Design Matters!

An empirical analysis of online deliberation on different news platforms

Katharina Esau / Dennis Friess / Christiane Eilders

Institute of Social Sciences at the University of Düsseldorf, Germany

Paper to be presented at the IPP2016: The Platform Society

22-23 September 2016, University of Oxford, United Kingdom

Abstract

Ever since the internet has provided easy access to online debates, advocates of deliberative democracy have hoped for improved debates. This paper investigates which particular platform features are most likely to promote deliberation. We assume that moderation, asynchronous discussion, a well-defined topic, and the availability of information enhance the level of deliberative quality of user comments. A comparison between different kinds of news platforms that differ in terms of design (a news forum, news websites, and Facebook news sites) shows that deliberation (rationality, reciprocity, respect, and constructiveness) differs significantly between platforms. The news forum yields the most rational and respectful debate. While user comments on news websites are only slightly less deliberative, Facebook comments perform poorly in terms of deliberative quality. However, comments left on news websites and on Facebook show particularly high levels of reciprocity among users.

Keywords: platform design, deliberative quality, online deliberation, comparative research

Author Notes

Katharina Esau is Research Fellow at the Department of Communication at the University of Düsseldorf (Germany). Contact: katharina.esau@hhu.de

Dennis Friess is Research Fellow at the Department of Communication at the University of Düsseldorf (Germany). Contact: dennis.friess@hhu.de

Christiane Eilders is Professor for Communication and Media Studies at the Department of Communication at the University of Düsseldorf (Germany). Contact: christiane.eilders@hhu.de
Introduction

The rise of the internet offers various opportunities for political communication. One elaborately discussed form of political communication on the internet is online deliberation. From a theoretical point of view, the internet provides an infrastructure for the public sphere which advocates of deliberative theory have dreamed of (Graham & Witschge, 2003). Not surprisingly, deliberative democracy is one of the most influential theoretical concepts in the ongoing debate on the relationship of democracy and internet technology (Chadwick, 2009). However, due to the widespread popularity of deliberative theories and an increasingly fragmented research landscape, there are many different concepts of deliberation drawing upon a broad range of interpretations of the theoretical literature (Bächtiger & Pedrini, 2010; Dahlberg, 2007). The definitions of deliberation vary in terms of focus, refinement and ambition. Nevertheless, most of the authors share the basic idea that deliberation is a demanding type of communication which fulfills standards like rationality, reciprocity, constructiveness and mutual respect among participants.

Previous research has analyzed deliberation in many different environments, e. g. within parliaments (e. g. Steiner, Bächtiger, Spörndli, & Steenbergen, 2004), deliberative polls (e. g. Fishkin, 2009) or jury systems (e. g. Wolf, 2010). In these environments, deliberation was conceptualized as synchronous face-to-face discussions between a manageable number of participants. With regard to mediated communication (e. g. Gastil, 2008), deliberative quality was analyzed as the degree to which a journalistic text complies with the standards of deliberative communication (e. g. quality and plurality of reasoned arguments), but it did not consider any kind of discourse between participants. Only with the advent of web 2.0 technology, the analysis of mediated deliberation could include the exchange between participants as in face-to-face discussions. In recent years, studies have analyzed online deliberation, e.g. in Usenet newsgroups (e. g. Wilhelm, 1998), government forums (e. g. Coleman, Hall, & Howell, 2002), newspaper websites (e. g. Zhou, Chan, & Peng, 2008) and on social networking sites (e. g. Stroud, Scacco, Muddiman, & Curry, 2015). The variety of this short account of recent communication environments shows that the internet is a network of different communication spaces rather than one monolithic public sphere.

One frequently discussed space for deliberation are news platforms where users comment on journalistic content whether e. g. via news websites or via social networking sites such as Facebook (Strandberg & Berg, 2013; Stroud et al., 2015). While journalistic content
provides occasions and starting points for user deliberation, comments posted on news platforms not always live up to the ideal of deliberation (Coe, Kenski, & Rains, 2014). Therefore, news organizations have adopted different strategies for channeling, filtering and shaping user generated content. While some news outlets have closed comment sections completely (e.g. Popular Science Online), others have outsourced discussions to social networking sites (e.g. Frankfurter Allgemeine Zeitung Online), set up special discussion forums (e.g. Süddeutsche Zeitung Online) or have pioneered new ways of engaging commenters (e.g. Guardian Online). These strategies concern the access to and the architecture of the discourse spaces. Translating the strategies into platform design features, reveals a set of conditions that may promote deliberative styles of discourse. For instance, moderation techniques ensure that comments are respectful and availability of quality information supports reasoned debate. Against this backdrop, we argue that the design of online platforms affects the level of deliberative quality on different online platforms.

In order to investigate this effect, discussions on different platforms have to be compared. However, only few studies have systematically compared deliberation across different online platforms (e.g. Jensen, 2003; Rowe, 2015). This paper aims at filling this research gap by investigating the level of deliberative quality across different online news platforms. We focus on a specific news forum, news websites, and Facebook news sites. News stories on two topics are considered (Refugees Crisis and Military Engagement in Syria). While all platforms have invited users to comment on the topics, they differ in moderation, asynchronicity, availability of information, and level of focus in topic definition. By analyzing deliberation across different online platforms, we link our research to the strand of literature which has empirically analyzed various design features and their effects on the quality of online deliberation (Janssen & Kies, 2005; Towne & Herbsleb, 2012; Wright & Street, 2007). Accordingly, the main research question this paper seeks to answer is:

*How does platform design affect the level of deliberative quality?*

Starting with an overview of existing literature on online deliberation within the context of news content, we discuss the relation between platform design and deliberative quality and present five hypotheses and one further research question. They are tested using data from quantitative content analysis of the level of deliberative quality including 1,801 user comments collected from a news forum, news websites and Facebook. The findings are discussed against the backdrop of normative public sphere theory.
Previous Research

As internet technologies have a high penetration rate in most democratic societies, empirical research on online deliberation has experienced a sharp increase in recent years and an enormous body of theoretical and empirical literature is available now (e.g. Black, Welser, Cosley, & DeGroot, 2011; Davies & Gangadharan, 2009; Davis, 2010; Gerhards & Schafer, 2010). However, the concept itself is rather fuzzy (Delli Carpini, Cook, & Jacobs, 2004). Reflecting the complexity of the deliberation concept and the broadness of the research field, several scholars have introduced different frameworks for a systematic investigation of (online) deliberation (e.g. Bächtiger & Wyss, 2013; Friess & Eilders, 2015; Wessler, 2008). In the most recent framework, Friess and Eilders (2015) distinguish between three elements of deliberation (Figure 1). Based on a review of empirical online deliberation research they argue that the focus has been on (1) the institutional input or institutional design that enables and fosters deliberation, (2) the communicative throughput or process which corresponds to certain quality standards of deliberation, and (3) the productive outcome with the expected results of deliberation. This framework not only reflects the structure of research activity in the field of online deliberation, it may also guide empirical research on the relations between these three elements of deliberation.

Figure 1. Framework for the Analysis of Online Deliberation

This paper focuses on the link between institutional input (1) and communicative throughput (2). It investigates how the design of online platforms influences the level of deliberative
quality in discussions linked to news articles. Therefore, two strands of research are particularly relevant: findings on the level of deliberative quality of user comments and findings on the impact of design on the quality of deliberation.

**Findings on Throughput: Deliberative Quality of User Comments on News**

Nowadays, users not only read journalistic content, but also comment and engage in discussions with other users. Advocates of deliberative democracy hope that online news platforms and other online communities have the potential to provide spaces for democratic debates or even for a new “virtual” public sphere (Papacharissi, 2002). In this vein, online deliberation research asks to what extent political online discussions live up to the norms of deliberation (Habermas, 1984, 1990).

Although there already is a remarkable range of empirical studies on online deliberation, findings are still contradictory. Some studies show that online discussions, while not fully complying with the ideal of deliberation, meet many characteristics of deliberative debate (e. g. Rowe, 2015; Ruiz et al., 2011; Singer, 2009; Strandberg & Berg, 2013). Other studies which are often guided by a more skeptical view of internet technology observe communication which is characterized by incivility and flaming instead of reasoning and respect (Coe et al., 2014) and homophily or polarization instead of rational consensus (e. g. Anderson, Brossard, Scheufele, Xenos, & Ladwig, 2014; Sunstein, 2002; Wilhelm, 2000).

The empirical ambivalence is well illustrated by the findings of Zhou, Chan and Peng (2008). Analyzing news comments within a Chinese newspaper website, they conclude that while a political public sphere in cyberspace is emerging, and quality of discourse has improved, it remains limited in terms of argumentative complexity, and the articulation of disagreement (Zhou et al., 2008). Strandberg and Berg (2013) have also presented mixed findings in their analysis of user comments on Finish news websites. While they found sufficient degrees of respect, findings reveal a low degree of reciprocity and reasoning.

As a consequence of the inconsistent picture of online deliberation reported above, some scholars have begun to explain those differences. For example, Ruiz et al. (2011), taking an international comparative perspective in analyzing user comments on five cross-national

---

1 Among others, these characteristics include topic relevance, reasoning, reciprocity, mutual respect, and constructiveness. However, the empirical operationalization of deliberative quality varies heavily among different studies (Friess & Eilders, 2015). This makes it difficult to compare the findings.
newspaper websites, identified different types of discourses which varied significantly in deliberative quality. *New York Times Online* (US) and *Guardian Online* (UK) showed a greater deal of reasoned, respectful, and reciprocal user comments than *El País Online* (ES), *Le Monde Online* (FR), and *Repubblica Online* (IT). Drawing on Hallin and Mancini (2004), they explain the differences with different national media system traditions. Also concerned with detecting differences in discourse quality, Rowe (2015) has compared user comments on Facebook and news websites. His analysis showed that discussions on *Washington Post Online* where significantly more on-topic, reasoned and reciprocal than discussions on the same news content posted on Facebook. However, his analysis did not explain differences in the light of different platform designs.

Summing up, the above mentioned studies have been sensitive to deliberation on different news platforms, but they did not explicitly link platform design to the level of deliberative quality. As design affects the quality of user contributions (Towne & Herbsleb, 2012), the relation between particular design features and characteristics of deliberative discussions needs to be investigated more closely. This claim ties in with a key belief in literature on deliberative design: it is not about the ability of the internet to sustain democratic debate in general, but a question of the conditions under which deliberation is promoted (Wright & Street, 2007).

**Findings on Input: Deliberative Platform Design**

A growing body of research has identified various social as well as technical factors affecting deliberation (e.g. Coleman & Moss, 2012; Himelboim, Gleave, & Smith, 2009; Janssen & Kies, 2005; Stromer-Galley & Martinson, 2009; Towne & Herbsleb, 2012; Wise, Hamman, & Thorson, 2006; Wright & Street, 2007). A review of the empirical findings helps disclosing particular design features which are likely to have an effect on the level of deliberative quality. Since previous research has considered manifold design features, it is beyond the scope of this paper to discuss all of them. Therefore, we focus on those design features which differ between news platforms and thus play a crucial role in this paper.

Empirical findings suggest that moderation can have positive effects on the quality of deliberation. Analyzing different discussion forums, Wright and Street (2007) conclude that moderation is a crucial design feature to enable respectful online discussions. Similarly, Coleman and Gøtze (2001) state that moderation is important for the success of many-to-many asynchronous dialogue. Stroud et al. (2015) investigated whether the engagement of
journalists affects the quality of user comments beneath news articles. Findings of the quasi-experiment indicate that the engagement of journalists in user dialogs on Facebook influence the deliberative behavior of commenters positively. Against this backdrop, journalists or moderators can act as democratic intermediaries, guiding the debate in a more elaborate way (Edwards, 2002). In the same vein, Noveck (2004, p. 24) argues that “effective facilitation” is the only way to manage competing voices. Other studies suggest that moderation decreases the number of message postings (Rhee & Kim, 2009). On the other hand, Wise and colleagues show that the presence of moderation has a positive effect on the intention to participate in an online community (Wise et al., 2006).

Another important design feature of online platforms concerns the temporal dimension of computer-mediated communication, namely the level of synchronicity or asynchronicity. Janssen and Kies (2005, p. 321) stress that real-time discussions (e. g. chat rooms) as synchronous discussions are more likely to provide small-talk and jokes, while asynchronous discussion spaces which have no time constraints (e. g. forums) are more likely to provide rational-critical debate (for alternative definition of synchronicity see e. g. Jucker & Dürscheid, 2012). Those claims are supported by Stromer-Galley and Martinson (2009) who found that synchronous online chats are problematic for creating quality discourse. They conclude that short messages lead to under-developed arguments, lack of coherence and show a high level of personal attack (Stromer-Galley & Martinson, 2009, p. 197). In the same vein, Strandberg and Berg (2015) provide evidence from an online experiment which suggests that synchronous discussion is a crucial design factor for a deliberative debate.

Since deliberation rests upon the weighing of different arguments and viewpoints, the provision of relevant information is crucial for deliberation. Gudowsky and Bechtold (2013) emphasize the important role that different types of information play in participatory processes. Therefore, the availability of information has also been a key element of the deliberative poll studies conducted by Fishkin and colleagues (e. g. Fishkin & Luskin, 2005; Iyengar, Luskin, & Fishkin, 2005). While information are obviously a source for rational reasons, information could also serve as “discussion catalysts” stimulating deliberation (Himelboim et al.,

---

2 Janssen and Kies (2005) stress the type of moderation. They argue that “the moderator can be a ‘censor’ – for example by removing opinions that are at odds with the main ideology of the discussion space – or he can be ‘promoter of deliberation’ by, for example, implementing a system of synthesis of debate, by giving more visibility to minority opinions, by offering background information related to the topics etc.” (Janssen & Kies, 2005, p. 321).
2009). Studying 20 political online forums, Himelboim, Gleave and Smith found show that 95 percent of the most recognized users posted valid information. Additionally, common information helps to share mental models and fosters coherent communication (Towne & Herbsleb, 2012, p. 104).

The last crucial design feature to be discussed is the clarity of the topic. Noveck (2009, p. 171) points out that the more specific the question, the better targeted the response will be. Reviewing several design principles for deliberative systems, Towne and Herbsleb (2012, p. 102) recommend to divide large tasks into well-defined topics or questions in order to support constructive communication. Since the division of large tasks into small and specified units is one of the key lessons from crowd sourcing projects like Linux or Wikipedia, they argue that this kind of structuration should also be used for deliberative systems in order to generate substantial contributions (Towne & Herbsleb, 2012, p. 103). However, analyzing the effects of journalistic engagement in the comment sections on Facebook, Stroud et al. (2015) could not find significant effects on deliberation from guiding the debate via concrete questions.

Summing up, the design features moderation, asynchronous discussion, availability of information, and well-defined topic have been shown to be particularly influential for the deliberative quality online discussions. The next section will discuss how these features differ between the platforms compared in this paper and how the discussions on these platforms will consequently differ in terms of level of deliberative quality. Before stating the hypotheses, however, it needs to be pointed out that there are additional factors influencing the degree of deliberativeness. Karlsson (2012, p. 65) points out that “online political discussion is mainly shaped not by political institutions, or designers of online platforms or moderators, but by the participants themselves, utilizing forums strategically in relation to their needs and aims.” While various design features could help to foster deliberation, there is no guarantee that they will do so since the context factors, social dynamics and norms (Freelon, 2015) cannot be directly shaped by the designer. As this paper focuses on design features, the additional factors are not considered in the hypotheses.
Hypothesis: Platform Design and Level of Deliberative Quality

Based on findings on deliberative design, we assume that the level of deliberative quality varies among online platforms, each showing different patterns of design. In this paper, we focus on three kinds of news platforms: a news forum, news websites and Facebook news sites. These platforms differ in terms of moderation, asynchronicity, availability of information, and level of focus in topic definition (see Table 1). The news forum under study belongs to the news media organization Süddeutsche Zeitung (SZ) Online. It was designed to sustain reasoned and focused debate and meets most of the above mentioned criteria of deliberative design. In contrast, comment sections on news websites meet many, but not all design criteria. Facebook news sites, lastly, perform worst in meeting the deliberative design criteria. On social networking sites, media organizations may encourage user comments on news, but they have no influence on platform design. Accordingly, there is no moderation, less information, and a weak level of focus in topic definition to enhance deliberation. Moreover, communication on Facebook is less asynchronous compared to the other platforms. The differences between the three platform designs in general are addressed in our first hypothesis:

H1: The highest level of deliberative quality will be found in the news forum, a moderate level of deliberative quality will be found on news websites, and the lowest level of deliberative quality will be found on Facebook.

Table 1. Deliberative Design across News Platforms

<table>
<thead>
<tr>
<th>Design features</th>
<th>News forum</th>
<th>News websites</th>
<th>Facebook</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SZ Online</td>
<td>Welt Online</td>
<td>Spiegel Online</td>
</tr>
<tr>
<td>Moderation (H2)</td>
<td>++</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Asynchronous discussion (H3)</td>
<td>++</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>Availability of information (H4)</td>
<td>++</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Well-defined topic (H5)</td>
<td>++</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: - negative or no effect on deliberative quality; + moderately positive effect; ++ strong positive effect

While our first hypothesis focuses on the level of deliberative quality on different platforms in general, we also assume more specific effects of particular design features on certain characteristics of deliberative quality. On the side of the design features, this concerns moderation, asynchronous discussion, availability of information, and well-defined topic and on the
side of the individual characteristics of deliberative quality, respect, reasoning, constructiveness, and reciprocity are of interest. The specific effects of the different design features will be described in the following.

**Moderation and Respect**

We assume that moderation has a positive impact on deliberative debate, as long as moderators are independent and there are clear rules for participants and moderators. More specifically, it is argued that moderation promotes mutual respect in online interactions. Further, we distinguish between pre- and post-moderation. Post-moderation activities can be found on all three news platforms under study. Therefore, it is not suitable for comparative analysis. Pre-moderation however, is carried out in the news forum and on the news websites, but not on Facebook. In the news forum which has been designed to provide factual, respectful, on-topic debate, we find the strongest form of pre-moderation: supported by automated processes, all comments are reviewed before publication. Assaults and other forms of verbal aggression are removed automatically. In addition, the news forum provides clear guidelines and rules of discussion. The same automated pre-moderation technique is used on the news website *Welt Online*, while the other two news websites, *Spiegel Online* and *Zeit Online*, follow a different strategy which is less systematic and not automated. Due to these differences between the dissimilar types of platforms and between *Welt Online* and the other news websites, we expect different levels of respect in accordance with different moderation techniques:

H2: The highest level of respect will be found in the news forum and on *Welt Online*, a lower level of respect will be found on the news websites *Spiegel Online* and *Zeit Online*, and the lowest level of respect will be found on Facebook.

**Asynchronous Discussion and Reasoning**

Asynchronous discussion constitutes a favorable technical and organizational architecture for rationality and reasoning since it allows participants to spend more time to elaborate their arguments and justify their positions (e. g. Janssen & Kies, 2005). In our comparative design, all three web spaces allow for asynchronous discussion. However, there are significant differences in the temporality of the communication structures of Facebook on the one hand, and news forums or news websites on the other hand. Due to technical infrastructure and social practices, response time on Facebook is shorter while the rate of comments is higher.
which may result in the experience of a quasi-synchronous discussion where many participants contribute posts at the same time. The deliberative quality, especially reasoning, may decrease under these conditions. Hence, we expect a negative influence of a quasi-synchronous discussion on the level of reasoning:

H3: The debate on Facebook will show less reasoning as compared to the news forum and news website discussions.

**Availability of Information and Reasoning**

Another important requirement for reasoning is the availability of quality information (e. g. background information, facts, and stats) (e. g. Gudowsky & Bechtold, 2013; Towne & Herbsleb, 2012). The highest amount of information is provided in the news forum where several articles as well as further material are made available to participants. On news websites, the main source of information is the journalistic article which may be supplemented by hyperlinks to other articles. Journalistic articles and further information provided by the editorial staff are not as readily available on Facebook than on news websites or in the news forum. Based on these differences, we hypothesize that:

H4: The highest level of reasoning will be found in the news forum, a lower level of reasoning will be found on news websites, and the lowest level of reasoning will be found on Facebook.

**Well-defined topic and Constructiveness**

Another characteristic of deliberative quality is constructiveness. This means that discussants try to find solutions to the problem at hand. It is assumed that a well-defined topic has a positive impact on the amount of constructive contributions. While the general topics of the discussion are held constant for all three platforms (*Refugees Crisis* and *Military Engagement in Syria*), there are differences regarding the level of focus in topic definition. The news forum provides a specific question which is intended to initiate a focused debate that might

---

3 While communication on Facebook is per se asynchronous the employed design creates a quasi-synchronous communication environment. In 2011, Facebook started using “Live commenting”, a technology which supports “opportunities for spontaneous online conversations to take place in real time” (see: [https://www.facebook.com/notes/facebook-engineering/live-commenting-behind-the-scenes/496077348919]). User comments are immediately visible for other users and users receive notifications about other users' comments. Since Facebook is also available on mobile devices such as tablets and smartphones and has entered many people daily lives, it seems to be a 24/7 network which may develop pressure to respond more quickly.
even yield solutions for the problem addressed, e. g. “Bundeswehr against IS: rash decision or urgently needed?”. In contrast, news websites do not guide the discussion. However, users may be assumed to read at least parts of the full article, thus they also receive some guidance about issues that need to be discussed and problems that need to be solved. On Facebook, there is only a short teaser (summary, title and picture) which is linked to the full article. The very brief presentation of the topic in the teaser may trigger general opinion expression without knowledge of the journalist’s arguments rather than contributions with specific solutions. Due to these differences in the level of focus in topic definition, we hypothesize:

H5: The highest level of constructiveness will be found in the news forum, a lower level of constructiveness will be found on news websites, and the lowest level of constructiveness will be found on Facebook.

**Reciprocity as consequence of platform design**

Theorists of deliberative democracy argue that deliberation is a social process of giving and taking, which includes both listening and responding (Barber, 1984, p. 175). Therefore, deliberation is a reciprocal process: arguments should not just be articulated, but rather also listened and replied to. In the context of online deliberation, it is crucial to capture whether (general engagement) and how (e. g. critical, argumentative engagement) a comment addresses another comment. However, as sufficient empirical findings on design features affecting reciprocity are missing, we ask how the level of reciprocity varies across the three platforms.

RQ1: How does reciprocity vary across the news platforms?

**Methodology**

This study assesses the level of deliberative quality among platforms with different design patterns. Therefore, we conducted a content analysis of comments left a) in a news forum, b) on news websites, and c) on Facebook news sites concerning the same journalistic content on two topics, (Refugees Crisis and Military Engagement in Syria), both of them controversially discussed in December 2015. The comments were collected from topic-related articles which addressed a specific problem, included conflict and required decision, characteristics assumed to be preconditions for deliberative discussion (Gutmann & Thompson, 2004). The following sections describe the sampling process and the operationalization of deliberative quality of user comments.
Sample

A sample of news articles published in December 2015 with the related user comments was drawn from the online platforms of four German news media: *Süddeutsche Zeitung Online*⁴, *Spiegel Online*⁵, *Welt Online*⁶, *Zeit Online*⁷. The selected news media are national elite newspapers and a news magazine considered as opinion leaders in the national media system (Jarren & Vogel, 2011). Their online offshoots are listed among the most popular German online resources (AGOF, 2015)⁸.

The first step of the sampling process consisted of a total of 18 news articles⁹ from which 3,341 comments were collected and entered into a database where they were numbered chronologically. Each comment was also given a number to signify from which platform and which news article it was taken. In the second step for each article up to 100 sequential comments were randomly selected for content analysis adding up a sample of 1,801 comments (979 on Facebook, 591 on news websites, 231 in the news forum).

A first descriptive analysis shows that on average per article, 212 comments were written via Facebook, 201 via news website and 77 via news forum. The differences exhibit that different platforms generate different amounts of comments even though comments related to the same journalistic content.

**Conceptualization and Coding scheme**

For the purpose of this study, we considered seven measures of deliberative debate grouped in four dimensions in order to assess the deliberative quality on the different platforms (Table 2). The coding scheme is based on Habermas' (1990) discourse ethics and on

---

⁴ *SZ Online*: comments via news forum (www.sueddeutsche.de/thema/Ihr_Forum) and Facebook (www.facebook.com/ihr.sz).
⁵ *Spiegel Online*: comments via news website (www.spiegel.de) and Facebook (www.facebook.com/spiegelonline).
⁶ *Welt Online*: comments via news website (www.welt.de) and Facebook (www.facebook.com/welt).
⁷ *Zeit Online*: comments via news website (www.zeit.de) and Facebook (www.facebook.com/zeitonline).
⁸ Although these criteria apply to other newspapers as well, e. g. *Frankfurter Allgemeine Zeitung Online*, comparative analysis was not possible due to closed comment sections on the news websites.
⁹ 3 articles via news forum (SZ Online), 6 articles via news website (Spiegel Online, Welt Online, Zeit Online), and 9 articles via Facebook (SZ Online, Spiegel Online, Welt Online, Zeit Online).
recent empirical studies on online deliberation (e. g. Black et al., 2011; Steiner et al., 2004; Stromer-Galley, 2007; Trénel, 2004).

Table 2. List of Dimensions, Measures and Measure Definitions

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Measure</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rationality</td>
<td>Topic relevance</td>
<td>This measure captured if a comment hit the topic of the discussion space.</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td>This measure captures whether a comment presents at least one reasoned argument (justification of a statement).</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>General engagement</td>
<td>This measure captures whether a comment addresses another comment.</td>
</tr>
<tr>
<td></td>
<td>Argumentative engagement</td>
<td>This measure captured whether a comment is addressing a specific argument claimed by another comment.</td>
</tr>
<tr>
<td></td>
<td>Critical engagement</td>
<td>This measure captured if a comment is critical towards another comment.</td>
</tr>
<tr>
<td>Respect</td>
<td>Respectful communication</td>
<td>This measure captures whether users interact respectfully with each other. Respectful communication is therefore defined as the absence of aggressive and offensive language.</td>
</tr>
<tr>
<td>Constructiveness</td>
<td>Constructive contribution</td>
<td>This measure captures whether a comment contains constructive elements such as proposals of solutions.</td>
</tr>
</tbody>
</table>

Inter-Coder Reliability

The sample of comments was analyzed by a team of twelve coders. After several training sessions and revisions of the coding scheme, the coders analyzed a sub-sample of 40 comments randomly selected from comments on all platforms. Two indicators of inter-coder agreement were applied: the ratio of coding agreement (RCA) (Holsti, 1969) and Cohen’s Kappa (k) (Cohen, 1960) were appropriate (Table 3). Reliability is in line with other content analyses of deliberative quality (Rowe, 2015; Steenbergen, Bächtiger, Spörndli, & Steiner, 2003). In total, coders made 440 judgments and agreed in 83.5 % of the cases, which is a satisfactory result.

10 Due to low variance, Cohen’s k could not be computed for all variables.
Table 3. Reliability Scores by Coding Category (N=40)

<table>
<thead>
<tr>
<th>Measure</th>
<th>RCA</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic relevance</td>
<td>0.99</td>
<td>-</td>
</tr>
<tr>
<td>Reasoning</td>
<td>0.64</td>
<td>0.347</td>
</tr>
<tr>
<td>General engagement</td>
<td>0.92</td>
<td>-</td>
</tr>
<tr>
<td>Argumentative engagement</td>
<td>0.77</td>
<td>0.542</td>
</tr>
<tr>
<td>Critical engagement</td>
<td>0.89</td>
<td>-</td>
</tr>
<tr>
<td>Respect</td>
<td>0.90</td>
<td>0.82</td>
</tr>
<tr>
<td>Constructiveness</td>
<td>0.86</td>
<td>0.771</td>
</tr>
</tbody>
</table>

Results

Altogether, the results coincide with the previous research on online deliberation related to news (Graham & Wright, 2015; Rowe, 2015; Ruiz et al., 2011; Singer, 2009; Strandberg & Berg, 2013). Overall, results show that the comment sections under study comply with many of the characteristics of deliberative discussions (see Table 4). With regard to rationality, comments are mostly on-topic (topic relevance 82%), and almost half of them provide justified statements (reasoning 46%). With regard to reciprocity, two thirds showed explicit or implicit references to other users (general engagement 68%), one out of three comments showed critical engagement with other users and just under a third showed argumentative engagement with others (critical engagement 39% and argumentative engagement 30%). Surprisingly, despite of the very controversial discussion topics, most comments were respectful without personal attacks or other forms of verbal aggression towards other users (respect 90%). Finally, a small number of 102 comments tried to provide a solution to the problem (constructiveness 6%).

Table 4. Deliberative Quality in User Comments, all Platforms (N=1801)

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic relevance</td>
<td>1472</td>
<td>81.7</td>
</tr>
<tr>
<td>Reasoning</td>
<td>822</td>
<td>45.6</td>
</tr>
<tr>
<td>General engagement</td>
<td>1223</td>
<td>67.9</td>
</tr>
<tr>
<td>Critical engagement</td>
<td>705</td>
<td>39.1</td>
</tr>
<tr>
<td>Argumentative engagement</td>
<td>548</td>
<td>30.4</td>
</tr>
<tr>
<td>Respect</td>
<td>1621</td>
<td>90.0</td>
</tr>
<tr>
<td>Constructiveness</td>
<td>102</td>
<td>5.7</td>
</tr>
</tbody>
</table>
In line with the differences in platform design, we assumed that the highest level of deliberative quality would be found in the news forum, a moderate level of deliberative quality would be found on news websites, and the lowest level of deliberative quality would be found on Facebook (H1). In order to test this general hypothesis we computed an additive deliberation index\(^\text{11}\) including all of the seven above mentioned characteristics. We found that news forum comments show the highest level of deliberative quality \((M = 4.02, SD = 1.40)\), user comments on newspaper websites show a slightly lower level of deliberative quality \((M = 3.94, SD = 1.42)\) and Facebook comments show the lowest level of deliberative quality \((M = 3.30, SD = 1.40)\). The differences between the news forum and Facebook \((t (1208) = 6.96, p < .001)\) and between news websites and Facebook \((t (1568) = 8.69, p < .001)\) are significant. Therefore, the first part of H1 is confirmed. However, the results show that the differences between news forum and news websites are not significant \((t (820) = .68, n. s.)\). Therefore, the second part of H1 is rejected: Deliberative quality is lower for Facebook than for the other platforms, but not significantly different between the news forum and news websites.

With regard to the specific effects of particular design features on particular characteristics of deliberative quality, we assumed an influence of moderation on the level of respect (H2). The results show no significant differences between the news forum (98\%) and the news website \textit{Welt Online} (97\%) as both platforms used the same pre-moderation technique (see Table 5). Further, discussions on the news websites \textit{Spiegel Online} and \textit{Zeit Online} (both at 87\%) were shown to be significantly less respectful compared to the news forum and the news website \textit{Welt Online} \((\chi^2 = 38.410, df. 1, p < .001)\). Additionally, an even lower level of respect (84\%) than on the other platforms was found on Facebook \((\chi^2 = 32.466, df. 2, p < .001)\) (see Table 6). Since the ranks reflect the assumptions, H2 is supported.

\textbf{Table 5. Level of Respect on News Forum vs. News Websites (N=822)}

<table>
<thead>
<tr>
<th>Measure</th>
<th>News Forum</th>
<th>News Websites</th>
<th>(\chi^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=231</td>
<td>Welt Online N=294</td>
<td>Spiegel Online N=101</td>
</tr>
<tr>
<td>Respect</td>
<td>227 (98.3)</td>
<td>286 (97.3)</td>
<td>n. s.</td>
</tr>
<tr>
<td></td>
<td>227 (98.3)</td>
<td>88 (87.1)</td>
<td>171 (87.2)</td>
</tr>
</tbody>
</table>

\(^{11}\) Although the internal consistency of the index was relatively low (Cronbach's alpha = 0.50), we use this measure of deliberative quality to reduce a very complex phenomenon for comparative analysis (Bächtiger et al., 2009).
H3 and H4 addressed effects of design on reasoning. H3 dealt with the effect of asynchronicity. The quasi-synchronous discussions on Facebook were assumed to result in lower levels of reasoning than on the two other platforms. Table 5 shows that whereas a majority of comments in the news forum (72%) and news websites (56%) address at least one argument, only a third of the Facebook comments (33%) provide arguments ($\chi^2 = 129.524$, df. 1). However, the differences between news forum and news websites are also significant ($\chi^2 = 18.179$, df. 1, $p < 0.01$) while H3 had assumed reasoning in news forum and news websites to be similar due to the same level of asynchronicity. Consequently, H3 has to be rejected because we could not show that asynchronicity leads to the expected pattern.

Table 6. Deliberative Quality across News Platforms ($N=1801$)

<table>
<thead>
<tr>
<th>Measure</th>
<th>News Forum N=231</th>
<th>News Website N=591</th>
<th>Facebook N=979</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic relevance</td>
<td>222 (96.1)</td>
<td>494 (83.6)</td>
<td>756 (77.2)</td>
<td>46.632 ($p &lt; .001$)</td>
</tr>
<tr>
<td>Reasoning</td>
<td>166 (71.9)</td>
<td>329 (55.7)</td>
<td>327 (33.4)</td>
<td>146.996 ($p &lt; .001$)</td>
</tr>
<tr>
<td>General engagement</td>
<td>125 (54.1)</td>
<td>450 (76.1)</td>
<td>648 (66.2)</td>
<td>39.862 ($p &lt; .001$)</td>
</tr>
<tr>
<td>Critical engagement</td>
<td>87 (37.7)</td>
<td>259 (43.8)</td>
<td>359 (36.7)</td>
<td>8.158 ($p &lt; .05$)</td>
</tr>
<tr>
<td>Arg. engagement</td>
<td>75 (32.5)</td>
<td>229 (38.7)</td>
<td>244 (24.9)</td>
<td>33.773 ($p &lt; .001$)</td>
</tr>
<tr>
<td>Respect</td>
<td>227 (98.3)</td>
<td>545 (92.2)</td>
<td>825 (84.3)</td>
<td>32.466 ($p &lt; .001$)</td>
</tr>
<tr>
<td>Constructiveness</td>
<td>26 (11.3)</td>
<td>24 (4.1)</td>
<td>52 (5.3)</td>
<td>16.578 ($p &lt; .001$)</td>
</tr>
</tbody>
</table>

Note: Frequency percentages for each platform in parentheses.

Since the difference in level of reasoning between news forum and news website cannot be explained by differences in asynchronicity, availability of information was assumed to account for the differences (H4). This design factor not only distinguishes the news forum from the news websites, but it is also known to foster reasoning. Testing H3 against H4 clarifies which factor explains the level of reasoning more precisely.

In H4 we predicted that due to the differences in amount of available information, the highest level of reasoning will be found in the news forum, a moderate level of reasoning will be found on news websites, and the lowest level of reasoning will be found on Facebook. A critical test to determine whether levels of reasoning vary with asynchronicity or with availability of information, relates to the question whether or not news forum and news websites show different levels of reasoning. If there are no differences between news forum and news
websites, asynchronicity is likely to predict reasoning because asynchronicity does not differ between the two platforms. If, however, news forum and news websites differ, it is the availability of information that accounts for the differences because the platforms have different levels in information availability. As the differences between platforms showed the ranking assumed for the effect of availability of information (Table 5), H4 was accepted.

Finally, H5 predicted that differences in the level of topic definition will lead to different levels of constructiveness among news platforms. The results show that news forum comments were most likely to provide constructive comments (11.3%). However, contrary to the expectations, we found more constructiveness on Facebook (5.3%) than on news websites (4.1%) ($\chi^2 = 16.578$, df. 2, p < 0.01). Due to the ranks of the platforms concerning constructiveness, H5 is rejected. Although the highest level of constructiveness was found in the news forum where the level of topic definition was highest, it cannot be said that the level of topic definition leads to different levels of constructiveness.

The last analysis deals with the research question concerning the effects of design features on reciprocity. The levels of reciprocity on the three platforms show a pattern that is different from the pattern associated with the other characteristics of deliberative quality. All three indicators of reciprocity are most frequently found on the news websites with 76% of comments addressing other comments (general engagement), 44% criticizing other comments (critical engagement), and 39% addressing a specific argument claimed in another comment (argumentative engagement) (Table 5). Facebook users are only slightly less active than website users in terms of general engagement (66%), but show lower levels of critical (37%) or argumentative engagement (25%) in their interaction. In contrast, forum users less often interact in terms of general engagement (54%), but show higher levels of critical or argumentative engagement in their interaction (38% and 33%).

Discussion

This study started from the assumption that deliberative quality depends on platform design (H1). The empirical findings of our comparative analysis across three types of news platforms support this assumption. Deliberation is most likely to be found in the news forum particularly designed to initiate user discussions. News websites complied less with deliberative design criteria and showed a lower level of deliberative quality. Facebook was last in meeting deliberative design criteria and performed poorer than the other platforms in sustaining of deliberative debates. Since Facebook and news websites did not differ significantly,
however, H1 has been partly rejected. We had further assumed a relation between particular design features and particular characteristics of deliberative quality. Findings support our assumptions on the influence of moderation, availability of information and level of focus in topic definition: Moderation had a positive effect on respect (H2), the availability of information increased the level of reasoning (H4), and a well-defined topic in the news forum resulted in more constructiveness (H5). However, there was no evidence for a positive influence of asynchronous discussion on the provision of reasons (H3).

With regard to our research question on reciprocity among users, findings contradicted the pattern found for most other elements of deliberative quality. While Facebook performed poorly concerning the overall level of deliberative quality as well as the above mentioned characteristics of deliberative quality, it proved to promote a high degree of general engagement among users. Conversely, the news forum platform intended to initiate deliberative debates, showed the lowest scores on this characteristic of deliberative quality. Interestingly, users interacted more without a specified question and without additional information provided than under conditions complying with the ideal deliberative design characteristics. As reciprocity is a key affordance of deliberative debate, this finding deserves more attention in future research. Experimental studies testing variations of design features might clarify under which conditions users connect with each other in discussing an issue. However, platform design is not the only factor influencing the degree of deliberative quality. In order to deepen our understanding of the conditions of deliberative debate, we need surveys and in-depth interviews aiming at motivations and other audience characteristics (e.g. Springer et al. (2015).

It goes without saying, that one study cannot answer all the questions in the young tradition of online deliberation. This study addressed effects of platform design. Yet, within the scope of this research objective, several limitations have to be mentioned. Firstly, this study focused on the few design features that differed between the platforms under scrutiny while neglecting other influence factors. Deliberative quality is also known to depend on identification of users (Coleman & Moss, 2012, p. 8), group size (Himelboim, 2008), group heterogeneity (Zhang, Cao, & Tran, 2013) and response rate (Wise et al., 2006).

Limitations also concern the operationalization of concepts. Unlike other more sophisticated operationalizations of reasoning (e.g. Steenbergen et al., 2003) our dichotomous measure simply captures whether a comment presents at least one justification of a statement. As online comments are short and hardly provide extensive justifications (compared to e.g.
essays or parliamentary debates), we did not analyze the argumentative quality in detail. We therefore might overestimate the degree of reasoning in the debates under study. Furthermore, our results support previous studies (Bächtiger et al., 2009), which showed that deliberative quality is a multi-dimensional concept. Drawing on this, the analysis was based on a deliberative quality index that although low in internal consistency, was helpful in reducing the complex phenomenon of deliberation for comparative analysis.

This said, it is clear that further research has to close the gaps mentioned above and to develop better instruments for assessing deliberative quality. Research on online deliberation has only started to become a discernable tradition. This study sheds light on a very small segment of online deliberation. Nonetheless, it yields some relevant findings. Although both the European refugee crisis und the German military operation in Syria are highly politicized and contested topics, the quality of the debate on all three platforms was mostly on topic, fairly reasonable and reciprocal compared to other studies (Graham & Wright, 2015; Rowe, 2015; Ruiz et al., 2011; Singer, 2009; Strandberg & Berg, 2013). With regard to enhancing deliberative discussions for large segments of the population, the results support the claim that careful design considerations improve the deliberative quality of online discussions (e. g. Wright & Street, 2007).

The findings of this study suggest that deliberative discourse in the virtual public sphere of the internet is indeed possible. This is good news for advocates of deliberative theory. However, findings also suggest that public discourse has to be organized by carefully considering the design of platforms where discourse is taking place. This has implications for journalism practice, namely for designing news platforms. Established media organizations are key actors in this process. They can hence shape the debates and foster deliberative quality by providing ideal conditions. In the light of these findings, outsourcing discussions to social networking sites such as Facebook is not advisable due to the low level of deliberative quality compared to other news platforms. In consequence, it requires significant efforts and resources for news media organizations. However, since this study presented evidence that design matters, advocates of an utterly free virtual public sphere may be disappointed. It becomes clear that deliberation is more likely to emerge if the design is adapted to particular criteria. This contradicts the basic idea of freely open debate among equals where the only force in place is “the forceless force of the better argument” (Habermas, 1975, p. 108).
References


Zhang, W., Cao, X., & Tran, M. N. (2013). The structural features and the deliberative quality of online discussions. *Telematics and Informatics, 30*(2), 74–86.