td>
<td>32-65</td>
</tr>
<tr>
<td>Comments per wall-posts (mean)</td>
<td>12.5</td>
<td>33.85</td>
<td>46.5</td>
</tr>
</tbody>
</table>

* Categorical variables were compared using Pearson’s Chi-Squared test. Significant differences were followed by post-hoc analysis using Pearson’s Chi-Squared test or Fisher’s exact test when appropriate. Value of p<0.005 was considered statistically significant, and if p≥0.005 but p<0.01, it was considered that the difference indicated trend.
In the meantime, we also found the frequency of respectful language towards outside individuals or groups (RL II) was significantly different (Table 5). While usage of foul language type II increased over time, respectful language type II – decreased. Specifically, respectful language type II was more often used before the conference, compared to during (p=0.01) and after (p=0.002) the event. The data suggests that before the conference participants were positive about the up-coming event and expected it to succeed. Positive moods were expressed in respectful language towards outside-discourse individuals, addressing them with trust, respect, and encouragements.

Hence, in the sense of respect towards outside actors (RL II), discourse before and during the conference was relatively more respectful compared to the discussions after the event. In other words, it seems that before the event pro-environmental support groups dominated discourse: a majority of participants were united in expressing support for global governments and political leaders, expecting the COP15 to succeed. Yet, sad and disappointing outcomes of the event significantly shifted the direction of the discussions: all the support
participants expressed earlier was changed by dissatisfaction – hate groups changed support groups.

CONCLUDING REMARKS

To summarize the results, it can be said that in the analysed case, self-moderated discussions on Web 2.0 based communication environments are not qualitative enough in terms of respect and they cannot be useful for policy-makers; instead it seems that audiences in these environments are polarized and tend to foster formation of hate and support groups, which encourage dogmatic and narrow thinking without acknowledgement of other possible options; therefore, the main precondition of qualitative deliberation – force of better argument – cannot be exercised. In other words, they do not lead to a consolidation of democratic processes, rather to the destabilization of public sphere and democratic processes in general.

In this paper I did not demonstrate a direct link between the categories of respect and audience polarization online, because I did not analyse the actual structure of discussions and did not consider participants’ characteristics; however, leaning on some theoretical assumptions, this link can be obvious. Specifically, polarized environments tend to support one side and do not involve different positions, which most likely would stimulate confrontations and disrespectful behaviour (unless the discussions are of a very high deliberative level, which is not very likely online), this was a case in my sample. Moreover, polarized environments can be easily transformed to hate or support groups. It was evident in the sample. Although participants were rather respectful to each other, they tended to be more disrespectful towards outside actors. This possibly means that participants supported each other while confronting some other positions articulated in the general COP15 discourse.

The next concluding remark of this paper is related to the variations of level of respect. Namely, support or hate groups are very reactive. They closely follow the general context and reflect the most recent events. Data demonstrated how quickly support groups could be transformed into hate groups.

Another important discovery of this study is that when assessing the level of respect of Web 2.0 based online communication envi-
environments, it is not enough to simply follow DQI and measure only frequency of foul or respectful language towards discourse participants, instead, more importantly, respect towards outside participants should be considered, as it may help to draw some important conclusions about the structure and climate of online discussions.

For the very end I want to highlight that, in general, Web 2.0 communication environments do provide citizens with a new kind of power; however, in order to use it in a good way a number of characteristics of online communication culture have to be recognized and solved.

REFERENCES


