



Discourse integration in positional online news reader comments: Patterns of responsiveness across types of democracy, digital platforms, and perspective camps

new media & society
2024, Vol. 26(11) 6796–6814
© The Author(s) 2023



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/14614448231183704
journals.sagepub.com/home/nms



Julia Jakob , Chung-hong Chan,
Timo Dobbrick and Hartmut Wessler

University of Mannheim, Germany

Abstract

Online discourse integration, or the degree to which online user comments are responsive, that is, address or refer to other debate participants, is a normatively valued yet neglected quality dimension of online discussions. This preregistered study features the first cross-country/cross-platform investigation of online discourse integration, using manual and computational content analysis ($N=9835$ and $N=30,753$ positional news reader comments). Unexpectedly, about one quarter of the comments was responsive in both majoritarian and consensus-oriented democracies (Australia/United States vs Germany/Switzerland) and on platforms that separate or mix public and private contexts (websites vs Facebook pages of mainstream media), even though other deliberative quality criteria were previously shown to vary by country and platform. Comments that are responsive to fellow commenters in the opposing perspective camp were more likely to contain negative evaluations of those addressed, whereas comments responsive within the same perspective camp were more likely to contain positive evaluations.

Keywords

Consensus democracy, context collapse, cross-national comparison, majoritarian democracy, online discussion, reciprocity, responsiveness, socio-technical affordances, user comments

Corresponding author:

Julia Jakob, Mannheim Centre for European Social Research, University of Mannheim, A5, 6 Building A, 68159 Mannheim, Germany.

Email: julia.jakob@uni-mannheim.de

In deliberative theory, reciprocity is one of the most central normative requirements for high-quality public discussions (Pedrini et al., 2013). It refers to the act of “listening or reading what others say and responding to it” (Esau and Friess, 2022: 1). Such a mutual exchange is considered key to productive public discussions because it encourages individuals to reflect on and engage with the reasons and positions of fellow debaters (Graham and Witschge, 2003). Thereby, reciprocity can increase individuals’ issue knowledge and make them more tolerant of diverging views, thus supporting the epistemic function of public deliberation (Esau and Friess, 2022). Moreover, reciprocity is linked to the deliberative norms of respect and inclusion because the key to reciprocity is to take individuals and their opinions seriously (Morrell, 2018; Pedrini et al., 2013; Scudder, 2020).

Online deliberation research assesses the quality of online discussions against deliberative norms. However, despite its theoretical importance, our empirical understanding of reciprocity in online discussions is limited because most online deliberation research neglects the relational structure of online debates (Esau and Friess, 2022). This preregistered study¹ contributes to filling this gap. We theoretically conceptualize and empirically measure *responsiveness* as one key aspect of reciprocity in online discussions. Responsiveness gauges whether online commenters address or refer to other debate participants in their posts, and thus captures the procedural component of deliberative reciprocity. While responsiveness is measured in the individual comments, at the aggregate level of the discussions, we refer to the degree to which online user comments are responsive as *online discourse integration*.

Empirically, our study has two starting points: *First*, existing research on the relational structure of online discussions lacks a systematically comparative approach. Studies of individual forums in individual countries show that between 30% and 70% of the comments in online discussions can be responsive (Camaj, 2021; Collins and Nerlich, 2015; Strandberg and Berg, 2013; Ziegele et al., 2020). But while this spread may partly be caused by different operationalizations, it is unclear which factors explain the remaining differences (Esau and Friess, 2022). *Second*, existing research disregards the camp-related dynamics in online discussions. Studies show that (online) public discussions typically take place between several larger *perspective camps*, that is, between coalitions of debaters who share the same general perspective on the discussed issue (Bodrunova et al., 2019; Wessler et al., 2008). Polarization studies suggest that exchanges between individuals from different perspective camps are rare in online discussions—and that such cross-cutting exchanges are dominated by negativity (Marchal, 2022; Yarchi et al., 2021). However, online deliberation research has hitherto not examined whether responsiveness occurs mainly across or within different perspective camps and which sentiments dominate the respective encounters.

In this study, we operationalize online discussions by studying “positional” user comments. Positional user comments contain a clear perspective on the issue under discussion, which means that they reveal opinions or judgments that support a distinct position on this issue, such as for example, a pro or anti stance on Muslim immigration. Specifically, we analyze a set of online news reader comments that contain two diametrically opposed perspectives on a contested issue, as will be specified in the “Methodology” section. Positional comments are a hard case for discourse integration because they are

comparatively unlikely to be responsive. Heatherly et al. (2017), for example, show that individuals who hold strong partisan positions are less likely to participate in cross-cutting online discussions. In a comparative analysis using manual and computational content analysis, we examine how the *type of democracy* in a country (Lijphart, 2012) and the *degree of context collapse* afforded by a discussion platform (boyd, 2011) shape online discourse integration, that is, the level of responsiveness in these positional comments. We investigate comments from two consensus-oriented democracies (Germany and Switzerland) and from two majoritarian democracies (Australia and the United States). In each country, we analyze comments from a platform that separates public and private contexts (the website comment sections of mainstream news media) and from a platform that blurs public and private contexts (the Facebook pages of mainstream news media). Independent from the national and platform-specific context, we study whether online news commenters engage with fellow debaters from their own or from the opposing perspective camp and explore the *valence of such responsive engagements*. The comments we analyze were posted below online news articles that thematize how Western societies should shape their relationship to Muslims and Islam, published between August 2015 and July 2016. The topic was highly salient and equally relevant in all four countries during this period, which enables the systematic comparison.

Literature review

Theoretical concepts: reciprocity, responsiveness, and discourse integration

Of different democratic theories, deliberative theory puts the greatest emphasis on reciprocity in public discussions (Gutmann and Thompson, 2002). Since using the most demanding theoretical standard to assess empirical realities increases the potential “extent of improvement brought about by following the respective recommendations” (Wessler et al., 2022: 378), deliberative theory is a suitable basis for this study.

Public deliberation describes the practice of *mutual* justification among individuals with different positions on public issues (Gutmann and Thompson, 2002). Thereby, reciprocity refers to the “taking in . . . of another’s claim or reason and giving a response to that claim” (Graham and Witschge, 2003: 176). Thus, reciprocity as conceptualized in deliberative theory has both a substantive and a procedural component (Gutmann and Thompson, 2002). The procedural component refers to individuals being *responsive* to each other, which means that they address or refer to another debate participant (Friess et al., 2021). The substantive component of reciprocity involves genuine listening and reflection (Dobson, 2012; Morrell, 2018). It requires a *fair consideration* of the reasons and perspectives of fellow citizens, which involves “really considering what others have to say” (Scudder, 2020: 505). Esau and Friess (2022) emphasize the core of the procedural and substantive component of reciprocity by distinguishing between *simple replying* and *deliberative reciprocity*: While simple replying involves responsiveness between individuals, deliberative reciprocity necessitates replies that are coherent, respectful, and reasoned.

This study focuses on responsiveness as the procedural component of reciprocity because this enables us to study the discussions in our dataset in their entirety. While the substantive component of reciprocity is hard to capture, measures of responsiveness

online are well established (Friess et al., 2021) and can be used to develop well-performing computational models for large-scale content analyses. We argue that studying online discussions comprehensively is important with respect to the structural nature of these debates because it sheds light on how integrated these discourses are as a whole. Therefore, while we measure responsiveness in the individual comments, at the aggregate level of the discussions, we refer to the degree to which online user comments are responsive as *online discourse integration*. The term emphasizes our focus on the overall structure of online debates rather than on the responsive relationship between individual commenters. The limitations of our procedural approach will be considered in the “Discussion” section.

Empirical context: what shapes responsiveness in online discussions?

While there is still “a lack of research explaining the different levels of reciprocity [in online discussions]” (Esau and Friess, 2022: 2), previous studies identified several factors that influence this dimension of online debate quality. The level of responsiveness varies, for example, across the comment sections of different media outlets (Ruiz et al., 2011; Ziegele et al., 2020). Furthermore, news factors of the articles commented on and the deliberative quality of previous comments can affect responsiveness: The level of responsiveness in news reader comments is higher, for example, when the related news article covers issues with a high societal impact or with negative consequences, and when the covered issue relates to personal experiences of the commenters (Ziegele et al., 2020). User comments that contain arguments or constructive elements like problem solutions, that ask for additional information from others, or that express a critical attitude on the discussed issue are more likely to receive replies (Esau and Friess, 2022). Since there is little comparative research on responsiveness in online discussions, we aim to contribute to these insights through a systematic cross-country–cross-platform analysis. From theoretical works and previous research, we identified three factors that seem particularly consequential in shaping responsiveness: The type of democracy, the degree of context collapse afforded by an online platform, and different perspective camps.

Type of democracy. Cross-national research on online discussions is scarce (Jakob et al., 2023a, 2023b). Therefore, responsiveness in user comments has hitherto not been compared systematically across countries, let alone across different types of democracy. Still, macro-level differences in online discourse integration can be inferred from theoretical works, from comparative research on parliamentary debates, and from cross-national studies on other aspects of online debate quality such as argumentation and civility. Lijphart (2012) distinguishes two prototypical types of democracy, namely, consensus-oriented and majoritarian democracy. The core difference between these two types is the degree of power sharing in the political system: While majoritarian democracies are governed by the dominant party, power is shared in multiparty coalitions in consensus-oriented democracies. Thus, while majoritarian democracies are inherently adversarial, consensus-oriented democracies require high levels of compromise and political dialogue. Therefore, political discourse tends to be more responsive and accommodating in consensus-oriented than in majoritarian democracies, in which political actors must not

necessarily talk with each other and an “us-versus-them” logic dominates political debates (McCoy and Somer, 2019; Steiner et al., 2004). We argue that this pattern could also emerge in online news reader discussions because citizens adopt the debate styles and norms that their political leadership exemplifies. Research shows that citizens are strongly influenced by political elite cues (Stapleton and Dawkins, 2022) and initial studies indicate that a transformation of discussion norms from politicians to the electorate is indeed taking place. Jakob et al. (2023b), for example, show that the argumentative quality of online discussions is significantly higher in consensus-oriented than in majoritarian democracies—a pattern that had previously also been found in parliamentary debates (Steiner et al., 2004). Similarly, both parliamentary and online discussions were shown to be more respectful in consensus-oriented than in majoritarian democracies (Jakob et al., 2023a; Steiner et al., 2004). This suggests that the “discussion norms of different political systems indeed transmit to citizens from political elite interactions through observational learning” (Jakob et al., 2023b: 594). Since this seems to be the case for the deliberative norms of argumentation and civility, it should also be true for responsiveness. Therefore,

H1. User-generated debates (i.e. positional comments) are more integrated (i.e. responsive) in consensus-oriented than in majoritarian democracies.

In our empirical analysis, consensus-oriented democracies are represented by Germany and Switzerland and majoritarian democracies are represented by the United States and Australia (see Supplementary Appendix I on The Open Science Framework (OSF) for country selection details).

Context collapse. At the platform level, we build on technical affordance theories (Nagy and Neff, 2015) and assume that the degree of context collapse afforded by an online platform may explain differences in responsiveness across forums. Deliberative theorists decouple problem-oriented public debates from private, essentially social conversations (Schudson, 1997). In digital spaces, however, public and private spheres are increasingly blurring (boyd, 2011). In many online discussion arenas, an individual’s diverse networks—including friends, professional contacts, and other acquaintances—pool into one large audience together with the public (Vitak, 2012). While this is also shaped by individual media use habits, more broadly, this context collapse can be considered a structural platform-related affordance that distinguishes the website comment sections from the public Facebook pages of news media outlets, both of which we study empirically: News website comment sections typically connect strangers brought together by their interest in an article (Esau et al., 2017). They network opinions rather than people and embody the notion of publicness. In contrast, when commenting on news articles on the Facebook pages of media organizations, individuals must expect that their contributions will also be displayed to their private friend network by the Facebook algorithm (Hughes et al., 2012)—even though their comments may be primarily addressed at the strangers who assemble on the Facebook news pages. Therefore, according to Rowe (2015), the collapse of public and private contexts is comparatively stronger on the public Facebook pages than in the website comment sections of news media. Studies show that while audiences (Nelson and Webster, 2017) and commenters (Rowe, 2015) on news websites are rather diverse in their

political views, individuals tend to befriend those with similar political opinions and orientations, personal values, and personality traits on Facebook (Cargnino and Neubaum, 2021; Lönnqvist and Itkonen, 2015). Other studies demonstrate that online discussions are comparatively more responsive “in forums where the level of consensus among participants is low” (Karlsson, 2012: 64) and particularly little responsive among like-minded people (Freelon, 2015; Valera-Ordaz, 2019). Combined, this evidence suggests that online discussions could be more responsive under separated contexts, for example, because individuals feel a greater need to defend their views among strangers than within a network that includes like-minded friends. In an initial study building on the explanatory factor of collapsed contexts, Rowe (2015) shows that online news reader comments are indeed significantly more responsive when public and private contexts are separated, comparing comments from the website and the public Facebook page of the *Washington Post*. To further substantiate his finding, we test this hypothesis on a set of comments from a larger pool of media outlets:

H2. User-generated debates (i.e. positional comments) are more integrated (i.e. responsive) in arenas that separate public and private contexts more clearly than in those that mix the two.

Valence of responsiveness across perspective camps. From a deliberative perspective, responsiveness is especially important between individuals who disagree with each other because it is the first step toward mutual understanding (Morrell, 2018). This is contrasted by the rather persistent idea of online discussions functioning as echo chambers in which people mainly engage with like-minded individuals rather than with those who hold different views (Sunstein, 2007). Research, however, suggests that this echo chamber effect is overestimated. Heatherly et al. (2017: 1283), for example, show that “although like-minded conversations occur more frequently, substantial levels of cross-cutting exchanges are also occurring on SNSs”—a finding that is corroborated by many other studies (e.g. Bond and Sweitzer, 2022; Karlsen et al., 2017; Yarchi et al., 2021). Empirically, such cross-cutting or heterogeneous discussions are usually operationalized as discussions between citizens with opposing political ideologies or party affiliations (Amsalem et al., 2022; Heatherly et al., 2017). However, even when individuals belong to the same political spectrum or political party, they can hold different views on public issues. Research shows that cross-cutting (online) public discussions typically take place between several larger perspective camps, that is, between coalitions of debaters who share the same general perspective on the discussed issue, rather than merely between political camps (Bodrunova et al., 2019; Wessler et al., 2008). Yet, existing online deliberation research does not account for whether a comment is responsive toward an individual from the same or from a different perspective camp. This is a considerable shortcoming because even responsive comments may not foster a constructive and fruitful discussion when they carry a hostile or negative valence (Rossini, 2022). Wollebæk et al. (2019) show that people who are angry are generally more likely to participate in online debates. Correspondingly, studies on online polarization suggest that negative sentiments dominate online discussions, both within and across broader perspective

camps (Marchal, 2022; Yarchi et al., 2021). However, they showed that exchanges between “opposed users were significantly more negative than like-minded ones” (Marchal, 2022: 376), at least on some platforms (Yarchi et al., 2021). We aim to merge these findings with the literature on online deliberation. Thus, we explore the extent to which online news reader comments are responsive toward debate participants from the same or the opposing perspective camp and study the valence of such responsive encounters. We hypothesize the following:

H3a. User comments that are responsive across perspective camps are more likely to contain negative than positive sentiments toward the commenters referred to.

H3b. User comments that are responsive within the same perspective camp are more likely to contain positive than negative sentiments toward the commenters referred to.

Methodology

We analyzed a dataset of 30,753 news reader comments using manual and computational content analysis. The comments were collected from the website comment sections and Facebook pages of leading news media in Australia, the United States, Germany, and Switzerland. They were posted below news articles that thematize how Western societies should shape their relationship to Muslims and Islam, published between August 2015 and July 2016. In Switzerland, we analyzed only the German-language discourse. The hypotheses, study design, and analysis plan were preregistered on The Open Science Framework (OSF) before the analysis – where our data, software scripts and appendices are also available.²

Topic of discussion

Public discussions on Western societies’ relationship to Muslims and Islam were especially salient in the investigated countries in 2015/2016, a period marked by the challenges of an unprecedented global refugee movement propelled by refugee flows from Muslim-majority nations (United Nations High Commissioner for Refugees (UNHCR), 2015). In Germany and Switzerland, the issue was of heightened concern because Europe was particularly affected by the developments. In Australia and the United States, it was a high-profile issue in advance of the 2016 elections in these countries. The cross-national relevance of the issue makes it well suited for our comparative analysis.

Data collection and sampling

The 30,753 news reader comments analyzed in this study stem from a large-scale research program on mediated contestation. The comments were collected and sampled in a carefully designed three-step procedure:

First, we built a base corpus of 400 online news articles covering issues related to the public role of religion and secularism in society (100 per country). These articles were published between August 2015 and July 2016 by leading news websites in Australia, the

United States, Germany, and Switzerland. The hosting news organizations were deemed among the most relevant in terms of national reach and importance for the domestic political discourse by 16 or more scientific experts who our research team surveyed in each country. Furthermore, 15 or more colleagues in each nation pinpointed prominent debates on the public role of religion and secularism in the respective country and provided signature words linked to these discussions. In an expert-informed topic modeling process, these signature words were then used to extract the 400 thematically significant online news articles from a comprehensive news discourse dataset (see Rinke et al., 2022 for further details).

Second, we filtered all online news articles from the base corpus that (a) had an associated reader comment section directly on the news website (115 of the 400 articles) and/or (b) had been shared by the publishing news organizations on their Facebook page (76 of the 400 articles). The news reader comments on these articles were then collected, amounting to 30,343 website and 54,230 Facebook comments.

Third, we sampled from the collected comments. This study focuses on online discussions on how Western societies should shape their relationship to Muslims and Islam rather than on the broader field of issues related to the public role of religion. Two coders carefully read and assessed all news articles with a comment section either on the website or on Facebook. Consensually, they decided whether each article was related to the subtopic and should be sampled. Overall, 91 articles covered the relevant topic. They were published by *The Australian Broadcasting Corporation*, *The Guardian Australia*, *The Sydney Morning Herald*, *CNN*, *The New York Times*, *The Wall Street Journal*, *The Washington Post*, *Der Spiegel*, *Die Tagesschau*, *Die Zeit*, *Der Blick*, *20 Minuten*, *Schweizer Radio und Fernsehen*, and *Tagesanzeiger*. Of these, 64 news articles had a website comment section and 41 were shared by the outlet on Facebook. In sum, 31,753 comments were posted on these articles. Since our computational measurement cannot classify visuals, we investigated only the written portion of these posts and filtered out comments without text. Ultimately, we analyzed 30,753 comments (18,488 website and 12,265 Facebook comments).³

Variables

This section describes the variables used in the analysis (see Supplementary Appendix II on The Open Science Framework (OSF) for codebook). The codebook was kept simple to facilitate the automated measurement. The descriptive frequencies reported per category are based on the manually coded data.

Issue perspective. Online discussions on how Western societies should shape their relationship to Muslims and Islam include voices from diametrically opposed perspective camps: While commenters with a cosmopolitan issue perspective plead for a society that is open and welcoming to Muslims and their religion, those with a communitarian issue perspective expect Muslims to adapt to the (Christian) cultural traditions of Western societies and to keep their religion private. We studied only comments that contain a clear issue perspective. This was particularly important to explore the valence of comments that are responsive within and across perspective camps (H3a and H3b). We thus

coded whether a comment contained a (1) *cosmopolitan issue perspective* ($n=2566$), a (2) *communitarian issue perspective* ($n=2180$), or whether (0) *no clear perspective* was detectable ($n=5089$). Subsequently, only comments that contained an issue perspective were coded for responsiveness.

Responsiveness. The coding unit was the comment. A comment was considered (1) *responsive toward a person* when it personally addressed or referred to another debate participant ($n=1167$) or (2) *responsive toward a collective* when it addressed or referred to a group of other debate participants ($n=109$). This could take the form of mentioning commenters by their real-world or usernames as well as directly addressing them through questions, requests, or pleas. The same comment could be responsive toward a person and a collective (this was the case for 12 comments). If none of this applied, the comment was (0) *not responsive* ($n=3482$).

Camp affiliation. For camp affiliation, we used the commenters (or groups of commenters) addressed or referred to in the responsive comments as the coding unit. They belonged to the (2) *same camp* when they supported the issue perspective that the author of the comment identified with in their post (cf. variable “issue perspective”) ($n=65$). They belonged to the (1) *opposing camp* when they supported the issue perspective that the author of the comment did not identify with in their post ($n=991$). Otherwise, (0) *no clear camp affiliation* was detectable ($n=220$). The camp affiliation had to be inferred from the text of the comments.

Valence. The instances of responsiveness coded previously were used as the coding units. The reference contained a (1) *positive evaluation* when it was approving or supportive of the commenter(s) referred to ($n=10$) and a (2) *negative evaluation* when it was disapproving of or dissenting with the commenter(s) referred to ($n=460$). If the reference carried neither a positive nor negative sentiment, (0) *no clear evaluation* was detectable ($n=806$).

Manual and computational content analyses

We used manual and supervised machine learning approaches to measure the variables. This entails the steps of (1) manually labeling gold standard data and (2) building and evaluating the machine learning models.

Gold standard data. Three coders produced a gold standard dataset of 9835 news reader comments, where all comments were randomly selected from the entire dataset of 30,753 comments. First, they coded whether each comment contained an issue perspective (48%). For all 4746 comments that did, they then coded whether each comment was responsive (27%). For all 1264 responsive comments, the coders then scored the camp affiliation of the fellow debaters referred to and the valence carried by this reference (for 1276 instances of responsiveness in the 1264 responsive comments). The coders received intensive training. Table 1 shows the intercoder reliabilities from the pretest. Since acceptable Krippendorff’s alpha values of $>.7$ were reached for issue perspective and

Table 1. Inter-coder reliabilities from pretest ($N=200$ per variable).

Variable	Percent agreement	Krippendorff's alpha
Issue perspective	73.0	.71
Responsiveness	88.5	.76
Camp affiliation	74.5	.56
Valence	71.5	.61

responsiveness, each comment was randomly assigned and scored by one coder in the main coding (single coding). In contrast, reliability remained low for camp affiliation and valence. Both variables could only be coded for comments previously scored for issue perspective and responsiveness. Therefore, a limited amount of prescored data was available to train the coders and the cost to produce additional test coding data was prohibitive, as indicated by the percentages at the beginning of this paragraph. This made coder training for these variables challenging. Thus, to ensure high data quality, camp affiliation and valence were triple-coded. Each comment was scored by all three coders, and the majority code used in the gold standard. If there was no majority, the coders determined the final code consensually (this was the case for only six coding instances).

Machine learning procedures. The gold standard data were split into a training and test set on an 80:20 basis. The training set was used for training the machine learning models and the test set was used to evaluate predictive performance. We needed to train four models to predict the issue perspective (for filtering comments with an issue perspective), responsiveness (H1 and H2), camp affiliation (H3a and H3b), and valence (H3a and H3b).

For this, we applied the so-called transfer learning technique. It requires a pre-trained language model, which is trained on a large corpus of texts for other tasks, for example, for general natural language understanding. There are several of these pre-trained models available, with Google BERT being the most common one. As we studied comments in two languages (English and German), an extension to BERT called XLM-RoBERTA (Conneau et al., 2019) was used because of its state-of-the-art multilingual capabilities.

We fine-tuned the pre-trained XLM-RoBERTA model with our training data. Thereby, the model learned the domain knowledge from the training data (e.g. recognition of issue perspective) while also having the pre-trained general language understanding. Thus, unlike with the usual “end-to-end” training (Barberá et al., 2021), our machine learning model did not need to learn the usage of language from scratch. For details about the machine learning procedure, please refer to our software scripts on OSF.

We found that our models adequately capture issue perspective (F1: 82.4) and responsiveness (F1: 89.7), but cannot predict camp affiliation and valence, for which the trained classifiers sorted all cases into the majority class. This problem probably arose due to the heavy imbalance in labels in the hand-coded data: There are only 10 instances of positive evaluation and 65 instances of addressing someone from the same perspective camp. The logical step would be to increase the hand-coded data size and thus the number of rare instances. However, we saw this as infeasible because the

prevalence of these instances is so low (<0.1% for positive evaluation). For example, to obtain one additional case of positive evaluation, one would probably need to code a further 1000 comments. Even when one would code all our available data (30,753 cases), probably, only 28 cases of positive evaluation would occur—an amount likely still insufficient for the machine learning. The modified analytical strategy to solve this issue is elaborated in the next section.

Data analysis. Our preregistered plan was to code all variables with the machine learning classifiers and to analyze these data. Due to the failure to automatically classify camp affiliation and valence, following Fong and Tyler (2021), we changed our plan to a two-step strategy: *In the first step*, we performed a traditional manual content analysis of the randomly selected and humanly coded data ($N=9835$ comments). *In the second step*, we then performed a computational content analysis in which we classified and analyzed the two variables for which the machine classification was possible in the whole dataset ($N=30,753$ comments). This means that H1 and H2 were tested in both the manual and computational content analysis because issue perspective and responsiveness could be classified automatically. In contrast, H3a and H3b were only tested in the manual content analysis because camp affiliation and valence could not be classified automatically. The advantage of this approach is that the computational analysis is a robustness check of the effects tested in H1 and H2.

When both steps are available, the estimator from the manual content analysis (“manual estimator”) is an accurate point estimate but with a much wider interval estimate. One would also expect that the estimator from the computational content analysis (“computational estimator”) would introduce measurement errors. These errors would bias the estimate toward zero, in other words decrease the anticipated statistical power (Fong and Tyler, 2021). For the analysis that used both steps, we adjusted the significance level according to Bonferroni. The overall family-level significance level was maintained at 5%. This setup increases the false negative rate (i.e. it is more difficult to reject the null hypothesis) but not the false positive rate (i.e. it is not easier to reject the null hypothesis).

We performed Bayesian multilevel logistic regressions. These regressions can reliably estimate country-level ecological effects as proposed in H1 and H2 (De Leeuw et al., 2020). Furthermore, our computational estimator is based on our entire dataset (not a random sample), and therefore, the traditional frequentist approach is not a valid choice (Western and Jackman, 1994). For all estimations, we added vary terms (L2) for the country and article identifier because comments from the same country and the same article are likely to be correlated. As we lacked prior information on how the predictors would behave, the default non-informative priors were used in all models. The R package brms (Bürkner, 2017) was used for the analysis.

Results

Our first hypothesis assumed that user-generated debates would be more integrated in consensus-oriented than in majoritarian democracies. The descriptive data show that the level of responsiveness in the positional news reader comments is similar across the countries studied: The share of comments personally addressing or referring to

Table 2. Bayesian multilevel logistic regression predicting the responsiveness of positional comments by democracy types: Estimator from manual and computational content analysis.

Independent variable	Manual estimator (97.5% HDI)	Computational estimator (97.5% HDI)
Intercept	-1.04 (-1.53 to -0.40)	-1.14 (-1.64 to -0.48)
Majoritarian democracy (vs consensus-oriented)	0.02 (-0.81 to 0.75)	0.12 (-0.66 to 0.88)
Varying intercept		
Country, <i>SD</i>	0.21 (0 to 1.25)	0.23 (0 to 1.55)
Story, <i>SD</i>	0.49 (0.36 to 0.65)	0.51 (0.40 to 0.65)

Note. HDI: Highest Density Interval; SD: Standard Deviation.

Table 3. Bayesian multilevel logistic regression predicting the responsiveness of positional reader comments by arena types: Estimator from manual and computational content analysis.

Independent variable	Manual estimator (97.5% HDI)	Computational estimator (97.5% HDI)
Intercept	-1.07 (-1.38 to 0.71)	-1.17 (-1.46 to -0.83)
News website (vs Facebook)	0.05 (-0.19 to 0.31)	0.09 (-0.07 to 0.25)
Varying intercept		
Country, <i>SD</i>	0.13 (0 to 0.73)	0.14 (0 to 0.83)
Story, <i>SD</i>	0.49 (0.36 to 0.65)	0.52 (0.41 to 0.66)

Note. HDI: Highest Density Interval; SD: Standard Deviation.

another debate participant or a group of fellow debaters in Germany (25.4%) and Switzerland (27.1%) is on par with the United States (26.3%) and Australia (24.9%). Thus, there does not appear to be a systematic difference between consensus-oriented and majoritarian democracies. Table 2 shows the estimators from the manual and computational content analysis (97.5% Highest Density Interval (HDI)), adjusted for country and story differences. Neither estimator provides enough evidence to support the assumption that positional comments would be more responsive in consensus-oriented than in majoritarian democracies (manual estimator: 0.02, 97.5% HDI -0.81 to 0.75; computational estimator: 0.12, 97.5% HDI: -0.66 to 0.88). Even with the more powerful computational estimator, the effect size (adjusted odds ratio: 1.12) is too small. Thus, H1 was rejected.

Our second hypothesis suggested that user-generated debates would be more integrated in arenas that separate public and private contexts more clearly than in those that mix the two. With respect to the platforms, the descriptive data show that there is a 5.2% difference in positional reader comments being responsive between the website comment sections (24%) and the Facebook pages (29.2%) of mainstream news media. Thus, unlike expected in our hypothesis, user-generated debates seem to be more integrated in the arena that mixes public and private contexts than in the one that separates them more clearly. Table 3 displays the estimators from the manual and computational content

Table 4. Distribution of valence and camp affiliation of commenters addressed in instances of responsiveness (hand-coded gold standard data).

	Negative evaluation	No clear evaluation	Positive evaluation	All
No clear camp affiliation	3 (1.4%)	216 (98.2%)	1 (0.5%)	220 (17.2%)
Opposing camp	457 (46.1%)	532 (53.7%)	2 (0.2%)	991 (77.7%)
Same camp	0 (0%)	58 (89.2%)	7 (10.7%)	65 (5.1%)
All	460 (36.1%)	806 (63.2%)	10 (0.8%)	1276 ^a

^aThe number of instances of responsiveness (1276) deviates from the number of responsive comments (1264) because a comment could contain several instances of responsiveness (only 12 comments did).

analysis, adjusted for country and story differences. Like indicated by the descriptive data, these estimators do not support our assumption that positional comments would be more responsive in arenas that separate rather than mix public and private contexts (manual estimator: 0.05, 97.5% HDI: -0.19 to 0.31; computational estimator: 0.09, 97.5% HDI: -0.07 to 0.25). Thus, H2 was also rejected. Yet, the analysis did also not confirm an effect in the opposite direction. Even with the high-power computational estimator, this effect (adjusted odds ratio: 1.09) is too small.

Finally, we assumed that user comments that are responsive across perspective camps would be more likely to contain negative than positive sentiments toward the commenters referred to (H3a), whereas user comments that are responsive within the same perspective camp would be more likely to contain positive than negative sentiments (H3b). As explained in the “Methodology” section, these hypotheses were only tested in the manual content analysis because the machine learning classifier could not reliably detect camp affiliation and valence.

Table 4 displays the distribution of the camp affiliation of and the valence toward the commenters addressed in the hand-coded instances of responsiveness. Only few instances of responsiveness carry a positive evaluation of the commenters referred to (0.8%) and if this is the case, it usually happens when those commenters belong to the same perspective camp as the commenting individual (7 out of 10). In contrast, many instances of responsiveness carry a negative evaluation of the commenters addressed (36.1%) and these negative sentiments were overwhelmingly directed at commenters who belonged to the opposing perspective camp (99.3% of the negative evaluations). This suggests strong support for our hypotheses.

Table 5 displays the regression results for H3a and H3b—in fact, both hypotheses are strongly supported. Positional reader comments that are responsive across perspective camps are indeed more likely to contain negative than positive sentiments toward the commenters referred to (H3a). In contrast, positional reader comments that are responsive within the same perspective camp are more likely to contain positive than negative sentiments toward the commenters referred to (H3b). The effect size is extremely large in both cases (adjusted odds ratio for negative evaluation by commenters from the opposing camp: 40; adjusted odds ratio for positive evaluation by commenters from the same camp: 8.25). Thus, the valence seems to be strongly associated with camp affiliation.

Table 5. Bayesian multilevel logistic regression predicting the valence (evaluation) by camp affiliation: Estimator from manual content analysis.

Independent variable	Manual estimator: Negative evaluation (95% HDI) ^a	Manual estimator: Positive evaluation (95% HDI)
Intercept	-3.86 (-4.80 to -2.96)	-0.45 (-287 to 1.91)
Opposing camp (vs no camp)	3.69 (2.93 to 4.57)	-0.97 (-2.24 to 0.31)
Same camp (vs no camp)	-0.49 (-1.91 to 0.82)	2.11 (0.92 to 3.32)
Varying intercept		
Country, <i>SD</i>	0.24 (0.01 to 1.07)	4.9 (2.05 to 9.94)
Story, <i>SD</i>	0.47 (0.17 to 0.78)	0.61 (0.02 to 1.91)
Comment, <i>SD</i>	0.92 (0.1 to 1.7)	

^aOf the 12 comments that contain several instances of responsiveness, 2 are responsive across camps. Each of the two comments contains one instance of responsiveness with a negative evaluation and one with a neutral evaluation. The unit of analysis in this case is the instance of responsiveness. The multilevel structure thus needs to be adjusted by entering comment as L2 and story as L3. This does not affect the analysis of positive evaluation. Note. HDI: Highest Density Interval; SD: Standard Deviation.

Discussion

In this preregistered study, we investigated responsiveness as one key aspect of reciprocity in online discussions. Responsiveness gauges whether online commenters address or refer to other debate participants in their posts. While responsiveness was measured in the individual comments, at the aggregate level of the discussions, we refer to the degree to which online user comments are responsive as *online discourse integration*. Empirically, we studied positional online news reader comments that contained a clear issue perspective.

We hypothesized that user-generated debates would be more integrated in consensus-oriented than in majoritarian democracies (H1), which we studied by comparing comments from Germany and Switzerland versus Australia and the United States. Furthermore, we assumed that user-generated debates would be more integrated in arenas that separate rather than mix public and private contexts (H2), comparing comments from the websites and Facebook pages of mainstream news media. Unlike hypothesized, the patterns of online discourse integration were consistent across the different types of democracy and platforms. The level of responsiveness was low in all four countries and both discussion arenas, with only about one quarter of the comments being responsive and no significant differences. This was unexpected because previous research shows that both the type of democracy (Jakob et al., 2023a, 2023b) and different platform architectures (Esau et al., 2017) shape other deliberative quality criteria such as argumentation and civility in online discussions. Moreover, our results did not substantiate initial evidence by Rowe (2015), who found a positive effect of separated contexts on responsiveness in online discussions. Thus, our findings further nuance technical affordance theories (Nagy and Neff, 2015) and theories about the importance of political systems for public debates (Steiner et al., 2004). One possible theoretical interpretation of the null effects would be that online discourse integration is not shaped by the type of democracy or the degree of context collapse afforded by a platform's architecture. However, they could also be explained by

other factors such as the specific selection of democratic countries, media outlets or online platforms in our dataset, the news value of the articles the investigated comments relate to (Ziegele et al., 2020), or the deliberative quality of previous comments (Esau and Friess, 2022). Future research should aim to disentangle the different influences at play here. Overall, the low levels of responsiveness we found across countries and platforms raise questions about the role of news media in democratic public discourses, especially with respect to how these media could foster more integrated public debates. Could they, for example, moderate online discussions more strongly to encourage responsiveness? And would a more constructive and conciliatory reporting style foster responsiveness between people with different opinions?

Independent of the country and platform context, we found that a large majority of the positional news reader comments that *were* responsive addressed or referred to fellow debaters who held the opposing issue perspective—and that these instances of responsiveness were overwhelmingly negative in sentiment. Only few comments were responsive toward individuals who belonged to the same perspective camp, but when this was the case, the comments were significantly more likely to contain positive than negative evaluations of the commenters addressed. These results supported both parts of our third hypothesis (H3a/b). The fact that most responsiveness occurred between commenters who belonged to opposing rather than to the same perspective camp substantiates previous research which showed that cross-cutting exchanges are indeed taking place in online discussions (e.g. Heatherly et al., 2017; Karlsen et al., 2017). Our findings thus contribute to disconfirming the widespread notion of online discussions being fragmented into echo chambers of like-minded individuals (Sunstein, 2007). Simultaneously, they confirm the results of studies on online polarization, which suggested that responsive references between individuals from different perspective camps would be dominated by negative sentiments (Marchal, 2022; Yarchi et al., 2021). From a normative perspective, this constitutes cause for concern: On one hand, online commenters with opposing perspectives do talk to each other, which is a prerequisite to solve conflicts and find commonly acceptable solutions (Gutmann and Thompson, 2002). On the other hand, however, the predominantly negative sentiment of such responsive references likely hampers the constructive debate that is needed to achieve these goals (Rossini, 2022).

Conceptually, this latter finding reminds us that responsiveness measures only the procedural but not the substantive component of reciprocity (Gutmann and Thompson, 2002). Responsiveness gauges whether online commenters address or refer to other debate participants, capturing “the relational and structural aspects of [public] communication” (Esau and Friess, 2022: 2). It is, however, not concerned with *what* online debaters talk about or *how* they communicate with each other, that is, it does not capture whether commenters truly consider the reasons and perspectives of fellow citizens (Scudder, 2020). This study focused on responsiveness as the procedural component of reciprocity because this allowed us to study the discussions in our dataset in their entirety. While the substantive component of reciprocity is hard to gauge empirically (Friess et al., 2021), due to its focus on the structural aspects of online discussions, responsiveness is well suited for the development of computational classifiers that can score large corpora of social media data. Due to this procedural approach, however, our study cannot speak to the substantive quality of responsive engagements. Esau and

Friess (2022: 2) recently suggested moving away from the procedural construct of responsiveness toward “the more demanding concept of *deliberative reciprocity* (coherent, respectful and reasoned replies)”. While our computational analysis provides unique insights into how integrated online discourses are as a whole, it cannot capture such a complex and multidimensional theoretical construct. Future research could aim to better balance this “trade-off between powerful, scalable computational strategies, and the theoretical sensitivity offered by small-scale manual analyses” (Baden et al., 2020: 165). Another limitation of our study is that our preregistered plan to measure all variables computationally was prevented by the fact that supervised machine learning approaches are limited by the amount of training data available. This became problematic in measuring the rare events of positive sentiment and addressing someone from the same perspective camp. Finally, our study focused on analyzing comments that contained a clear issue perspective because we considered these comments to be the most challenging case for discourse integration. The levels of responsiveness could be higher in comments that do not express a clear issue perspective, but focusing on positional comments enabled us to examine the dynamics between different perspective camps that previous research had not considered. Overall, our study provided an unprecedented systematic cross-country–cross-platform comparison of responsiveness in online discussions. Future research should build on our insights and continue to disentangle the factors that shape this dimension of online debate quality, while also applying more complex measures of deliberative reciprocity.

Authors' note

All authors have agreed to the submission, and they hereby confirm that the article is not currently being considered for publication by any other print or electronic journal.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant number 260291564.

ORCID iD

Julia Jakob  <https://orcid.org/0000-0003-2340-5193>

Supplemental material

Supplemental material for this article is available online at The Open Science Framework (OSF): <https://doi.org/10.17605/OSF.IO/XQ8HG>.

Note

1. The preregistration is available at: <https://osf.io/sjgf8>
2. The preregistration, data, software scripts and appendices are available at: <https://doi.org/10.17605/OSF.IO/XQ8HG>
3. Instead of the 30,753 comments analyzed in this study, the pre-registration mistakenly lists the sample size before the third data selection step as the analytical basis ($N=84,573$).

References

- Amsalem E, Merkley E and Loewen PJ (2022) Does talking to the other side reduce inter-party hostility? Evidence from three studies. *Political Communication* 39(1): 61–78.
- Baden C, Kligler-Vilenchik N and Yarchi M (2020) Hybrid content analysis: toward a strategy for the theory-driven, computer-assisted classification of large text corpora. *Communication Methods and Measures* 14(3): 165–183.
- Barberá P, Boydston AE, Linn S, et al. (2021) Automated text classification of news articles: a practical guide. *Political Analysis* 29(1): 19–42.
- Bodrunova SS, Blekanov I, Smoliarova A, et al. (2019) Beyond left and right: real-world political polarization in Twitter discussions on inter-ethnic conflicts. *Media and Communication* 7(3): 119–132.
- Bond RM and Sweitzer MD (2022) Political homophily in a large-scale online communication network. *Communication Research* 49(1): 93–115.
- boyd D (2011) Social network sites as networked publics: affordances, dynamics and implications. In: Papacharissi Z (ed.) *A Networked Self: Identity, Community and Culture on Social Network Sites*. New York: Routledge, pp. 39–58.
- Bürkner P (2017) brms: an R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1): 1–28.
- Camaj L (2021) Real time political deliberation on social media: can televised debates lead to rational and civil discussions on broadcasters' Facebook pages? *Information, Communication & Society* 24(13): 1907–1924.
- Cargnino M and Neubaum G (2021) Are we deliberately captivated in homogeneous cocoons? An investigation on political tie building on Facebook. *Mass Communication and Society* 24(2): 187–209.
- Collins L and Nerlich B (2015) Examining user comments for deliberative democracy: a corpus-driven analysis of the climate change debate online. *Environmental Communication* 9(2): 189–207.
- Conneau A, Khandelwal K, Goyal N, et al. (2019) Unsupervised cross-lingual representation learning at scale. Available at: <https://arxiv.org/abs/1911.02116> (accessed 27 July 2022).
- De Leeuw SE, Azrout R, Rekker RSB, et al. (2020) After all this time? The impact of media and authoritarian history on political news coverage in twelve Western countries. *Journal of Communication* 70(5): 744–767.
- Dobson A (2012) Listening: the new democratic deficit. *Political Studies* 60(4): 843–859.
- Esau K and Friess D (2022) What creates listening online? Exploring reciprocity in online political discussions with relational content analysis. *Journal of Deliberative Democracy* 18(1): 1–16.
- Esau K, Friess D and Eilders C (2017) Design matters! An empirical analysis of online deliberation on different news platforms. *Policy & Internet* 9(3): 321–342.
- Fong C and Tyler M (2021) Machine learning predictions as regression covariates. *Political Analysis* 29(4): 467–484.
- Freelon D (2015) Discourse architecture, ideology and democratic norms in online political discussion. *New Media & Society* 17(5): 772–791.
- Friess D, Ziegele M and Heinbach D (2021) Collective civic moderation for deliberation? Exploring the links between citizens' organized engagement in comment sections and the deliberative quality of online discussions. *Political Communication* 38(5): 624–646.
- Graham T and Witschge T (2003) In search of online deliberation: towards a new method for examining the quality of online discussions. *Communications* 28(2): 173–204.
- Gutmann A and Thompson DF (2002) Deliberative democracy beyond process. *The Journal of Political Philosophy* 10(2): 153–174.
- Heatherly KA, Lu Y and Lee JK (2017) Filtering out the other side? Cross-cutting and like-minded discussions on social networking sites. *New Media & Society* 19(8): 1271–1289.

- Hughes DJ, Rowe M, Batey M, et al. (2012) A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior* 28(2): 561–569.
- Jakob J, Dobbrick T, Freudenthaler R, et al. (2023a) Is constructive engagement online a lost cause? Toxic outrage in online user comments across democratic political systems and discussion arenas. *Communication Research* 50(4): 508–531.
- Jakob J, Dobbrick T and Wessler H (2023b) The integrative complexity of online user comments across different types of democracy and discussion arenas. *The International Journal of Press/Politics* 28(3): 580–600.
- Karlsen R, Steen-Johnsen K, Wollebæk D, et al. (2017) Echo chamber and trench warfare dynamics in online debates. *European Journal of Communication* 32(3): 257–273.
- Karlsson M (2012) Understanding divergent patterns of political discussion in online forums: evidence from the European citizens' consultations. *Journal of Information Technology & Politics* 9(1): 64–81.
- Lijphart A (2012) *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*. New Haven, CT: Yale University Press.
- Lönnqvist J-E and Itkonen JVA (2015) Homogeneity of personal values and personality traits in Facebook social networks. *Journal of Research in Personality* 60: 24–35.
- Marchal N (2022) “Be nice or leave me alone”: an intergroup perspective on affective polarization in online political discussions. *Communication Research* 49(3): 376–398.
- McCoy J and Somer M (2019) Toward a theory of pernicious polarization and how it harms democracies: comparative evidence and possible remedies. *The Annals of the American Academy of Political and Social Science* 681(1): 234–271.
- Morrell M (2018) Listening and deliberation. In: Bächtiger A, Dryzek JS, Mansbridge J and Warren M (eds) *The Oxford Handbook of Deliberative Democracy*. Oxford: Oxford University Press, pp. 237–250.
- Nagy P and Neff G (2015) Imagined affordance: reconstructing a keyword for communication theory. *Social Media + Society* 1(2): 1–9.
- Nelson JL and Webster JG (2017) The myth of partisan selective exposure: a portrait of the online political news audience. *Social Media + society* 3(3): 1–13.
- Pedrini S, Bächtiger A and Steenbergen MR (2013) Deliberative inclusion of minorities: patterns of reciprocity among linguistic groups in Switzerland. *European Political Science Review* 5(3): 483–512.
- Rinke EM, Dobbrick T, Löb C, et al. (2022) Expert-informed topic models for document set discovery. *Communication Methods and Measures* 16(1): 39–58.
- Rossini P (2022) Beyond incivility: understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research* 49(3): 399–425.
- Rowe I (2015) Deliberation 2.0: comparing the deliberative quality of online news user comments across platforms. *Journal of Broadcasting & Electronic Media* 59(4): 539–555.
- Ruiz C, Domingo D, Micó JL, et al. (2011) Public sphere 2.0? The democratic qualities of citizen debates in online newspapers. *The International Journal of Press/Politics* 16(4): 463–487.
- Schudson M (1997) Why conversation is not the soul of democracy. *Critical Studies in Mass Communication* 14(4): 297–309.
- Scudder MF (2020) The ideal of uptake in democratic deliberation. *Political Studies* 68(2): 504–522.
- Stapleton CE and Dawkins R (2022) Catching my anger: how political elites create angrier citizens. *Political Research Quarterly* 75(3): 754–765.
- Steiner J, Bächtiger A, Spöndli M, et al. (2004) *Deliberative Politics in Action: Analyzing Parliamentary Discourse*. Cambridge: Cambridge University Press.

- Strandberg K and Berg J (2013) Online newspapers' readers' comments—democratic conversation platforms or virtual soapboxes? *Comunicação e Sociedade* 23: 132–152.
- Sunstein C (2007) *Republic.com 2.0*. Princeton, NJ: Princeton University Press.
- United Nations High Commissioner for Refugees (UNHCR) (2015) Global trends: forced displacement in 2015. Report, United Nations High Commissioner for Refugees. Available at: <https://www.unhcr.org/media/unhcr-global-trends-2015>
- Valera-Ordaz L (2019) Liberal individualist, communitarian, or deliberative? Analyzing political discussion on Facebook based on three notions of democracy. *International Journal of Communication* 13: 1056–1076.
- Vitak J (2012) The impact of context collapse and privacy on social network site disclosures. *Journal of Broadcasting & Electronic Media* 56(4): 451–470.
- Wessler H, Althaus SL, Chan C, et al. (2022) Multiperspectival normative assessment: the case of mediated reactions to terrorism. *Communication Theory* 32(3): 363–386.
- Wessler H, Peters B, Brüggemann M, et al. (2008) *Transnationalization of Public Spheres*. Basingstoke: Palgrave Macmillan.
- Western B and Jackman S (1994) Bayesian inference for comparative research. *American Political Science Review* 88(2): 412–423.
- Wollebæk D, Karlsen R, Steen-Johnsen K, et al. (2019) Anger, fear and echo chambers: the emotional basis for online behavior. *Social Media + Society* 5(2): 1–14.
- Yarchi M, Baden C and Kligler-Vilenchik N (2021) Political polarization on the digital sphere: a cross-platform, over-time analysis of interactional, positional and affective polarization on social media. *Political Communication* 38(1–2): 98–139.
- Ziegele M, Quiring O, Esau K, et al. (2020) Linking news value theory with online deliberation: how news factors and illustration factors in news articles affect the deliberative quality of user discussions in SNS' comment sections. *Communication Research* 47(6): 860–890.

Author biographies

Julia Jakob is a researcher in political communication at the University of Mannheim. Her work is particularly concerned with how the theory of deliberation can be further developed to accommodate the peculiarities of public communication in the digital age.

Chung-hong Chan is a Postdoctoral Research Fellow at the Mannheim Centre for European Social Research. An epidemiologist by training, he is interested in developing new quantitative methods for communication research.

Timo Dobbrick is a former research associate and collaboration partner at the Mannheim Centre for European Social Research. His focus is on data science, machine learning, and natural language processing.

Hartmut Wessler is a Professor for Media and Communication Studies at the University of Mannheim and a Principal Investigator at the Mannheim Centre for European Social Research. A recurring theme of his research relates to the possibilities of assessing the quality of mediated contestation against diverging normative models of democracy.