Hate Beneath the Counter Speech? A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos

By[1]: Julian Ernst[2]; Josephine B. Schmitt[3]; Diana Rieger[4]; Ann Kristin Beier[5]; Peter Vorderer[6]; Gary Bente[7]; Hans-Joachim Roth [8]

Abstract

The odds in stumbling over extremist material in the internet are high. Counter speech videos, such as those of the German campaign *Begriffswelten Islam* (Concepts of Islam; Bundeszentrale für politische Bildung, 2015a) published on YouTube, offer alternative perspectives and democratic ideas to counteract extremist content. YouTube users may discuss these videos in the comment sections below the video. Yet, it remains open which topics these users bring up in their comments. Moreover, it is unknown how far user comments in this context may promote hate speech—the very opposite of what counter speeches intent to evoke. By applying a qualitative content analysis on a randomly selected sample of user comments, which appeared beneath the counter speech videos of *Concepts of Islam*, we found that comments dominated, which dealt with devaluating prejudices and stereotypes towards Muslims and/or Islam. However, we also discovered that users in a large scale discussed the content of the videos. Moreover, we identified user comments, which hint at hateful speech either in comments themselves or the discourse the comments are embedded in. Based on these results, we discuss implications for researchers, practitioners and security agencies.

Keywords: hate speech, counter speech, user comments, qualitative content analysis, anti-Muslim racism

---

[1] All authors are working in the project *CONTRA: Countering Propaganda by Narration Towards Anti-Radical Awareness,* commissioned by the European Union, that aims to prevent extremist attitudes and behaviors among the young generation by fostering critical media competence.

[2] Julian Ernst is junior researcher at the Chair for Intercultural Education, University of Cologne (Twitter: @JulianErnst91)

[3] Josephine B. Schmitt, PhD, is postdoctoral researcher at the Chair of Media and Communication Psychology, University of Cologne (Twitter: @JottBeEs)

[4] Diana Rieger, PhD, is postdoctoral researcher at the Department of Media and Communication Studies, University of Mannheim (Twitter: @DianaRieger79)

[5] Ann Kristin Beier is a student assistant at the Chair for Intercultural Education, University of Cologne

[6] Peter Vorderer, PhD, is professor at the Department of Media and Communication Studies, University of Mannheim

[7] Gary Bente, PhD, is professor for Media and Communication Psychology, University of Cologne, and visiting professor at the Department of Communication, Michigan State University

[8] Hans-Joachim Roth, PhD, is professor for Intercultural Education at the University of Cologne.

Introduction

There are thousands of online forums, blogs, social media offers and, particularly, videos on the Internet which disseminate extremist ideologies (Glaser, 2013; Hussain & Saltman, 2014)—and the number is steadily increasing. The odds in stumbling over extremist material are accordingly high (Klein, 2012; Rieger, Frischlich, & Bente, 2013). Various political and societal organizations aim at challenging the extremist and antidemocratic (online) content with different concepts of so-called *counter messages* respectively *counter speeches*. These offer—with limitations—alternative perspectives and democratic ideas to counteract radical messages respectively propaganda material (Briggs & Feve, 2013).

YouTube is a one of the key platforms on the internet distributing counter speech videos. Beside easily publishing and distributing nearly any kind of video material, YouTube provides a feature for creating user comments. However, user comments are not always related to the content they are published beneath, nor are they always constructive, friendly or in line with a basic understanding of a democratic discussion culture. Oftentimes, there are comments propagating hate, discrimination and calls for violence towards certain (imagined) groups such as Muslims, Jews, Sinti and Roma, refugees, and many more (e.g., Ben-David & Fernández, 2016; Leets, 2002). Such comments are often called as *hate speech*.

Until now, it is a desideratum in what way hate speech exists in user comments related to counter speech videos. These videos are discussed as an 'antidote' against extremist ideas and are used in order to foster tolerant attitudes and inoculate against propaganda effects (Frischlich, Rieger, Morten, & Bente, in press). Nevertheless, research concerning the effectiveness of narrative persuasion and counter messages found widely inconsistent results (for an overview, see Braddock & Dillard, 2016; Braddock & Horgan, 2016). Frischlich et al. (in press) did not even a clear effect of counter narratives on attitudes towards extremist content. Against this background, it seems to be even more, important to investigate the context in which such videos are thought to deliver their potential. From a societal perspective, it seems necessary to shed light on the kind and content of user comments related to the counter messages and counter speeches since this may deliver important information about and implications for the perception of those counter speech videos as well as for the strategies to moderate user comments propagating hatred. A video aiming at countering

extremist ideas or emphasizing alternatives instead might get a different connotation when comments propagate hate speech right beneath it.

In the presented study we focus on user comments beneath videos tagged with #whatIS, which have been released on YouTube within the framework of the counter speech campaign *Begriffswelten Islam* (Engl.: concepts of Islam) published by the German Federal Agency of Civic Education[9] (Bundeszentrale für politische Bildung (bpb)) in cooperation with popular actors of German YouTube scene. The campaign aims at targeting stereotypical representations and discussions generating biased opinions towards Islam respectively Muslims in Germany (Bundeszentrale für politische Bildung, 2015a). The bpb wants to promote a critical discourse about the topic, concurring to the principles of civic education, and explicitly against the hate speech; since it is considered to be increasingly relevant in terms of anti-Muslim prejudices in general, hateful messages and comments in social media in particular (Bundeszentrale für politische Bildung, 2015a).

Based on a qualitative content analysis of a random sample of user comments related to the videos that are tagged with #whatIS, we basically aim to answer three research questions: First, to estimate the potentials of counter messages in order to foster a critical awareness or reflection, we inquire which themes and topics are brought up in comments posted below counter speech videos. Second, to relate these topics to the content of counter speech videos and we further investigate to which aspects of the videos the comments refer to. Third, in order to focus on a specific aspect of hate speech in comments below the counter message videos, we question which topics serve as the indicators of hate speech.

The current paper thereby aims to investigate potentials of the counter messages for fostering critical awareness for the topics presented in such videos as a positive outcome. The research also considers the negative potential of counter messages to pave the way for hate speech. We believe that focusing on comments as a contemporary way of political participation, and questioning them with a content analysis presents an insightful approach for

---

[9] The German Federal Agency of Civic Education (Bundeszentrale für politische Bildung/bpb) is a public institution pursuing the provision of „[…] citizenship education and information on political issues for all people in Germany" (Bundeszentrale für politische Bildung, 2012a). Besides offering information about various topics relevant for civic education in the form of publications and events, the bpb initiates and takes part in social media activities (Bundeszentrale für politische Bildung, 2012b).

Julian Ernst et al.: A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos

researchers, practitioners and security agencies in order to better understand what counter messages are able and not able to unfold.

The role of user comments in the perception of online content

User comments may be considered as a standard feature of the web 2.0. Comments made mainly by unknown users appear below the various kind of online content (e.g., news articles, posts on social networking sites, videos). Motivation to start commenting on a media content stem from expressing an emotion or an opinion, adding information, correcting inaccuracies or misinformation as well as giving a personal perspective (Stroud, van Duyn, & Peacock, 2016). Numerous studies are dealing with user comments as a form of political *online* participation (e.g., Vitak et al., 2010; Weber, 2014), its potential with regard to the political deliberation (e.g., Anderson, Brossard, Scheufele, Xenos, & Ladwig, 2014; Halpern & Gibbs, 2013) as well as an important indicator for the different forms of *offline* political participation (e.g., Kruikemeier, Van Noort, & Vliegenthart, 2016; Vitak et al., 2010).

In communication research, relationship of the mass and interpersonal communication has been discussed a lot (e.g., Katz & Lazarsfeld, 1955; Eveland, Morey, & Hutchens, 2011; Eveland & Schmitt, 2015; Gastil & Dillard, 1999; Hardy & Scheufele, 2005; Katz, 1957). According to the differential gains model (Hardy & Scheufele, 2005) interpersonal discussion of media content—either online or face-to-face—may foster media effects. Therefore, it seems important to take these discussions, for example, in forms of comments happening below certain media content into account in order to get a more complete picture of these potential effects. There are various studies pointing at potential effects of the user comments on perception of the related online content (e.g., Kim, 2015; Lee, 2012; Lee & Jang, 2010; von Sikorski & Hänelt, 2016; Weber, 2014). For instance, it was found that the valence of comments affects the perceived journalistic quality, the trustworthiness, and persuasiveness of an online news article (von Sikorski & Hänelt, 2016). Lee and Jang (2010) demonstrated in an experiment that participants' opinion regarding a specific online content varied significantly as a function of other users' comments on the news article. People who read comments opposed to news content changed their attitudes concerning the news when compared to

people who read the news article without comments or comments supporting the articles' opinion (Lee & Jang, 2010).

In particular, with regard to online content constructed and uploaded with the intention to inform about socially important topics, decrease prejudices, and foster reflection of sensitive topics, user comments can be crucial for the effectiveness—particularly if there are even doubts whether this online content (i.e., counter messages) is able to meet one's expectations (see e.g., Frischlich et al., in press). Before claiming further need for such educational online material in the form of counter messages—for instance in the realm of countering populism, islamophobia or even extremism—it is important to consider what kind of user comments are triggered by such counter message material and to investigate topics that occur beneath videos. Therefore, we ask: Which themes and topics are brought up in the user comments beneath the videos (RQ1)?

Additionally, to relate the content of counter message material to the comments below, and to derive best practices for the topics that should be mentioned in media content aiming at informing about sensitive topics, an investigation of the relationship between topics discussed in counter messages and comments depicted below is necessary. To account for this, we formulated RQ2: Which aspects of the video content do the user comments refer to?

Hate Speech in Online Environments

The propagation of hate and prejudices against certain groups and minorities in media is a very old phenomenon. However, in the last decades it has been enormously facilitated by the rise of the Internet as an interactive space with its diverse opportunities (Cammaerts, 2009). On the one hand, properties such as search engines, blogs, and social networks affect the flow of communication (Hughey & Daniels, 2013; Klein, 2012). On the other hand, anonymity on the web fosters the incivility in user comments (Halpern & Gibbs, 2013; Rösner & Krämer, 2016). Once, the interactivity of the web 2.0 was accompanied by the hope to make online content more relevant and significant for people as well as to encourage discussions of recipients about the presented online content. However, more interactivity could also provide more possibilities to generate *hate speech* (Cammaerts, 2009; Erjavec & Poler Kovačič, 2012; Hughey & Daniels, 2013).

Definition and Characteristics of Hate Speech

Hate speech has been shown to be a complex phenomenon that is studied by various scientific disciplines. Besides the research in the field of communication, legal, sociological as well as linguistic research shape the discourse of the subject (e.g., Ben-David & Fernández, 2016; Butler 2013; Meibauer, 2013; Dharmapala & McAdams, 2005; Schabas, 2000). The sociologist Judith Butler (2013) examines the character of speech acts transporting hate in specific discourses, possibilities of responses, and the role of law and the repression in terms of reproduction of hate speech. According to Butler, words or speeches can be hateful just by their usage—even if this effect is not intended by the speaker (Butler, 2013). Butler, referring to Austin, understands hate speech as a speech act, i.e. she emphasizes the performative dimension of speech as an *act*. By the example of law cases she shows, how the citation of hate speech inevitably reproduces it and leads to an implicit legitimization, and stabilization in a discourse (Butler, 2013). Moreover, Roth (2013) discusses the relevance of Butler's theoretical position for discourses related to migration, in particular lingual diversity. Jörg Meibauer (2013), who studied hate speech from a linguistic perspective, defined hate speech as an intentional act. It can be (1) direct or indirect, (2) open or hidden, (3) promoted by power or not, (4) in company with violence or not and (5) more or less strong respectively intense. Meibauer's definition goes along with other social-scientific definitions (e.g., Ben-David & Fernández, 2016; Erjavec & Poler Kovačič, 2012). Hate speech includes insults, abusive language and designations that devaluate members of certain societal or demographic groups as well as minorities (e.g., religious groups, people with handicaps; Meibauer, 2013). In online environment hate speech should further be understood "as the tactical employment of words, images, and symbols, as well as links, downloads, news threads, conspiracy theories, politics, and even pop culture" (Klein, 2012, p. 428). It is frequently considered to be related to the right-wing ideology. Erjavec and Poler-Kovačič (2013) identified particularly four strategies of people publishing hate speech in comments below online content. According to them, hate speech comments include (1) general attacks on human dignity, (2) assaults on other commenters' personality based on their supposed belonging to certain social minorities, (3) attacks on the journalists' personality based on their supposed political orientation, (4) offences on the respective media company based on its supposed political orientation.

Julian Ernst et al.: A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos

Hate Speech and Counter Speech

These days in Germany as well as in other European countries, hate speech is to a large extent directed against Muslims, or people who are addressed as Muslims (jugendschutz.net, 2015; Zick, Küpper, & Hövermann, 2011). Beside general prejudices and stereotypes against Muslims and Islam in general, fragments of conspiracy theories are used by persons publishing hate speech (Amadeu Antonio Stiftung, 2015, 2013; Arbeitsgemeinschaft Kinder- und Jugendschutz (AJS) Landesstelle NRW e.V. & Landesanstalt für Medien Nordrhein-Westfahlen (LfM), 2016). Moreover, the use of pejorative language towards Islam, Muslims, or people who are addressed as Muslims has been shown to be an important indicator for hate speech (Amadeu Antonio Stiftung, 2015, 2013). The term *taqiyya* indicating the accusation against Muslims to constantly hide their "true" intention, for instance, harming people of other faiths, is a popular example for the use of conspiracy theories, propagandized especially by players close to right-wing ideology (Shooman, 2014). It comprises a selective collection of evidences, rather aiming to prove than to falsify the presented theory (Hepfer, 2016): the accusation of Muslims doing *taqiyya* can hardly be invalidated but easily claimed (Shooman, 2014).

Current counter messages or counter speeches aim at challenging these transmitted ideas of hatred, prejudice or even extremism. A "counter-narrative spectrum" exists, ranging from strategic counter messages (Dafnos, 2014), over alternative narratives transmitting values of tolerance or freedom, up to counter narratives which de-construct and challenge extremist ideologies (Braddock & Horgan, 2016; Briggs & Feve, 2013). By explaining central concepts of Islam, the videos of the counter message campaign *Begriffswelten Islam* ('Concepts of Islam') wants to encounter stereotypical representations and negatively biased opinions towards Islam respectively Muslims in Germany (Bundeszentrale für politische Bildung, 2015a). Conceptualizing this specific campaign within the counter-narrative spectrum outlined above, the videos of Concepts of Islam can mostly be considered as one example of strategic counter messages but also entail elements of countering extremist ideas and providing alternative approaches. Although counter messages are produced with the intention to foster reflection of populist arguments and provide information, they also run the risk to evoke reactance in the viewers (Kim, Levine, & Allen, 2014). Previous research on

counter messages found that viewers with more favorable attitudes towards extremist ideas are more likely to dislike counter messaging attempts (Frischlich et al., in press). In consequence, the same people – the real target group of counter messages—could feel the urge to comment beneath such videos and report their disagreement with the presented arguments.

This is also mirrored in research demonstrating that people comment on news and other users' comments if they are somehow personally involved in the topic or if they perceive the content to be controversial or threatening (Weber, 2014; Ziegele, 2016). Based on these considerations as well as the mixed results regarding the persuasive effectiveness of counter messages (e.g. Frischlich et al., in press), it seems necessary to evaluate the potential for counter messages to trigger hate speech comments beneath. In order to better understand the nature of comments featuring hate speech, in particular in relation to the content of the counter message video material, we come up with the following research question: Which themes and topics in user comments indicate hate speech—towards Islam, Muslim, or people who are addressed as Muslims—beneath counter message videos (RQ3)?

Method

Material

To answer our research questions we focus on all eight videos of the YouTube video campaign *Concepts of Islam* (Bundeszentrale für politische Bildung, 2015a) which are marked with #whatIS as well as their respective user comments. The videos have been published within the period between 12th October 2015 (video "What does UMMA mean? #whatIS") and 16th January 2016 (video „Islam and Knowledge #whatIS", see Table 1).

For further analyses, all eight videos were transcribed with the transcription software *f4transcript* (dr. dresing & pehl GmbH, 2016a). We did not include the pictorial level of the videos. Only the content regarding the text level was documented. The user comments related to the audiovisual material were extracted from YouTube with the help of the *YouTube Data Tool* (Rieder, 2015) on 12th of July 2016. The resulting material consisted of eight data sets— one for each video—of all user comments, including the user names, comment counts, "like"

counts, reply counts as well as network data indicating the relationships between the commenting users.

Sampling of user comments.

For all eight videos, we randomly selected 155 user comments out of 5798 comments in total. The sampling procedure is based on the *profile-sampling* method developed by Reinders (2012) for qualitative research designs. Reinders' approach allows to qualitatively inquire a huge corpus of data material and at the same time considering its diversity and complexity. The basic idea of this method is to reduce the material by organizing the data in specific formal clusters (in the presents study we relied on the comment count per user as formal criterion[10])—so-called *profiles* (Reinders, 2012)—and to (randomly) select the final material or the analysis based on these clusters. In order to apply this technique to the material (i.e., user comments related to the videos) and to specify certain profiles, quantitative information regarding the material are necessary (Reinders, 2012). In the present study, we used quantitative data about the number of comments by each user that have been generated with the *YouTube Data Tool* (Rieder, 2015). Via comment counts per user we constituted three formal profiles: Profile I comprises users with one to two comments, Profile II consists of users with up to nine comments and Profile III contains users with more than nine comments.

The profiles were applied on each of the eight data sets. Each user matched to one of the three profiles with regard to his or her number of comments. In a first step, for each of the eight videos, three users, who commented on the respective video, were randomly selected for each profile. The limitation of three users for each profile is a formal criterion, which was chosen to maintain a reduction of the material's volume. In a second step, we randomly selected a maximum of three comments for each selected user of each profile[11]. In case of Profile I the amount of the randomly selected comments corresponds to the total count of comments published by the particular users. By considering the profiles as formal clusters in

---

[10] Another possibility could be to cluster the data based on other user characteristics such as age or gender, or characteristics of the comments themselves (e.g., length, number of words). However, based on the data that we collected, the number of comments per user seemed to be the most appropriate criterion.

[11] The limitation of maximal three comments per user is also a formal criterion to keep the volume of the amount of comments small.

Julian Ernst et al.: A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos

the context of the random selection of the material, we take a certain *user characteristic* into account (comments count per user: "talkatives" vs. "medium publishers" vs. "restrained commenters"). This characteristic, in turn, may be regarded as one important source of the diversity and complexity of the data corpus.

Table 1 gives an overview about the videos, the total number of comments as well as the number of randomly selected comments. Moreover, information regarding the general outreach of the video (dislikes, likes, number of views) can be found.

Julian Ernst et al.: A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos

Table 1. Overview about the eight videos, their publication date, their outreach (like and dislike count, number of views) as well as the

total number of comments and the number of sample comments

Video	Publication date	Dislikes	Likes	Number of views	Number of sample comments	Total number of comments (12.7.2016)
Info Islam: What does Dschahiliyya mean? #whatIS	28 October 2015	243	2.884	64.405	21	1664
Info Islam: What does CALIPHATE mean? #whatIS	12 November 2015	285	10.898	131.406	22	1288
Islam and Knowledge #whatIS	16 January 2016	310	7.402	85.744	22	1240
Info Islam: What does Jihad mean #whatIS	11 December 2015	842	1.460	23.179	21	539
What does UMMA mean? #whatIS	12 October 2015	221	4.852	56.129	20	525
Info Islam: What does halal/haram mean? #whatIS	10 January 2016	75	1.874	23.823	16	257

Info Islam: What does Bid'a mean? #whatIS	27 November 2015	33	345	5.977	21	159
Infos Islam: What does "territory of war" mean? #whatIS	19 December 2015	64	640	8.245	12	126
total (*N*)					155	5798
total (%)					2.673	100

Data analysis

Madden et al. (2013) emphasize the heterogeneous shape of user comments on media content. This led to a qualitative approach focusing mainly on inductively generated categories as the specifics of user comments are simply not anticipatable on the basis of research literature on the subject so far. Thus, to answer the research questions, we applied *qualitative* content analysis based on Mayring (2010) to both the video material as well as the user comments related to the material. This method shares many basic assumptions and steps of *quantitative* content analysis: it uses a pattern of categories including specific, comprehensible and reliable rules and sequences for the categorization of data. At the same time, it distances itself from a pure deductive logic by explicitly including inductive mechanisms, which qualifies the tool for the examination of data material one cannot make a statement about considering its specifics before. Qualitative content analysis does neither try to falsify hypotheses nor does it rely on a fixed category system set up before. It rather tries to *understand* the material implying to generate hypotheses about the material for eventual subsequent quantitative oriented research on the same data (Mayring, 2010).

In the present research, we used two types of qualitative content analysis: *summary of content* and the *structuring analysis of content* (Mayring, 2010). In a first step, the *summary of content* is applied to the transcripts of the videos. It reduces the material to its core content by paraphrasing und generalizing the data (see Table 2 and Table 3).

Table 2. Results of the qualitative analysis of the video content; similar topics are marked with the same color.

Info Islam: What does Dschahiliyya mean? #whatIS	Info Islam: What does CALIPHATE mean? #whatIS	Islam and Knowledge #whatIS	Info Islam: What does Jihad mean #whatIS	What does UMMA mean? #whatIS	Info Islam: Was does halal/haram mean? #whatIS	Info Islam: What does Bid'a mean? #whatIS	Infos Islam: What does "territory of war" mean? #whatIS
The Islamic State	Territorial dominion and competences of the Kalif		Dschihad and war				Classification of territory and extremist groups
Comments and infobox	Comments and questions	Comments and video description	Comments and video description	Statements in comments		Comments and opinions	Questions and comments
Pre-Islamic traditions	The Muslim community today and the community of Medina in the seventh century		Historical lines of Dschihad	Questions about Islam	Classification of acts in Islam from a religious-legal perspective	Historical references of Bidà	Historical development of the territorial dominion
Meaning of Dschahiliyya	Ideas of Caliphate		Ways of Dschihad	Definitions of Umma	Other uses of halal und haram	Differenz understandings of Bidà	Definitions of territorial dominions
Knowledge and		Knowledge and		Ways of Islam	Other meanings		Opinions on

information about the topic Islam		opinions about the topic Islam		of halal und haram		territorial dominions
		Concepts of Knowledge		Obligated acts		Religious freedom today
Usage of Dschahiliyya today **Traqnsfer of Dschahiliyya nowaday conditions**	Caliphate in the media	Opinions on the relation of Islam and Knowledge	Term Dschihad in recent discussion	Permitted acts	Discussions about Islam	Discussions about Islam
	Caliphate and controversy about succession	Madrassas as places of Islamic knowledge transfer Knowledge and transfer of knowledge in Islam		Responsibility for acts	Other videos and YouTubers	
		Contribution to society		Self-determination and the own behaviour		
		Muslim scientists and experts		Non-permitted and undesirable acts Recommended acts		
				Dispensation of acts		

Table 3. Overview of the main topics of the videos based on the summary of content. Colors refer on the colors in Table 2 (c.f. Table 2).

Main topics	A	B	C	D	E	F	G
	Territory, war and violence	Comments, questions, video descriptions	Historical references	Diverse definitions of concepts, terms etc.	Knowledge, transfer of knowledge and Islam	Acts and their permission in Islam	Discussions about Islam

In a second step, we conducted a *structuring content analysis* of the user comments (RQ1). The summary of the video served to set up a first framework of categories to approach the material particularly with regard to RQ2. During the categorization of the comments, the categories were inductively expanded (Mayring, 2010), modified and if necessary discarded—in close relation to the analyzed material. We used the software *f4analyse* (dr. dresing & pehl GmbH, 2016b) to set up the category system and analysis of the comments. For each category a definition was determined, as well as an explication of the coding unit, context unit and evaluation unit (cf. Table A in the Appendix).

In order to address the problem of the 'subjectivity of the coding process' two researchers conducted both content analyses. As recommended by Elo and colleagues (2014), one researcher was responsible for the analyses and the other carefully followed up on the categorization and coding process. Divergent opinions were continuously discussed.

Results

Which themes and topics are brought up in the comments?

Based on our sample, we identified 48 topics (categories) users discussed in the comment section. Eight categories were found to be applicable to the comments at least seven times and can therefore be considered as especially relevant. Further, another eight categories could have been applied four or more times on different comments. The remaining 32 categories are each related to three or less user comments. In the following we will focus on the eight most important categories we identified in the user comments: (1) devaluating prejudices and stereotypes towards Muslims and/or Islam, (2) references to the experts of the German Federal Agency of Civic Education, (3) comparison of Islam and Christianity, (4) conspiracy theories, (5) religion and faith in general, (6) Islam and Islamic State, (7) Quran, (8) YouTubers. These topics will be exemplified in Table 4 as well as explained briefly thereafter.

Table 4. The eight most relevant categories and examples for each category.

Category and definition	Examples[12]
Devaluating prejudices and stereotypes towards Muslims and/or Islam **Definition: Statements containing negative prejudices and stereotypes towards Muslims and/or Islam as well as statements, which thematize prejudices and stereotypes towards Muslims and/or Islam.**	*Example 1:* + USER[13] isn't there a chance in Germany being a Muslim without being related to all those bad things? *Example 2:* I don't have anything against Muslims but …" 3, 2, 1 GO! *Example 3:* At these paper prints there are just friendly Muslims. The female genital mutilating beating slaver Sultans have been overlooked. There really are some of those. Beside that very informative.
References to the experts of the German Federal Agency of Civic Education **Definition: Utterances/statements that relate to statements, contributions of or directly to the user account ex*perts of bpb* respectively the institution itself.**	*Example 1:* +*experts for bpb* The beginning was enough. Quickly turned off. Was a really cool feeling to say STOP. *Example 2:* + *experts for bpb* Did you really just say "expert academics"? Are you serious about that? *Example 3:* +USER The Federal Agency for Imagination provides enlightenment.

---

[12] The comments are originally written and published in German language. Examples used in the text are translated in English language. Typical stylistic elements have been tried to reproduce close to the starting material.

[13] The sign „+" marks the intentionally made reference to another user, which continuously anonymized in the study with "USER". It also indicates the integration of the comment in a discourse.

Comparison of Islam and Christianity *Definition:* **Statements, which contrast/compare Christianity and Islam.**	*Example 1: + USER* towards my opinion Islam is much more advanced than Christianity. That is obvious because the Islam was founded about 600 years later *Example 2:+ USER* in no way I wanted to defend Christianity (cause I consider any religion as wrong), it wasn`t my intention to make my comment look like this. I actually wanted to continue and expand your thought to not just criticizing Islam.
Conspiracy theories *Definition:* **Statements which relate to conspiracy theories if they contain threat scenarios, hermetical closed approaches for the explanation of different phenomena, allegations of future developments, happenings etc. and insinuations and suspicions towards (imagined) groups**	*Example 1: + USER* if you declare lies to truth it worked really well in your case. Strange, that all terrorists always die and you can blame Islam, cause they can`t say a word anymore. Nobody is interested in what you think *Example 2:* Thanks, thanks, thanks for beginning with Adam and Eve for the billionth time. And all that sponsored by taxes and Saudi-Arabia? Yes, we're all confused. Fine these people are untangling. *Example 3: + USER* we can ensure, that you cannot „buy" people. Still we don't know where the accusation of being a "state's propaganda agency" comes from.
Religion and faith in general *Definition:* **Statements referring to religion/faith in general, that are not explicitly related to Islam or Christianity.**	*Example 1: + USER* Exactly, religion was a red thread for life once upon the time. Like a constitution. If you harm someone, bad things will happen to you. On the other side if you are friendly and share, others share with you too. That alone was the principle of religion. *Example 2: + USER* ok, I got that but why are they supposed to do it?? Why can't people belonging to different religions live normally with each other?!! Is it so hard to live and work with another?!! That doesn't make any sense for me … as long as humanity is not united, humans keep staying nothing but animals to me!!!! But we are an animal questioning everything and not doing what nature might prescribe … *Example 3: + USER* you seem to be a believer too. You truly seem to have faith, you could convince or dissuade somebody from his or her religion with this way of discussion.

Islam and Islamic State ***Definition:*** **Statements with regard to the relation between Islam and the so called Islamic State, for example regarding to the readiness to use violence.**	*Example 1:* I think the IS does not belong to Islam (just my personal opinion). As far as I know is charity the highest law in Islam and killing the worst sin. Therefore I`m asking, if a Community which violates both can be part of Islam?
Quran ***Definition:*** **Statements referring to the Quran, Suras, lines respectively words that can be found in the Quran.**	*Example 1:* The Schahada:" I confess, that there is no god but the one and I confess, that Muhammad is his prophet." Would be absurd to claim, that there is a dispute, if the Quran was really handed down by god. There is no difference between Islam, that claims, the Koran is 100% sent by God, and the Bible, that contains narratives about God. *Example 2:* First of all I got a doubt. If the Koran really is God's will and an evidence for God, it is not kind nor good. QUOTE OF A (WEBSITE ABOUT QURAN Don't be afraid, normally I do not hang around on those websites, just wanted to use it for quotation.) And nevertheless if that is the true word of Allah, I'm sorry. These Verses and others make it impossible for me to believe in Allah, let alone worship. *Example 3: + USER* Ha, weird. Did you read it in Arabic or a translation? If it was a translation, which one? Do you also have direct quotations from the Quran? Your case sounds more like: "Hey, I've read what was said in the internet" instead of "I've read the Quran with an interpretation, without prejudices and preconceived opinion.

References to YouTubers	*Example 1:* Well done, LeFloid[14].
Definition: **Statements with references to "YouTubers", more explicitly, persons or accounts operating a channel and producing video footage for YouTube.**	*Example 2:* I think, you've done it well, I really liked your delivery I personally prefer a face to the voice I'm listening to just got two points of criticism: your computer display was visible and additionally I think it's difficult if you change your position with all those cuts just stay at one spot. *Example 3:* I understand, what you're trying to achieve with these informative videos. It is admirable but I hope you won't gloss over and please tell the people about Christianity and the Bible as well respectively what is said in it and what law would be appropriate according the Bible. Most people don't know anything about this topic as well.

---

[14] „LeFloid" is the pseudonymous of a well-known German YouTuber.

The category *devaluating prejudices and stereotypes towards Muslims and/or Islam* dominated the data. It was assigned to comments, which include statements containing negative prejudices and stereotypes towards Muslims and/or Islam, such as for instance calling Muslims "female genital mutilating beating slaver Sultans". Furthermore, the category was assigned to statements, which thematize prejudices and stereotypes towards Muslims and/or Islam. One user for example brought up that Muslims in Germany are often associated with bad news. In another comment a common pattern of expressing prejudices was ironically imitated: first explicitly dissociating from prejudices and subsequently attaching one (see Table 4).

A further—very present—category is the *reference to the experts of the Federal Agency of Civic Education*. The category has been defined as utterances/statements that relate to statements, contributions of, or directly to the German Federal Agency of Civic Education (bpb). Some users referred to the bpb itself, others directed their comments straight to its user account *experts for bpb* via the usage of YouTube's interactive feature for direct addressing ("+USER"). By means of this account the bpb published comments. Moreover, the bpb's experts were supposed to moderate the user discussions (Bundeszentrale für politische Bildung, 2015b).

Comparing Islam and Christianity also turned out as an important topic discussed by the users of the videos. The resulting category was assigned to user comments that contrasted Islam and Christianity, for example in order to provide another object of criticism beside Islam.

Conspiracy theories respectively statements arguing in a conspiratorial way have been frequently found. The category was constructed in order to characterize statements that refer to conspiracy theories and/or scenarios. We considered statements as related to conspiracy theories if they contained threat scenarios, simplistic approaches for the explanation of different phenomena, allegations of future developments, happenings etc. and insinuations as well as suspicions towards (imagined) groups[15]. Some comments, for example, included allusions such as the relation between the death of Islamist terrorists and the accusation of

---

[15] Despite the complex and multilayered definition of the category, the identification of conspiracy theories respectively their fragments in user comments is difficult to handle with a content analytical approach. Conspiracy theoretical lines of reasoning use implications and require (with reservations) background information to be taken into account (e.g. Hepfer 2016, p. 59, Shooman 2014, p. 44).

"the Islam" in context of terrorism or the idea of Saudi-Arabia sponsoring the bpb's campaign Concepts of Islam. Comments that only referred to conspiracy theories, by responding but not promoting them, have been included as well (for an example, see Table 4).

Besides references to Islam and Christianity, the topic of *religion and faith* was also brought up in the comments in a more general way. Comments have been assigned within this category if they contained statements referring to religion and faith in general that are not explicitly related to Islam or Christianity. Those comments deal for example with individual approaches to religion and faith, the changing role of religion in history, or general issues of conflicts related to religious orientations.

In our sample we further identified the topic *Islam and Islamic State*. We assigned this category to user comments if they referred to the so called Islamic State (IS, ISIS or ISIL) and its relation to Islam. More concretely, the category has been defined as statements with regard to the relation between Islam, and the so called Islamic State, for example regarding the readiness to use violence.

Moreover, references to the *Quran* were important topic for the recipients of the videos. This category contains comments including statements referring to the Quran, Suras, lines respectively words that can be found in the Quran. In one comment, for example, the *Schahada*, the Islamic confession to faith, was explicitly recited. Other comments provided sources about the Quran or deal with users' positions towards the Quran respectively their interpretation of parts of it (see Table 4).

Several comments need to be pointed out because of their *references to the YouTubers*, more explicitly, persons or accounts operating a channel and producing video footage for YouTube. Those comments mostly included thanks and praises directed at the YouTubers themselves, suggestions and advices for the improvement of the videos on a technical level or regarding the content.

Although not equally important with regard to absolute mentions of the category, a category was identified that has a specific importance when investigating hate speech and counter speech in user comments: *pejorative language*. We assigned the category to comments, which included statements containing pejorative language and/or insults. Pejorative language is not easily identifiable as such because the attribute requires the reference to a person valuing the act of speech as pejorative. Therefor those comments have

been tagged within the category, when we intuitively, in other words based on our personal knowledge about linguistic and social conventions, considered it as pejorative or insulting.

Conclusion

The eight most relevant categories pointed to three focal points in the comments. User comments bringing up themes related to religion predominated the field. Islam, in relation to specific aspects such as prejudices, the so-called Islamic State, Christianity and the Quran, turned out to be a main topic in the analyzed comments, but also religion and faith in general, to wide perspectives on faith independently from a specific religious orientation. References to the bpb and YouTubers emerged as another focus next to statements that refer to conspiracy theories and/or scenarios. Although the category was only assigned to a few comments, pejorative language related to certain persons or groups should additionally be mentioned due to its significant importance for the analysis of hate speech (c.f. results regarding RQ3).

Which aspects of the video content do users refer to in their comments?

To answer Research Question 2 we compared the topics that have been brought up by the users with the themes covered by the videos. We used the results of the *summary of content* to discover intersections or thematic deviations. Table 2 depicts the results of the summary, Table 3 contains the thematic bundling of Table 2. For each video, several themes were discovered (Table 2). Besides a few isolated themes, for example "Muslim scientists and experts" in the video "Islam and Knowledge #whatIS", most summarized topics could be connected to seven main topics: (A) *territory, war and violence,* (B) *comments, questions, video descriptions,* (C) *historical references,* (D) *diverse definitions of concepts, terms etc.*, (E) *knowledge, transfer of knowledge and Islam,* (F) *acts and their permission in Islam and* (G) *discussions about Islam* (cf. Table 3). Each thematic bundling is marked with a letter as a shortcut (A to G). In Table 5 each of the eight most relevant categories were compared to the respective video content and classified as "content" or "non-content", i.e. as relating to video content or not. Additionally, Table 5 gives details about the categories, classified as *content*, and their relations to one or more thematic packages. For instance, the category *devaluating prejudices and stereotypes towards Muslims and/or Islam* matches shortcut E, representing

the bundle *knowledge, transfer of knowledge and Islam*, because it comprehends lacks of knowledge about Islam/Muslims and resulting prejudices. For every allocation, we give short explanations in the column next to the category like this. Except for two of the frequently brought up topics, all were somehow related to general topics presented in the videos. The two categories *Comparison of Christianity and Islam* and *Pejorative language* were neither directly nor indirectly proposed through the videos.

Table 5. Relations of *the e*ight most relevant categories categories to the video content (tagged with content in Table 5). The relations are coded with "content", i.e. the category relates to a topic listed in Table 3, and "non content", i.e. the category does not relate to a topic listed in Table 3.

Category	Reference to main topics of videos (Table 3)	Explanation of the reference	Content or non-content
Devaluating prejudices and stereotypes towards Muslims/Islam	E	Thematization of lacks of knowledge about Muslims/Islam and resulting prejudices and stereotypes	Content
Reference to Experts of Federal Agency of Civic Education (bpb)	B	Mention of user "experts for BpB", which is supposed to receive questions concerning video`s content etc.	Content
Comparison of Christianity and Islam	-	-	Non-content
Conspiracy theories	E G	Mention of discussions about unrealistic threat scenarios as well as ignorance towards and lack of knowledge about Muslims/Islam	Content
Religion/Faith in general	C E	Mention of religion`s/faith`s role in former societies, general aspects of knowledge, recognition etc. in context of religion/faith.	Content
Islam and Islamic	A	Reference to the so called	Content

Julian Ernst et al.: A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos

State		Islamic State (IS), to its ideology and acts	
Koran	C D E F	Reference to Koran in various ways: historical classification of Verses, Concepts defined in the Koran and their interpretations, lack of knowledge about the Koran`s content, definition of permissions in the Koran etc.	Content
Reference to YouTuber	B G	Mention of YouTubers sharing the videos via their channels or participating in the discussion via comments, discussions on YouTube	Content

Note: **The second column presents the specific reference coded with the letter, each main topic was assigned with in Table 2. The third column explains the allocations. Categories with more than one letter relate to various main topics.**

Which themes and topics in user comments are indicators of hate speech beneath counter speech videos?

For the identification of comments that potentially contain hate speech (RQ3) we deduced relevant categories especially from characteristics proposed by the definitions of Meibauer (2013) as well as Erjavec und Poler-Kovačič (2013). In conclusion, hate speech may be characterized as follows: a hate speech statement has to contain (a) specific connection of hate to an (imagined) group, e.g. Muslims but also journalists or media companies (Erjavec & Poler Kovačič, 2013;Amadeu Antonio Stifung, 2013), (b) the usage of pejorative language with regard to a certain (out-)group (Meibauer, 2013), for example directing at Muslims/Islam (jugendschutz.net, 2015), (c) the presence of conspiracy theories (Amadeu Antonio Stifung, 2013; Hepfer, 2016; Shooman, 2014). Based on this we determined (a) categories including *references to groups* (i.e., *devaluating prejudices and stereotypes towards Muslims and/or Islam*, *comparison of Christianity and Islam*, *references to the experts of the bpb*, *references to YouTubers*), (b) *pejorative language* and (c) *conspiracy theories* as indicators for hate speech. As hate speech is a diverse and hardly tangible phenomenon (Meibauer, 2013) the usage of multi-layered coding-patterns seem to be

Julian Ernst et al.: A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos

necessary to identify hate speech. Therefore, we classified comments indicating hate speech if they contained *combinations* of the deduced categories (at least a combination of two categories).

Four combinations have been found in the user comments, indicating hate speech in the comment itself or pointing to hate speech in the context of the discourse the comment is embedded in. In our sample, we identified four relevant comments which suited the defined patterns.

Table 6. Categories related to hate speech. Table shows the intersections of comments indicating hate speech with relevant categories.

Comments	Categories Devaluating prejudices and stereotypes towards Muslims and/or Islam	Comparison of Christianity and Islam	Conspiracy theories	Pejorative language	References to the experts of bpb
+ *USER* you mean the knowledge that gay people, lesbian people, and people of different faith in all 56 Islamic dominated countries had to gain and still have to. Right. That is really a more than obvious argument that the Islam indeed is a fascist, suppressing, bloodthirsty and antidemocratic state ideology.	X		X		
Concerning your Christianity just think of the crusaders you idiot!!!		X	X		
+*experts for bpb* „racist terms like "Muselmen"" What kind				X	X

of stupid and false statement is that?

+*experts for bpb* 270.000.000 dead cause of jihad, apparently not related to the islam. More than 20.000 terror attacks since 9/11, the attackers yelling "allahu akbar" before the deed, of course not linked to the islam. The "islamic state" of course not at all linked to the islam. the koran of IS of course is not linked to the islam. suppressed, disenfranchised and veiled women are not linked to the islam at all. same thing with the hate against gays and jews and kuffar in general not linked to the islam at all. the scharia demanded everywhere where the islam makes up a certain part of the population (mostly from 7,5%-10%) is not linked to the islam at all. not accepting the free democratic basic order is not linked to the islam at all. butchering animals is not linked to the islam at all. riots and dead if somebody dares to caricature your from sexually transmitted disease suffering and pedophile jumping jack is of course not related to the islam. There is no religion that obviously supposed to not be related to itself, it is an impudence what kind of taqiyya is	X		X	X	X

disseminated here. And if your sick ideology is criticized
for what it is, then you are crying and yelling hate and
racism. As if a sick ideology would be a race. But this is
what happens if you inbreed for centuries and prohibit
science as haram: naive, towards the middle ages oriented
jihadis, wanting to islamize (rule) the whole *world*

(1) Devaluating prejudices and stereotypes towards Muslims and/or Islam and conspiracy theories.

The first displayed comment contains two aspects: proposed prejudices and stereotypes towards Muslims and/or Islam as well as fragments of conspiracy theories. The comment designates "the Islam" as a "fascist, suppressing, bloodthirsty, and antidemocratic" ideology purchased by state. According to the user, the Islam as a "state ideology" discriminates groups of people, "gay […], lesbian, and people of different faith", in those countries it rules. All these mentioned attributes and descriptions, unfolded and reframed strongly in the comment, refer to negative stereotypes towards Islam as a collective body. Hand in hand with accusing the "Islam" being an "ideology" for a "state" the comment develops a line of reasoning close to conspiracy theories. The discrimination of "gay people, lesbian people and people of different faith" are traded as an "obvious argument" for the supposed "fascism" and other apparent attitudes of "the Islam". The recent example does not use a logical based approach of falsification but a hermetical closed line of reasoning. The comment is structured by a straight habitus of verification, trying to shape Islam as a "state ideology", for example by picking up another user's proposition. This has been shown to be a typical pattern in conspiracy theoretical argumentations (Hepfer, 2016). Along with anti-Muslim prejudices and stereotypes this may be a hint for the existence of hate speech.

(2) Comparison of Christianity and Islam and pejorative language.

The following comment fits the pattern of relation to a group and pejorative language. Thus, the comment indicates the presence of hate speech. The phrase "your Christianity" signalizes that the comments' publisher assumed a connection between an addressed person and the group of Christians. The pejorative "you idiot" devaluated the addressed person with an insult. The label "crusader"—may be assumed to be a stereotypical designation for Christians threatening Muslims—points to the Christian crusades (religious and economic motivated wars against Muslim states) in the mediaeval times, maybe even to political rhetoric during the "War on Terror". It can be assumed as an argument for the devaluation of a person associated with Christianity; at the same time it implies Christian people threatening persons with a Muslim background. Based on the theoretical rationale, the conjunction of the group reference (Christianity), addressing a particular person as a part of a certain group

("*your* Christianity"), and the insult ("*you* idiot"), indicates hate speech in the comment and the discourse this comment is embedded in.

(3) References to the experts of the federal agency of civic education (bpb) and pejorative language.

The following comment starts with a reference to the user account *experts for bpb* which is—according to the commenter—supposed to contain pejorative devaluation of Muslims marked as such trough quotation marks: "Muselmen". The question following the quote evaluated the putatively quoted statement of *experts for bpb* as "stupid and false". If we consequently follow Judith Butler's definition of hate speech (2013) the designation "Muselmen" itself can be seen as transferring hate[16]. Therefore, the present comment itself would promote hate speech. Nevertheless, it has to be kept in mind that although the comment is pejorative and rude, it is not *explicitly addressing* Muslims as "Muselmen" as the quotation signalizes a reference to another user's comment and, thus, a certain distance to the term. Therefore, the comment itself may not necessarily classified as an act of hate speech. Still, the quotation referring to another comment in the user discussion indicated the possible presence of hate speech in the discourse the comment is embedded in.

(4) Prejudices and stereotypes towards Muslims and/or Islam, reference to the experts of the bpb, pejorative language and conspiracy theories.

The following comment contains an intersection of different categories. The direction to the account *experts for bpb* is followed by an enumeration of prejudices and stereotypes towards Muslims: among others the image of "suppressed […] and veiled women" has been promoted. The list is structured as sequence of negations added to every single apparent "facet" about "the Islam". The end of the sequence is marked by extreme insults of the religion. The personal pronouns "you" and "your" do not distinctively provide information about a specific addressee. Nevertheless, the usage of those pronouns signalizes that the comment's publisher points to a group of persons, probably the *experts for bpb* addressed as Muslims or Muslims in general.

---

[16] This classification concerns the German-speaking world (Acke, 2010).

Fragments of conspiracy theories implicitly become apparent in the comment by designating the Islam as "ideology". This term reduces Islam as religion to pure policy and corresponds to the assumption of Muslims as a collective body trying "to islamize (rule) the whole world". Explicitly the threat scenario of a "hidden danger" going out from "the Islam" and the apparent phenomenon of "taqiyya" (Shooman, 2014) is invoked, too. Furthermore, some of the author's propositions are complemented by numerical data such as the following: "mostly from 7,5%-10%", "270.000.000 dead cause of jihad" or "20.000 terror attacks since 9/11"—presumably to strengthen the argumentation.

To sum up, the conspiracy theoretical line of reasoning, manifested among others in the threat scenario of taqiyya, in combination with the clear devaluation of Muslims through the use of stereotypes and harsh insults strongly indicated the presence of hate speech in the comment.

Conclusion

Four comments corresponding to four combinations of categories could have been identified among the analyzed user comments as indicating hate speech. The indication varied in its intensity. The comments assigned to Pattern 1 and Pattern 4 provided strong indications for hate speech by explicitly insulting Muslims as a collective, or the accusation of conspiracy. Whereas the comments corresponding with Pattern 3 und Pattern 2 do not necessarily indicate the presence of hate speech *in the comments itself* but the *discourse they are embedded in*.

Discussion

With a qualitative approach, the present study aimed at exploring a randomly selected sample of user comments beneath counter speech videos of the bpb published on YouTube. With regard to Research Question 1, we shed light on the relevance of specific topics in the user comments. Based on our analysis, three focal points in the comments can be emphasized: (1) themes related to religion, (2) references to the experts of the bpb and YouTubers as well as to (3) conspiracy theories. Regarding the first aspect (1), we identified a high presence (more than seven assignments) of anti-Muslim prejudices and stereotypes in the analyzed

Julian Ernst et al.: A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos

comments. This is a quite relevant finding as the video campaign Concepts of Islam explicitly claims to counteract anti-Muslim prejudices and stereotypical disputes (Bundeszentrale für politische Bildung, 2015a). Based on this result, it seems plausible to ask whether the explicit naming of prejudices and stereotypes in counter speech videos lead to their reproduction in content related user comments. First, research on the *sleeper effect* (Hovland & Weiss, 1951) showed that attitudes associated with a message, for instance low credibility of a source or adverse cues (e.g., extremism, violence), were remembered less after a while, while content memorization of the message itself remained stable (Kumkale & Albarracín, 2004). Consequently, a message content with low credibility or put into a very specific context could gain persuasiveness just because important contextualizing information is forgotten over the time. The same could hold true for naming prejudices in the context of counter speech videos: The context in which these topics were named could become recalled less while the topic itself would still be remembered. Second, this effect could be amplified for the negative topics within the counter messages since negative information is more likely to be recognized, both in terms of content as well as in terms of the source (Robinson-Riegler & Winton, 1996). Third, Butler pointed out the performativity of speech acts. Even a citation of hate speech intended to be shaped in a critical way, is capable not just to reproduce the hateful content but also to solidify it (Butler, 2013).

Moreover, concerning the second point (2), we found a large number of references to the user account *experts for bpb*. The account, announced in the videos as an instance of moderation and source for answering questions, seems to be a special contact for users. While having a closer look on several comments tagged with this category—especially passages found to be ironic, or questioning the account's status of being experts—are striking. The relevant number of references might be connected to the promotion of the bpb—in particular the account *experts for bpb*—in the videos. The account might have been addressed simply because users knew its name from the video content. The tendency of a disrespectful style in the comments might be explained by the account's appearance as representative of a public institution. *Experts for bpb* might not be identified as a single individual but a collective body because of the plural form (expert*s*) and the clear reference to the institution in the account's name (*bpb*). The inhibition threshold of being ironic and disrespectful might be reduced by this de-individualized characteristic.

The third relevant aspect (3) in the comments are the references to conspiracy theories. The great relevance might be traced back to the implicit naming of conspiracy theoretical lines of reasoning. For example in the video "Islam and Knowledge #whatIS", which emphasizes acquisition, transition, quality and reliability of knowledge in Islam and about Islam. But conspiracaly lines of reasoning also appear among comments related to other videos of the campaign. Maybe the topic of Islam in general evokes not just prejudices and stereotypes but also the necessity to legitimize these prejudices. For instance, the frequently proposed claim that Islam is a religion, which inherently intends to threaten people of other faith, needs an argumentative starting point: e.g. the concept of taqiyya (Shooman, 2014). Conspiracy theories like the taqiyya allegation can fill this gap of justification, and these are possibly attractive for this purpose because of their simple reproducible principle of verification (Amadeu Antonio Stiftung, 2013). At the same time, they may easily spread—more or less subtle—anti-Muslim beliefs, which has been shown to be particularly problematic in networks that are socially homogenous (Vicario et al., 2016).

In Research Questions 2 we asked how the topics are discussed within the user comments related to video content. We showed that users in a large part discussed the content of the video. This hint at the great importance of interpersonal discussion: on the one hand, as a means of online political participation (e.g., Vitak et al., 2010) and, on the other hand, for the often-suggested two-step flow of communication (e.g., Eveland, 2004; Eveland & Schmitt, 2015; Katz, 1957). Both aspects may be considered as intended by the bpb as publisher of the videos (Bundeszentrale für politische Bildung, 2015a). Besides the categories that clearly related to the videos' content, the analysis revealed two categories of great importance in the comments that were not connected to main topics dealt with in the videos: *comparison of Islam and Christianity* as well as the aspect of *pejorative language*. In particular, the appearance of the former category raises questions. Although Christianity was not explicitly mentioned in these videos, this category is one of the quantitatively most relevant categories with regard to the analyzed comments. The relevance of this category could be explained by the video's focus on Islam as a *religion*. Christianity might have been brought up here because it is the most common and known religion in Germany—and, thus, a relevant contrast category when considering a religion users in Germany that might be less familiar with it, and feel more threatened by: Islam. The factor that Christianity is thematized

only in comparison with Islam, which indicates the potential function of Christianity for the commenters: a reference point for the formation of an opinion. Users might use Christianity to underline differences between two religions, and also similarities or common grounds. Besides, the modality of comparison in the comments might point out a claim of an equality in criticism: Equivalently to Islam, Christianity should be a target of a criticism.

However, as we showed it in Research Question 3, this kind of intergroup comparison has also the risk to contain hateful speech. Results of social-psychological studies, which indicate that in-group favoritism may come along with out-group negativity (Bourhis & Gagnon, 2001; Mummendey & Otten, 2001; Tajfel & Turner, 1979; Turner & Reynolds, 2001) and may result in severe intergroup conflicts and hate (Halevy, Bornstein, & Sagiv, 2008; Woolf & Hulsizer, 2004), support the prevalence of a comparison between Christianity and Islam in the comments. In these comments Islam is often being criticized, whereas Christianity was mostly unaffected by this criticism. A possible explanation for the presence of this aspects of pejorative language may be related to the rarely assigned category of "thematization of respectful contact" (see Table A in the appendix), which, on the contrary, refers to the video content. The call for respect could have evoked a very opposite of its intended effect. It seems plausible to assume that pejorative language might have its origin in reactance to this explicit call for respect. In particular, due to the context in which (hateful) comments appear – the largely anonymous censorship-free environment of the WWW – pejorative language is likely to occur (Klenk, 2016).

Concerning Research Question 2 our qualitative approach proved its adequacy. The explorative way of working enabled us to consider also the topics in the various comments, which were not mentioned in the videos. Among others two categories classified as not related to the content (non-content), *comparison of Islam and Christianity* and *pejorative language,* which turned out being especially relevant for the indication of hate speech in comments in the Research Questions 3.

In Research Question 3, we asked if user comments beneath counter speech videos contain themes and topics that hint at the presence of hateful speech either in the analyzed comments or the discourse the comments are embedded in. With the diversity and complexity of the criteria, which is suggested in the literature identifying hate speech (e.g., Butler, 2013; Maibauer, 2013), we applied relatively strict standards in order to identify hints of hate speech

Julian Ernst et al.: A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos

in our sample of user comments. Comments indicated hate speech if they contained specific combinations of theoretically deduced categories. Based on this, we identified four patterns (combinations of categories) that indicate hate speech in the user comments: (1) Devaluating prejudices and stereotypes towards Muslims and/or Islam and conspiracy theories, (2) comparison of Christianity and Islam and pejorative language, (3) references to the experts of bpb and pejorative language, (4) prejudices and stereotypes towards Muslims and/or Islam, references to the experts of bpb, pejorative language and conspiracy theories.

The comments which have been matched with Pattern 1 and 4 heavily indicated hate speech through insults towards Muslims as a group and presence of conspiracy theories (Klein, 2012; Meibauer, 2013). As already mentioned in the discussion of Research Question 1, conspiracy theories might be used to legitimize the devaluation of a group. The allegation of taqiyya, which is explicitly referred to in the comment matching Pattern 4 can be regarded as a conspiracy theory by implying that Muslims might plan to hide their "true" intention in order to harm people of other faith, which at the same time, serves as a justification for further accusations and devaluations. Moreover, the reference to taqiyya indicated here that the publishing user might also visit websites which are most likely belonging to the political (far) right (Shooman, 2014).

Comments matched with Pattern 2 and 3 did not necessarily indicate hate speech in the comments, but they pointed to hate speech in the discourse they are embedded in. Further research is needed to analyze respective discourses and, thereby, concrete amount and intensity of hate speech that is included in more detail.

With regard to implications of hate speech in user comments, our results raise three additional questions: (1) For which purpose the account *experts for bpb* is addressed in such way? (2) How can experts of civic education moderate hateful comments? (3) How far might the presence of such an agent increase the motivation in order to publish hate speech or even cause to initiate it? Further examinations of discourse structure are necessary to explore the dynamics between moderators and users which are apparently promoting hatred in their comments more in detail. Although we did not find comments that question the YouTubers' position respectively his or her opinion and knowledge, it seems to be relevant, and interesting to more deeply analyze the differences between comments referring to *experts for bpb* (who are responsible for the videos as publishers) and comments addressing to

YouTubers (who are displayed in the videos as actors). The perception of YouTubers as *individuals* would strengthen the hypothesis of a de-individualized perception of *experts for bpb* which is related to the expressed pejorative language in the comments.

Limitation and future perspectives

The present study is limited in terms of giving statements about the whole data set as we analyzed only a randomized selection of user comments. Although, through the profile sampling (Reinders, 2012) the complexity of the material has been considered the sample does not allow representative remarks. For example, by basing on our relatively strict criteria for hate speech, we found that "only" four comments among the 155 selected comments pointing to hate speech. Nevertheless, it shows that—according to the mere number of comments—about three percent of comments beneath counter speech videos indicating hate speech. Further analyses of a larger selection of these materials should give an idea about the total amount of comments including hate speech relatively to comments that do not include hate speech. Moreover, they could help to identify antecedents, motivations, and consequences of the propagation of hate speech.

Still, qualitative analysis of the selected data allows us to deduce implications for further investigations of the material: qualitative and quantitative ones. The summary of the videos' main themes hint at topics in the comments. Moreover, patterns of comments indicating hate speech give important implications for a more complete survey of the material. While this specific study focused on a small sample of videos of a specific campaign, future studies could take larger datasets into account. An automated sentiment analysis, for example, may complement our present research.

Moreover, the random selection of comments neglects the discourse structure of the material. Users are not only commenting concerning the video content, but also referring to other users' comments (Dynel, 2014). As already mentioned above, future qualitative studies could focus on specific discourses between users. For example, they could be regarding the aspect of hate speech in the context of interpersonal discussions or the impact of the moderating role of the user account experts for bpb. To unfold concrete interactions containing hate speech could provide specific knowledge about the dynamics which are

Julian Ernst et al.: A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos

limiting or encouraging hate speech. A more "sensitive" tool such as discourse analysis (Diaz-Bone, 2005), conversation analysis (Deppermann, 2008) or the documentary method (Bohnsack, 2014) might be appropriate to expatiate implications in phrases, and to analyze turn taking between different users.

Furthermore, it is unclear *why* there are so many comments promoting anti-Muslim prejudices and stereotypes and even hate speech. Are counter messages systematically targeted by extremists or at least sympathizers? Or do they somehow attract extremist sympathizers? Focusing on the publishers of hate comment could be an interesting approach to detect, whether and how fare, for instance, the presence of anti-Muslim stereotypes and conspiracy theories in the comments is related to organized extremist propaganda activities. A network analysis, reconstructing the interactivity of users beneath different videos, could be a suitable approach to shed light on this question.

The unconsidered stays at the visual level of the video and its perception in the comments. By taking into account videos that have a textual and a visual level, it might be also worthwhile to investigate comments related to visualizations in the specific videos.

In the context of our qualitative content analysis we focused only on the mere appearance of topics, themes in the videos, and themes in the comments. Research on the effects of counter message videos provides a first evidence that counter messages are able to raise awareness about the mentioned topics, and further fostering reflection of the material (Frischlich et al., in press). However, these results were obtained in laboratory studies, and through group discussions. Hate speech – and other comments beneath such videos – happen in a totally different context – namely at home, and anonymously. Since previous research demonstrated that user comments also depend on the valence of the news (von Sikorski & Hänelt, 2016), different genres of counter speech videos could be compared with the comments they evoke. Further research is needed to analyze *how* the facts and information provided by the media content are perceived and discussed in the comments: Did the users understand the content of the video correctly? How is this reflected in their comments? Do they re-frame the provided information? How does the narration use in the video impact on the user discussion? Are one-sided narrations more influential than two sided in term of their persuasive character? How do problematic user comments interfere respectively moderate the already weak effectiveness of counter messages?

Practical implications

Result of our study indicated both positive and negative consequences of counter speech video. On the one hand, we could show that people are motivated by media content in order to discuss the topics and themes that media covers. Thus, as discussed by the researchers in the field of political communication, interpersonal discussion about (political) media content may result in increasing knowledge of the presented facts, especially more (political) online and offline participation (e.g., Eveland, 2004; Eveland, 2005); all of these results are considered to be desirable from a societal perspective (Delli Carpini & Keeter, 1996).

On the other hand, presence of anti-Muslim prejudices in user comments related to counter messages—demanding to tackle those—raised further questions, to what extent the videos achieved their original goal, and how counter messages could be designed to avoid this phenomenon. In here lies the specific potential to think about the ways to foster a discourse as well as nurturing its civilized manner. One potential starting point could be the role of *trustworthy, credible* moderators in such comments. Working with these moderators could be employed and also analyzed in a more systematic manner in order to investigate whether such moderators could help in rendering hate speech and increasing the reflective and respectful nature of user comments. Although, we identified "only" four comments in our sample that hint at hate speech—according to our conceptualization—the mere presence of hateful comments may influence the perception of the video. As previous research has shown, that the valence of user comments may influence peoples' attitudes towards their evaluation of the credibility and the trustworthiness of respective media content (e.g., Lee & Jang, 2010; von Sikorski & Hänelt, 2016). Therefore, the videos aiming at countering extremist ideas might get a different connotation when comments propagate hate speech right beneath it. Moreover, it has been found that aggressive wording of commenters causes more aggressive wording by other commenters (Rösner & Krämer, 2016). Therefore, it seems plausible to assume that presence of hate speech below a video may motivate others to add hateful comments. The reading of these comments may also have severe psychological consequences for the recipients. Recipients of hate speech report negative emotions (e.g., shock, anger, bad mood, embarrassment) as short-term consequences of the perception of hate speech, on the long run

they assume lasting effects on their self-esteem and psyche as well as on the defensive attitudes against the speakers' group (Leets, 2002).

The rather positive perception of the YouTubers and the fact that the presence of a public institution as publisher of counter speech videos may result in hate speech raise the question if it would be better for the respective institution to stay in the background to avoid hate speech in connection with counter messages. Against the background of the above-mentioned assumption that the name of the account representing the institution (experts of the bpb) leads to a de-individualization of the people behind the account and, therefore, a more disrespectful interaction with it, slight changes of the account's name, or the appearance of a concrete protagonist in the videos related to the account, could potentially evoke less disrespectful comments and, more importantly, less hate speech.

However, based on our analysis also the question arises if this kind of videos are suitable to foster tolerant attitudes and inoculate against propaganda effects. Previous research on effectiveness of counter messages, which speaks rather for counter messages to unfold an indirect effect: They increase the rejection of propaganda material but not necessarily increase the reflection of the tolerant, democratic arguments (Frischlich et al., in press). It seems that counter messages work best when they are presented in critical moments, for instance, when the ideas of extremist video material are deconstructed or when people are in need for answers – be it through perceived exclusion, a high need for cognition or when having stumbled over propaganda beforehand. Therefore, in order to counteract extremist ideas, to avoid hate speech as well its effects, seems more appropriate to—at least—embed these videos in a broader pedagogical program. For comments beneath such videos, it seems important to provide context and guidance, to diminish the occurrence of hate speech and make comments appear in the right context.

Julian Ernst et al.: A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos

References

Acke, H. (2010). >Muslim_in<. Nduka-Agwu, A., & Lann Hornscheidt, A. (eds.). Rassismus auf gut Deutsch. Ein kritisches Nachschlagewerk zu rassistischen Sprachhandlungen. [Racism in plain German. A critical reference book for racist speech acts]. Frankfurt am Main: Brandes & Apsel Verlag. 326-331.

Amadeu Antonio Stiftung. (2015). "Geh sterben!". Umgang mit Hate Speech und Kommentaren im Internet. [„Go and die!". Dealings with hate speech and comments in the internet]. Retrieved July 29, 2016, from https://www.amadeu-antonio-stiftung.de/w/files/pdfs/hatespeech.pdf

Amadeu Antonio Stiftung. (2013). Viraler Hass. Rechtsextreme Kommunikationsstrategien im Web 2.0. [Viral hate. Right-wing Extremist Communication strategies]. Retrieved September 15, 2016, from https://www.amadeu-antonio-stiftung.de/w/files/pdfs/viraler-hass.pdf

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "Nasty Effect:" Online Incivility and Risk Perceptions of Emerging Technologies. *Journal of Computer-Mediated Communication*, *19*(3), 373–387. doi: 10.1111/jcc4.12009

Arbeitsgemeinschaft Kinder- und Jugendschutz (AJS) Landesstelle NRW e.V., & Landesanstalt für Medien Nordrhein-Westfahlen (LfM). (2016). Hate Speech - Hass im Netz. [Hate speech – hate in the internet]. Retrieved November 28, 2016, from http://www.ajs.nrw.de/wp-content/uploads/2016/06/160617_HateSpeech_WEB2.pdf

Ben-David, A., & Fernández, A. M. (2016). Hate Speech and Covert Discrimination on Social Media: Monitoring the Facebook Pages of Extreme-Right Political Parties in Spain. *International Journal of Communication*, *10*(0), 27.

Bohnsack, R. (2014). Rekonstruktive Sozialforschung. Einführung in qualitative Methoden. [Reconstructive social research. Introduction in qualitative methods]. (9th ed.). Opladen & Toronto: Verlag Barabara Budrich.

Bourhis, R. Y., & Gagnon, A. (2001). Social orientations in the minimal group paradigm. Brown, R. & Gaertner, S. L. (eds.), Blackwell Handbook of Social Psychology: Intergroup processes (pp. 133–152). doi: 10.1002/9780470693421.ch7.

Braddock, K., & Dillard, J. P. (2016). Meta-analytic evidence for the persuasive effect of narratives on beliefs, attitudes, intentions, and behaviors. *Communication Monographs*, 1-24. doi:10.1080/03637751.2015.1128555

Braddock, K., & Horgan, J. (2015). Towards a guide for constructing and disseminating counter-narratives to reduce support for terrorism. *Studies in Conflict & Terrorism*, 39(5), 381-404. doi:10.1080/1057610X.2015.1116277

Briggs, R., & Feve, S. (2013). Review of programs to counter narratives of violent extremism: What works and what are the implications for government? London: Institute for Strategic Dialogue. Retrieved August 24, 2016, from https://www.counterextremism.org/resources/details/id/444/review-of-programs-to-counter-narratives-of-violent-extremism-what-works-and-what-are-the-implications-for-government

Bundeszentrale für politische Bildung (2015a). Begriffswelten Islam. [Concepts of Islam]. Retrieved from http://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/213243/webvideoformate-begriffswelten-islam

Bundeszentrale für politische Bildung. (2015b). Netiquette für die Webvideoreihe "Begriffswelten Islam". [Netiquette for the video campaign Concepts of Islam]. Retrieved from http://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/214763/netiquette-fuer-die-webvideoreihe-begriffswelten-islam

Bundeszentrale für politische Bildung (2012a). Strengthening Democracy – Fostering a Civil Society. The Federal Agency for Civic Education: Our Mission and Activities. Retrieved from http://www.bpb.de/die-bpb/138853/our-mission-and-activities

Bundeszentrale für politische Bildung (2012b). How we work: Key Activities. Retrieved from http://www.bpb.de/die-bpb/138867/key-activities

Butler, J. (2013). Haß spricht: Zur Politik des Performativen. [Original title: Exitable Speech. A Politics of the Performative]. German translation of Menke, Katharina und Krist, Markus (4th ed.). *edition suhrkamp: Vol. 2414*. Frankfurt am Main: Suhrkamp.

Cammaerts, B. (2009). Radical pluralism and free speech in online public spaces. The case of North Belgian extreme right discourses. *International Journal of Cultural Studies*, *12*(6), 555–575. doi: 10.1177/1367877909342479

Dafnos, A. (2014). Narratives as a means of countering the radical right; Looking into the Trojan T-shirt project. *Journal EXIT Deutschland*, 3, 156-188.

Dharmapala, D., & McAdams, R. H. (2005). Words That Kill? An Economic Model of the Influence of Speech on Behavior (with Particular Reference to Hate Speech). *The Journal of Legal Studies*, *34*(1), 93–136. doi: 10.1086/425599

Deppermann, A. (2008). Gespräche analysieren. Eine Einführung. [Analysing conversations. An introduction]. (4th ed.). Wiesbaden: VS Verlag für Sozialwissenschaften.

Diaz-Bone, R. (2005). Diskursanalyse. [Discourse analysis]. In: Mikos, L. & Wegener, C. (Eds.): Qualitative Medienforschung. Ein Handbuch. [Qualitative research on media. A handbook]. Konstanz: UVK Verlagsgesellschaft.

dr. dresing & pehl GmbH (2016a). f4transcript. https://www.audiotranskription.de/f4.htm

dr. dresing & pehl GmbH (2016b). f4analyse. https://www.audiotranskription.de/f4-analyse

Dynel, M. (2014). Participation framework underlying YouTube interaction. *Journal of Pragmatics*, *73*, 37–52. doi: 10.1016/j.pragma.2014.04.001

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative Content Analysis: A Focus on Trustworthiness. *Sage Open*, *4*(1), 1-10. https://doi.org/10.1177/2158244014522633.

Erjavec, K., & Poler Kovačič, M. (2012). "You Don't Understand, This is a New War!" Analysis of Hate Speech in News Web Sites' Comments. *Mass Communication and Society*, *15*(6), 899–920. doi: 10.1080/15205436.2011.619679

Eveland, W. P. J. (2004). The Effect of Political Discussion in Producing Informed Citizens: The Roles of Information, Motivation, and Elaboration. *Political Communication*, *21*(2), 177–193. doi: 10.1080/10584600490443877

Eveland, W. P. J., Morey, A. C., & Hutchens, M. J. (2011). Beyond deliberation: new directions for the study of informal political conversation from a communication perspective. *Journal of Communication*, *61*(6), 1082–1103. doi: 10.1111/j.1460-2466.2011.01598.x

Eveland, W. P. J., & Schmitt, J. B. (2015). Communication Content and Knowledge Content Matters: Integrating Manipulation and Observation in Studying News and Discussion Learning Effects. *Journal of Communication*, *65*(1), 170–191. doi: 10.1111/jcom.12138

Frischlich, L., Rieger, D., Morten, A., & Bente, G. (Eds.) in cooperation with the Federal Criminal Police Office, Department Terrorism/Extremism (in press). Videos gegen Extremismus? „Counter-Narrative" auf dem Prüfstand [Videos against extremism? "Counter-Narratives" being put to test]. Luchterhand

Gastil, J., & Dillard, J. P. (1999). Increasing political sophistication through public deliberation. *Political Communication*, *16*(1), 3–23. doi: 10.1080/105846099198749

Glaser, S. (2013). Rechtsextremismus online. [Right-wing extremism online]. Mainz: Druckerei Bunter Hund.

Halevy, N., Bornstein, G., & Sagiv, L. (2008). "In-Group Love" and "Out-Group Hate" as Motives for Individual Participation in Intergroup Conflict A New Game Paradigm. *Psychological Science*, *19*(4), 405–411. doi: 10.1111/j.1467-9280.2008.02100.x

Halpern, D., & Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior*, *29*(3), 1159–1168. doi: 10.1016/j.chb.2012.10.008

Hardy, B. W., & Scheufele, D. A. (2005). Examining Differential Gains From Internet Use: Comparing the Moderating Role of Talk and Online Interactions. *Journal of Communication*, *55*(1), 71–84. doi: 10.1111/j.1460-2466.2005.tb02659.x

Hepfer, K. (2016). Verschwörungstheorien: Eine philosophische Kritik der Unvernunft [Conspiracy theories: philosophical criticism on irrationality]. (Licensed ed. for the German Federal Agency of Civic Education). *Schriftenreihe der Bundeszentrale für politische Bildung: Band 1701*. Bonn: Bundeszentrale für politische Bildung.

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, *15*(4), 635–650.

Hughey, M. W., & Daniels, J. (2013). Racist comments at online news sites: a methodological dilemma for discourse analysis. *Media, Culture & Society*, *35*(3), 332–347. doi: 10.1177/0163443712472089

Hussain, G., & Saltman, E. M. (2014). Jihad trending: A comprehensive analysis of online extremism and how to counter it. Quiliam Foundation.

jugendschutz.net (Ed.). (2015). Islamismus im Internet: Propaganda-Verstöße-Gegenstrategien. [Islamism in the internet: propaganda-violations-counter-strategies]. Mainz.

jugendschutz.net (Ed.). (2015). Rechtsextremismus online: beobachten und nachhaltig bekämpfen. Bericht über Recherchen und Maßnahmen im Jahr 2014. [Right-wing extremism online: observing and sustainable fighting. Report on research and acts in the year 2014]. Mainz.

Katz, E., & Lazarsfeld, P. F. (1955): Personal Influence: The Part Played by People in the Flow of Mass Communications. The Free Press, New York.

Katz, E. (1957). The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis. *Public Opinion Quarterly*, *21*(1), 61–78. doi: 10.1086/266687

Kim, Y. (2015). Exploring the Effects of Source Credibility and Others' Comments on Online News Evaluation. *Electronic News*, *9*(3), 160–176. doi: 10.1177/1931243115593318

Kim, S. Y., Levine, T. R., & Allen, M. (2014). The intertwined model of reactance for resistance and persuasive boomerang. *Communication Research*. doi: 10.1177/0093650214548575

Klein, A. (2012). Slipping Racism into the Mainstream: A Theory of Information Laundering. *Communication Theory*, *22*(4), 427–448. doi: 10.1111/j.1468-2885.2012.01415.x

Klenk, F. (2016). Boris wollte mich verbrennen. [Boris wanted to burn me.]. FALTER 45/16, 8.11.2016. Retrieved 5.12.2016, from https://www.falter.at/archiv/wp/boris-wollte-mich-verbrennen

Kruikemeier, S., Van Noort, G., & Vliegenthart, R. (2016). The Effect of Website Interactivity on Political Involvement. *Journal of Media Psychology*, *28*(3), 136–147. doi: /10.1027/1864-1105/a000200

Kumkale, G. T., & Albarracín, D. (2004). The sleeper effect in persuasion: A meta-analytic review. *Psychological Bulletin*, *130*(1), 143–172.

Lee, E.-J. (2012). That's Not the Way It Is: How User-Generated Comments on the News Affect Perceived Media Bias. *Journal of Computer-Mediated Communication*, *18*(1), 32–45. doi: 10.1111/j.1083-6101.2012.01597.x

Lee, E.-J., & Jang, Y. J. (2010). What Do Others' Reactions to News on Internet Portal Sites Tell Us? Effects of Presentation Format and Readers' Need for Cognition on Reality Perception. *Communication Research*, *37*(6), 825–846. doi: 10.1177/0093650210376189

Julian Ernst et al.: A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos

Leets, L. (2002). Experiencing Hate Speech: Perceptions and Responses to Anti-Semitism and Antigay Speech. *Journal of Social Issues*, *58*(2), 341–361. doi: 10.1111/1540-4560.00264

Madden, A., Ruthven, I., McMenemy, D. (2013). A classification scheme for content analyses of YouTube video comments. *Journal of Documentation*, *69*(5), 693-714. soi: 10.1108/JD-06-2012-0078

Mayring, P. (2010). Qualitative Inhaltsanalyse: Grundlagen und Techniken. [Qualitative content analysis: basics and techniques. (11th updated and revised ed.). Weinheim: Beltz. Retrieved from http://www.content-select.com/index.php?id=bib_view&ean=9783407291424

Meibauer, J. (2013). Hassrede - von der Sprache zur Politik. [Hate speech – from language to politics]. In J. Meibauer (Ed.), Hassrede/Hate Speech. Interdisziplinäre Beiträge zu einer aktuellen Diskussion [Interdisciplinary Contributions to a recent discussion]. (pp. 1–16). Gießen: Gießener Elektronische Bibliothek. Retrieved from http://geb.uni-giessen.de/geb/volltexte/2013/9251/

Mummendey, A., & Otten, S. (2001). Aversive Discrimination. Brown, R. & Gaertner, S. L. (eds.), Blackwell handbook of social psychology: Intergroup processes (pp. 112–132).

Reinders, H. (2012). Qualitative Interviews mit Jugendlichen führen: Ein Leitfaden. [Qualitative inteviews with youths: a guideline]. (2., revised edition). *Sozialwissenschaften 10-2012*. München: Oldenbourg. Retrieved from http://www.oldenbourg-link.com/isbn/9783486715903

Rieder, B. (2015). *YTDT Video Network*. Retrieved from https://tools.digitalmethods.net/netvizz/youtube/mod_videos_net.php

Rieger, D., Frischlich, L., & Bente, G. (2013). *Propaganda 2.0. Psychological Effects of Right-Wing and Islamic Extremist Internet Videos*. Köln: Luchterhand Verlag.

Robinson-Riegler, G. L.; Winton, W. M. (1996). "The role of conscious recollection in recognition of affective material: Evidence for positive-negative asymmetry". *The Journal of General Psychology*. 123 *(2): 93–104.* doi:10.1080/00221309.1996.9921263.

Roth, H.-J. (2013). Sprache – Sprechen – Schweigen. Historische und theoretische Positionen zum Verhältnis des Sprechens über sprachliche Vielfalt. [Language – speech – silence.

Historical and theoretical positions towards the speech about lingual diversity]. Roth, H.J., Terhart, H., Anastasopoulos (eds.): Sprache und Sprechen im Kontext von Migration. Worüber man sprechen kann und worüber man (nicht) sprechen soll. [Language and speech in the context of migration. What one can talk about and what one should (not) talk about]. Wiesbaden: Springer VS. 13-42.

Rösner, L., & Krämer, N. C. (2016). Verbal Venting in the Social Web: Effects of Anonymity and Group Norms on Aggressive Language Use in Online Comments. *Social Media + Society*, *2*(3), 2056305116664220. doi: 10.1177/2056305116664220

Schabas, W. (2000). Hate speech in Rwanda: the road to genocide. *McGill Law Journal*, *46*(1), 141–171.

Shooman, Y. (2014). Antimuslimischer Rassismus und Islamfeindlichkeit im World Wide Web. [Anti-Muslim racism in the World Wide Web]. In I. Attia, A. Häusler, & Y. Shooman (Eds.), *Unrast transparent Rechter Rand: Vol. 14. Antimuslimischer Rassismus am rechten Rand*. [Anti-Muslim racism at the right margin]. (2nd ed.). Münster: Unrast-Verl.

Sikorski, C. von, & Hänelt, M. (2016). Scandal 2.0 How Valenced Reader Comments Affect Recipients' Perception of Scandalized Individuals and the Journalistic Quality of Online News. *Journalism & Mass Communication Quarterly*, *93*(3), 551–571. doi: 10.1177/1077699016628822

Stroud, N. J., van Duyn, E., & Peacock, C. (2016). News Commenters and Comment Readers. Retrieved December 1, 2016, from https://engagingnewsproject.org/enp_prod/wp-content/uploads/2016/03/ENP-News-Commenters-and-Comment-Readers1.pdf

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks/Cole

Turner, J. C., & Reynolds, K. H. (2001). The social identity perspective in intergroup relations: Theories, themes, and controversies. In Brown, R. & Gaertner, S. L. (Eds.), Blackwell handbook of social psychology: Intergroup processes. (pp. 133–152). doi: 10.1002/9780470693421.ch7.

Vicario, M. D., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., … Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559. doi: 10.1073/pnas.1517441113

Vitak, J., Zube, P., Smock, A., Carr, C. T., Ellison, N., & Lampe, C. (2010). It's Complicated: Facebook Users' Political Participation in the 2008 Election. *Cyberpsychology, Behavior, and Social Networking*, *14*(3), 107–114. doi: 10.1089/cyber.2009.0226

Weber, P. (2014). Discussions in the comments section: Factors influencing participation and interactivity in online newspapers' reader comments. *New Media & Society*, *16*(6), 941–957. doi: 10.1177/1461444813495165

Woolf, L. M., & Hulsizer, M. R. (2004). Hate Groups for Dummies: How to Build a Successful Hate-Group. *Humanity & Society*, *28*(1), 40–62. doi: 10.1177/016059760402800105

Ziegele, M. (2016). Nutzerkommentare als Anschlusskommunikation. [User comments as interpersonal communication].Wiesbaden: VS Verlag für Sozialwissenschaften. Retrieved from http://www.springer.com/de/book/9783658128210

Zick, A., Küpper, B., & Hövermann, A. (2011). Die Abwertung der Anderen: Eine europäische Zustandsbeschreibung zu Intoleranz, Vorurteilen und Diskriminierung. [Devaluation of the others: a description of state in Europe concerning intolerance, prejudices and discrimination]. Berlin: Friedrich-Ebert-Stiftung Forum Berlin. Retrieved July 22, 2016, from http://library.fes.de/pdf-files/do/07905-20110311.pdf

Julian Ernst et al.: A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos