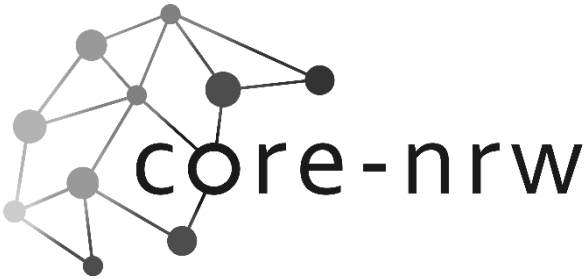
Withdrawal to the shadows: dark social media as opportunity structures for extremism

Frischlich, Lena; Schatto-Eckrodt, Tim; Völker, Julia

Veröffentlichungsversion / Published Version
Forschungsbericht / research report



Netzwerk für Extremismusforschung
in Nordrhein-Westfalen

Connecting Research on Extremism
in North Rhine-Westphalia

Withdrawal to the Shadows: Dark Social Media as Opportunity Structures for Extremism

Lena Frischlich | Tim Schatto-Eckrodt | Julia Völker

Abstract

Dark social media has been described as a home base for extremists and a breeding ground for dark participation. Beyond the description of single cases, it often remains unclear what exactly is meant by dark social media and which opportunity structures for extremism emerge on these applications. The current paper contributes to filling this gap. We present a theoretical framework conceptualizing dark social media as opportunity structures shaped by (a) regulation on the macro-level; (b) different genres and types of (dark) social media as influence factors on the meso level; and (c) individual attitudes, salient norms, and technological affordances on the micro-level. The results of a platform analysis and a scoping review identified meaningful differences between dark social media of different types. Particularly *social counter-media* and *fringe communities* positioned themselves as "safe havens" for dark participation, indicating a high tolerance for accordant content. This makes them a fertile ground for those spreading extremist worldviews, consuming such content, or engaging in *dark participation*. *Context-bound alternative social media* were comparable to mainstream social media but oriented towards different legal spaces and were more intertwined with governments in China and Russia. *Private-first channels* such as Instant messengers were rooted in private communication. Yet, particularly Telegram also included far-reaching public communication formats and optimal opportunities for the convergence of mass, group, and interpersonal communication. Overall, we show that a closer examination of different types and genres of social media provides a more nuanced understanding of shifting opportunity structures for extremism in the digital realm.

Keywords. dark participation, dark social media, extremism, platform analysis, platform regulation, opportunity structures, scoping review, theory of planned behaviour

CONTENTS

1 Introduction

Social media have become intertwined with our everyday lives. In the wake of measures to combat the COVID-19 pandemic since 2020, such as home-schooling or lockdowns, the use of social media has once again increased sharply (Newman et al. 2021). In Germany, the context of the current study, around 91 per cent of 14–29-year-olds use social media at least from time to time (Beisch & Koch 2021).

Social media offer new opportunity structures for "light," democratic participation, but also for "*dark participation,*" the misuse of digital communication technologies for manipulative purposes (Quandt 2018), for instance through extremists (Rieger et al. 2020). Following increased pressure by civil society and policymakers, major social media concerns such as Google, Meta, or Twitter are increasingly acting against extremist content. At the same time, so-called *dark social media* such as the far-right Twitter alternative *Gab* or the YouTube competitor Bitchute as well as hate-filled imageboards such as 8kun are flourishing (Baele et al. 2020; Chandrasekharan et al. 2017; Rogers 2020).

Dark social media has been described as a home base for extremists and a breeding ground for propaganda, hate speech, and conspiracy theories. For example, the far-right terrorist who killed 51 mosque visitors in Christchurch in 2019 previously posted a statement on the imageboard 4Chan (Comerford 2021). Further, since the outbreak of the COVID-19 pandemic, conspiracy theories have become increasingly popular on Telegram (Hoseini et al. 2021), an instant messenger that is also popular with Islamist extremists (Bloom et al. 2019).

Beyond the description of single cases, it often remains unclear what exactly is meant by dark social media, what types of dark social media exist, how frequent dark participation is on which type of (dark) social media, and what implications the surge of dark social has for extremism and thus for extremism prevention. The current paper contributes to filling this gap. With the help of a detailed platform analysis, as well as a scoping review on extremism and (dark) social media, we examine dark social media as opportunity structures for extremism.

The paper complements and deepens a previously published German-language CoRE-NRW Expert Report on this topic (Frischlich et al. 2022). Both publications are situated in the same theoretical background. However, while the CoRE-NRW Expert Report primarily addresses the detailed results of the analyses and their implications for extremism prevention, the present research paper focuses primarily on the methodological aspects and presents the results of both analyses in an integrative manner along the lines of different types of dark social media. Consequentially, the paper seeks to answer the following overarching research question: *RQ 1: What opportunity structures for extremism emerge in different types of dark social media?*

2 Extremism and Dark Participation

Extremists use social media in a variety of ways (for overviews, see Frischlich 2018; Rieger et al. 2020): for example, to motivate their own group, for planning and coordination their activities, to recruit new followers, or to attack "the enemy," for example by publishing personal data (*doxxing*) or spreading hate speech, and disinformation about others, events, and institutions. This abuse of participatory online technologies by malicious actors with sinister motives and despicable goals is also referred to as "*dark participation*" (Quandt 2018). Dark participation such as the spreading of hate speech and conspiracy theories can have dire consequences for individual and collective well-being (Quandt et al. 2022).

Extremism and dark participation are not the same: neither is all hate speech ideologically motivated (Erjavec & Kovačič 2012), nor is every conspiracy theory wrong or anti-democratic (Baden & Sharon 2021). However, conspiracy theories (Bartlett & Miller 2010; Rottweiler & Gill, 2020; Schneider et al. 2019) and hate speech (Costello et al. 2020; Johnson et al. 2019) are associated with extremist attitudes and they can serve extremist aims. The consumption of hate speech promotes prejudice (Hsueh et al. 2015) and reduces helping behaviour (Ziegele et al. 2018) towards minorities. The consumption of conspiracy theories promotes distrust of societal institutions (van Prooijen et al. 2022) and the willingness to engage in non-normative collective action including violence (Lamberty & Leiser 2019) while decreasing willingness to engage in political participation (Jolley & Douglas 2014). We thus use dark participation to gauge the prevalence of extremism in digital spaces.

2.1 Dark Social Media as Opportunity Structures for Extremism

The concept of "opportunity structures" is used primarily in political science, where it describes the interplay of constraining and facilitating factors promoting respectively inhibiting certain forms of behavior (Tarrow 1988). The flourishing of dark social media is influenced by a variety of such factors that can be described on three interwoven levels: (1) Regulation of and moderation on established social media as influence factors on the societal macro-level. (2) Types and genres of social media with their specific self-positioning on the meso level of distinct social media applications. (3) Attitudes, norms, and technical affordances shaping individual behavior on the micro-level. Notwithstanding, often the same structures that provide opportunities for democratic engagement can be abused for malicious means. Thus, although we focus on dark participation in this paper, this does not mean that there are no positive opportunities emerging in digital media, nor that dark participation does overweight other forms of engagement.

2.2 Regulation and Moderation as Influences on the Macro Level

In the last years, the pressure on major platform operators to remove extremist content has grown. Frameworks such as the *Christchurch Appeal*, the *European Digital Services Act*, or

the German *Network Enforcement Act* (NetzDG) all oblige tech companies to address extremism on their platforms. For instance, the German NetzDG, first implemented in October 2017, obliges platform operators to delete or block illegal content within 24 hours after it has been reported (NetzDG 2017).

Since its implementation, the NetzDG has been the subject of intense debate and criticism. For example, some critics note that the NetzDG only affects social media with more than two million registered users in Germany. This allows smaller platforms to circumvent regulation. Further, the responsibility for enforcing fundamental rights such as freedom of expression or the protection of privacy is handed over to (mostly) US corporations. Criteria for content deletion are neither democratically agreed upon nor even transparently documented. Plus, platforms could restrict the spectrum of what can be said as a precautionary measure to escape penalties (so-called *overblocking*). Finally, the NetzDG does already serve as a model for authoritarian governments to act against unpopular critics, as media law professor Wolfgang Schulz explained in an interview (Markert 2020).

Beyond content-removal, platform operators can also (permanently) block entire accounts (so-called *de-platforming*) (e.g., Jhaver et al. 2021). De-platforming is also discussed very critically, and its effectiveness is evaluated differentially (Rogers 2020). On the one hand, removing influential extremists from a social medium does lead to a significant loss of digital followers, as research on right-wing extremists (Jhaver et al. 2021; Rogers 2020; Wong 2018) and Islamist extremists (Conway et al. 2019) shows. Even though blocked users often come back under a new name, it takes time to regain the followers (Stern & Berger 2016). Especially when isolated communities that heavily rely on a single platform are banned, the disruptive effect can be large (Chandrasekharan et al. 2017).

On the other hand, de-platformed individuals can present themselves as victims of a digital "witch hunt." Moreover, most people nowadays rely on multiple digital communication channels and simply shift their activities to other channels after a ban (Baele et al. 2020). This works particularly well if the ban is announced in advance (Rogers 2020) and the respective actors can inform their followers about the new meeting spots. Thus, the flourishing of dark social media is a sign of the success of regulatory efforts but also offers new opportunity structures for extremists.

2.3 Genres and Types of (Dark) Social Media as Influences on the Meso Level

"Social media" is an umbrella term for various applications that (a) are based on digitally networked technologies; (b) enable their users to create a profile within the system or make information accessible in this system; and (c) interact with other profiles or information providers and thus establish or maintain social relationships (Taddicken & Schmidt 2017). Social media are very diverse and difficult to categorize into specific "genres," in part because their technological functions change rapidly (Taddicken & Schmidt 2017). Nevertheless, different social media bundle different functionalities in a certain way, which does allow to broadly

distinguish different genres. Building on Taddicken and Schmidt (2017) *publication-oriented social media*, *digital platforms*, and *chat-based* social media can be distinguished. Although single examples within these genres have similar technical functionalities, they do invite different usage routines and thus play a different role in the media diet of their users.

Publication-oriented social media. Publication-oriented social media, such as *YouTube* or *Twitter*, focus on publishing own content and responding to the content of others. Whereas in traditional outlets such as newspapers or television, editors and journalists serve as *gatekeepers* for the public sphere (e.g., Heinderyckx 2015; for an overview, see Shoemaker & Vos 2009), publication-oriented social media allow virtually anyone to broadcast their own content. The content is often (though not always) fully public and can be consumed without additional registration. Taddicken and Schmidt (2017) distinguish three subgenres: *blogs*; *microblogging services* (e.g., Twitter), and *non-text-based formats* (such as Soundcloud). In an extension of their taxonomy, we further consider *format-oriented publishing platforms* such as YouTube for videos or *Instagram* for images (and increasingly for short videos) and *live-streaming platforms* such as Twitch as genres of publication-oriented social media.

Digital platforms. Digital platforms are technical infrastructures that focus on networking and communication (Taddicken & Schmidt 2017). Digital platforms can combine various functions, such as an internal search, the uploading of images, texts, videos, streams and so on, the discussion of these contents, and private chats. We consider two subgenres here (a) *social networking sites* (e.g., Facebook) and (b) *digital discussion platforms* (e.g., Reddit). Social networking sites offer a vast number of communication channels, fostering the exchange within users' (digital) networks (Boyd & Ellison 2007; Ellison et al. 2009); while discussion platforms allow for egalitarian exchange with all other users.

Chat-based social media. Chat-based social media focus on communication between individual users or a small number of users who are often known to each other. Taddicken and Schmidt (2017) distinguish between *instant messengers* (e.g., WhatsApp) and *video-based systems* (e.g., Zoom, Skype). In this study, we focused on instant messengers due to their high popularity and their importance for extremist recruitment efforts. Furthermore, most instant messengers do also entail a video chat function nowadays. We also include *chat-oriented messengers* namely Discord and Snapchat as these are popular among young media users.

Types of dark social media. Beyond different genres, so-called "mainstream" social media (such as *Facebook*, *Twitter*, or *YouTube*) can be distinguished from *dark social media*. Here, we understand dark social media as offerings that differ from their mainstream counterparts either (1) regarding the establishment of specific communication norms or (2) by the privacy of the content. We suggest that different types of dark social media exist that can be broadly described as either *alternative social media* or *dark channels*.

We understand **alternative social media** as applications that position themselves as "correctives" to a hegemonically interpreted mainstream in a particular socio-cultural context offering a platform to those actors or positions that experience themselves as inadequately represented or marginalized within this mainstream. Thus, alternative social media are the platform counterpart to alternative news media (Holt et al. 2019). What is experienced as "alternative" or "mainstream" is heavily context dependent and different alternative social media relate differentially to "mainstream" applications. At least three subtypes can be distinguished:

(1) *Social counter-media.* These types of alternative social media orient themselves in functionality or design towards specific mainstream offerings while positioning themselves as a "safe haven" for content and people who feel ostracized or otherwise marginalized in "mainstream" social media. Social counter-media are often a self-declared bastion of "free speech" (like the concept of dark platforms in Zeng & Schäfer 2021). The microblogging service Gab, for example, is strongly oriented toward Twitter in terms of functionality, but addresses an ideologically defined Christian fundamentalist, far-right user base.

(2) *Context-bound alternative social media* also resemble the (among Western users) familiar social media in design and functionality. They do not position themselves as a "safe haven" for deviant positions, but rather address the entire population. Context-bound social media function as an alternative primarily through their orientation to their, often non-U.S.-based, contexts of origin, where different legal frameworks apply. For example, *VKontakte* serves as the Russian competitor for Facebook and declares to apply Russian law.

(3) Finally, *fringe communities* are not about imitating mainstream social media or addressing a mainstream audience. Instead, they target specific subcultures (Phillips 2012). The communication logic in fringe communities is very much shaped by the subcultures that use it and accordingly it is often difficult to grasp for outsiders. Well-known fringe communities such as *4Chan* are infamous for tolerating content and actors that have been or would be banned from other platforms.

Dark social channels (Madrigal 2012) describes digital channels used for (rather) private communication-such as messengers or closed group chats. Dark channels do not position themselves as an alternative to mainstream social media, but as an alternative to face-to-face communication or to writing letters. At least two forms can be distinguished: (1) *private-first channels*, such as instant messengers, which primarily served private communication. (2) *private rooms*, which are a part of larger social media (e.g., closed Facebook groups).

2.4 Behavior as a Function of Attitudes, Norms, and Affordances on the Micro-level

To understand how the use of certain (dark) social media might influence the behavior of users, we draw from the *theory of planned behavior* (for an overview, see Ajzen, 1991). The theory of planned behavior assumes that human behavior is motivated by three factors: (a) attitudes, particularly concrete behavior-related attitudes; (b) subjective norms regarding the behavior; and (c) behavioral control. Together, they influence behavioral intentions. Whether the behavior is then carried out is further shaped by behavioral control and the concrete options in each moment. If a social medium has no livestream function, nothing can be streamed. All three components can be applied to the context of (dark) social media.

The first thing influenced by users' attitudes is their choice of a given (social) medium. We live in 'high-choice' media system that allows users to easily select applications and contents following their preferences (Van Aelst et al. 2017). People make media choices partially habitually but also because they expect certain benefits, so-called *gratifications*, from it *(Uses and Gratifications Theory*, Katz et al. 1973). One of these gratifications is the confirmation of one's worldview as people generally prefer content and communities that validate their worldview over those who challenge them (*selective exposure*, Fischer & Greitemeyer 2010; Stroud 2010). Hence, the self-positioning of a (dark) social medium likely influences the audiences it attracts.

Whether a certain behavior is executed depends on the subjective norms in each situation. Two types of norms can be distinguished: (1) *injunctive norms* describing how one should behave and (2) *descriptive norms* that result from observing the behavior of relevant others, i.e., how people actually behave (Deutsch & Gerard 1955; Rivis & Sheeran 2003). Descriptive norms often have a substantially stronger influence on behavior (Manning 2009). In social media, injunctive norms are likely derived from legal regulations (e.g., the German NetzDG) as well as the respective "house rules," such as the netiquette or general terms of services (ToS), while descriptive norms are inducted from the behavior of other users.

Different norms can be salient and guide peoples' behavior (Cialdini et al. 1990). One factor that influences the salience of norms is one's anonymity. The *Social-Identity Deindividuation Effects* (SIDE) Model (Postmes et al. 1998) suggests that if personal characteristics, such as one's appearance or one's real name, are not recognizable, personal norms become less salient and the norms and identities of the community and context become more salient (*deindividuation*). Thus, under conditions of deindividuation, the norms of the community become particularly influential and it's likely that users orient more strongly on the behavior of other users.

Finally, behavior requires behavioral control, i.e., the opportunity for execution. Thus, the *technical affordances* of social media must be considered. The concept of affordances originally described the perceived environment and its usability, originally for example the food supply for an animal in its ecological niche (Gibson 1979). The concept of technical

affordances carries this idea further. Technological affordances are (1) based on technological functions, (2) influenced by the use of technology, and (3) can be used in different ways (Evans et al. 2017). For example, anonymity can provide protection from persecution by authoritarian states but also from legitimate law enforcement, a live stream can document human rights violations or serve propagandist means.

A first attempt to systematize technological affordances relevant for radicalization in online environments has been provided by Schulze et al. (*in print*). First, they name affordances that could promote (unintentional) **confrontation with** extremist **content**. These are (1) *algorithmic recommendation systems* that can be used by extremists (e.g., via hashtag hijacking) to spread their content (Miller-Idriss 2020; Schmitt et al. 2018) or enforce radical worldviews. For example, consuming an extremist video on YouTube can encourage the recommendation of further extremist material (Faddoul et al. 2020; O'Callaghan et al. 2015). (2) The *dialogue culture*, i.e., whether a social medium encourages exchange with known versus unknown persons (Costello et al. 2016). We suggest to also add (3) *targeting*, i.e., the possibility for content producers to send their content only to people who search for certain keywords or who belong to certain groups or have certain interests (e.g., men, 16–24, who are interested in violence). Targeting is usually related to a platforms business model and thus offered mostly in the context of advertisement.

Second, certain affordances shape **communication and** thus also the **relationships** that can be formed with extremists. These affordances include (4) subjectively significant *parasocial relationships* (Horton & Wohl 1956) with extremist *influencers*, that might arise through repeated engagement with an official profile or the host of a channel. Although parasocial interactions do not entail "real" interactivity, they can foster feelings of safety and friendship (for an overview, see Brown 2021) and motivate purchase decisions (Breves et al. 2021). Further (5) the design of *group communication* (e.g., whether groups have separate group chats, are presented on ones' digital profile and so on) can be more or less suitable to motivate the fusion of one's own identity with the extremist identity during a radicalization process (Hamid et al. 2019; Swann et al. 2014). Partially related, we suggest to also add (6) *symbolic self-presentation* opportunities to the list (e.g., profile pictures, nicknames). Expressing extremist identities can help to attract the attention of like-minded others (Bloom et al. 2017) and make the respective (extremist) group-identity (and its norms) more salient.

Finally, Schulze et al. (*in press*) highlight the need to consider affordances that influence the situational **salience of different norms**: (7) the level of *anonymity* respectively pseudonymity according to the principles of the SIDE model (Postmes et al. 1998), and (8) *community management* to deal with dark participation. Most plausible, it also plays a role whether a social medium *cooperates with security agencies*. Overarchingly, technical affordances are influenced by the specific genre of the social medium: instant messengers typically bundle different affordances than publication-oriented platforms.

3 Methods

We conducted two studies to examine the new opportunity structures emerging on different types and genres of dark social media: A comprehensive platform analysis and a scoping review on extremism, dark participation, and (dark) social media covering the published literature in German and English. In the following, we describe the data basis and methodological approaches of both studies in detail before discussing their results jointly for the different types of dark social media. For detailed insights into the individual studies, please refer to the (German) CoRE-NRW Expert Report (Frischlich et al. 2022) or contact the authors.

3.1 Study 1: Platform analysis

Methodological approach. We conducted a qualitative content analysis according to Mayring (Mayring & Fenzl 2014). Qualitative content analysis is a standard method in communication science (Brosius et al. 2016). During the analysis, pre-defined features of the object of interest (e.g., sentences, social media posts) are assigned to specific categories following pre-defined rules. Mayring's method is a strictly rule-based, intersubjectively verifiable procedure that can employ different approaches. Here, we choose a summarizing approach during which the material is reduced in an iterative manner allowing categories to emerge from the material.

First, we conducted four unstructured descriptions (one by each coder) of Telegram as well as an ethnographic review by the last author of Facebook. Based on the descriptions and the theoretical framework guiding this work, we developed our initial category system. Afterwards, all coders coded the messaging app Signal to determine intersubjective agreement. Disagreements were resolved in a coding conference and informed the revision of the category system. Two student coders coded all platforms using the revised category system. If new technical functions were identified (e.g., self-deleting messages), we integrated them into the category system and ensured that all platforms were recoded if neccessary. Additionally, we researched relevant facts (e.g., reach, operators, date of foundation) about the different platforms through studying the platforms' self-descriptions and (if available) their terms of service, respectively their "netiquette" or frequently asked questions. We used the online application Similarweb that provides estimated traffic statistics to gauge the reach of each of platform. Similarweb provided us with (1) global website traffic for covering the period from October to December 2020, and (2) German app traffic data for September 2021.

Data base. We identified social media based on the academic literature (e.g., Zannettou, Caulfield et al. 2018; Zeng & Schäfer 2021), reports from civil society institutions (Clifford 2018; Guhl et al. 2018; Guhl & Davey 2020) and media reports (Manakas 2021). We selected dark social media used by extremists, as well as mainstream social media covering different genres. In total, *N* = 19 platforms were selected for analysis (see Table 1). In the following, we focus primarily on dark social media. For detailed insights into functional differences between dark and mainstream social media, see Frischlich et al. (2022).

During coding, we first installed the respective app and created a user account. To protect the coders, we used anonymous accounts (i.e., a business phone number and a neutral e-mail address). To make the research context salient, we always used a variation of our initials followed by "researcher" as usernames (e.g., "xy.researcher"). Before the actual coding, the coders familiarized themselves with the respective social medium and studied the ToS. We tested all technical functions such as joining groups, chats, etc. but we did not directly interact with other users.

Table 1: Analysed social media.

Genre	Subgenre	Mainstream Social Media	Dark Social Media
Publication-oriented social media	Microblogging	Twitter	Gab
	Format-oriented publication platforms	Instagram, YouTube	BitChute, TikTok
	Live streaming	Twitch	DLive
Digital platform	Discussion platform	Reddit	4Chan, 8kun
Chat-oriented	Instant messenger	WhatsApp	Telegram, Signal, Threema
	Chat-oriented messenger	Discord, Snapchat	

Category system. The final category system included15 categories and 106 subcategories. The complete category system can be accessed in German via the Open Science Framework https://osf.io/x9sye/. Regarding confrontation with (potentially extremist) content, the following affordances were coded: (1) Algorithmic recommendation of content and contacts. (2) Search function and (3) topic organization (such as use of hashtags). In addition, (4) monetization options and targeting of specific audiences were measured. Furthermore, (5) the variety of communication channels (such as the possibility to upload own content, the availability of groups and chats) and the (6) availability of symbolic communication (emojis) and (7) options for symbolic self-expression (e.g., profile pictures, nicknames) were coded.

The options for uploading specific content (such as images, livestreams, etc.) were also mapped in detail. To consider the different levels of communication and relationships, upload options were coded separately for (8) profile owners, (9) chats, (10) groups and (11) from the perspective of the users. Further, we recorded whether (12) the platform favors influencers by prominently highlighting some accounts (such as verified Twitter accounts).

Regarding technological affordances that might influence the salience of certain norms we studied the platforms' own rules (e.g., netiquette, ToS, FAQs). We distinguished two types of anonymity: (13) "true" anonymity vis-à-vis the platform, for example through the possibility

of using the content without registering or by registering via email alone. (14) Pseudonymity on the platform itself, by the extent of social information that is visible on the platform (e.g., Facebook friends), geographic location, or the use of pseudonyms versus phone numbers. Finally, we coded (15) the implementation of injunctive norms by the presence of a netiquette and whether the platform claimed to cooperate with law enforcement.

3.2 Study 2: Scoping Review

Methodological approach. We conducted a scoping review to map the research field (Munn et al. 2018; Teare & Taks 2020). Scoping reviews are particularly useful for capturing the state of research in complex and interdisciplinary research fields (Schindler & Domahidi 2021). For example, to describe the state-of-evidence in a research field, map typical methods, or to identify research gaps (Munn et al. 2018).

Based on the research question, we first determined inclusion and exclusion criteria for the literature selection. The brand names of social media of different genres and types served as key terms, complemented by synonyms for dark social media. Boolean operators were used to search for texts in which these keywords were used in combination with different types of extremism (e.g., right-wing extremists*, left-wing extremists*, conspiracy theorists*, Islamists*, etc.) or dark participation (e.g., hate speech, conspiracy theories, extremism, etc.) or measures against dark participation (e.g., de-platforming, platform migration). In addition, we searched for links between extremists, dark participation, or measures against either of the two using various forms of digital participation (e.g., memes, instant messaging) and theoretically relevant concepts (e.g., affordances). Table 2 provides an overview of the search terms and linking logic.

We included scientific literature from various disciplines (e.g., communication science, computer science, IT security, pedagogy, political science, law, sociology), as well as so-called gray literature, such as reports from public authorities and NGOs. It should be noted that due to the duration of the project, no unpublished literature could be included. Although we did include pre-prints that have not yet been formally reviewed, a bias of the findings due to the so-called publication bias (the lower probability that null findings are published) cannot be ruled out (Kicinski et al. 2015; Sutton 2000).

We used two curated academic databases (*Academic Research Ultimo; Media & Communications*), a broad academic database (*Google Scholar)*, and a project database on the topics of democratic resilience, online propaganda, fake news, fear and hate speech curated by the first author. Since the order of search results, especially on Google, can differ depending on the person searching (Emmer & Strippel 2015), we always reviewed the first 50 hits. If relevant texts were identified, the search was continued until at least 20 hits contained no new or matching articles (“theoretical saturation”). We cleaned the database by excluding articles that dealt with unrelated topics (e.g., “Telegram for digital learning”), project announcements, job advertisements and PR materials, blog posts, or the like and removed duplicates.

Table 2: Search keywords and logic of the Scoping Reviews.

One of the terms		AND on of these terms		
Platform OR	**Participation**	**Ideologies OR**	**Content OR**	**Processes**
4Chan 8Chan 8kun Alternative Media BitChute Chan Dark Platform* Dark social media [Dunkle Soziale Medien] Discord DLive Facebook Gap Instagram Parler Reddit Snapchat Social Media [Soziale Medien] Social Networks [Soziale Netzwerke] Socio-technological systems [Sozio-technische Systeme] Technological affordances [Technische Affordanzen] Telegram TikTok Twitch VKontakte WhatsApp YouTube	Instant Messaging Images Meme* Post* Viral Content Videos	Conspiracy believers [Verschwörungstheoretik*] Far-right Far-left Islamic fundamentalism Islamist* Left-wing extrem* [Linksextrem*] QANON Querdenk* Right-wing extrem*[Rechtsextrem*] White Supremac*	Conspiracy theories [Verschwörungs*] Dangerous speech Extreme* Extremis* Extremist propaganda [Extremistische Propaganda] Hate speech Radical	De-platforming Extremism Prevention [Prävention] Plattform Migration

Notes. German search terms are provided in bracelets. Both German and English terms were used.

Data basis. Based on this procedure, *N* = 142 texts could be identified for the analysis. The majority originated from scientific sources, while just under one-fifth were grey literature. Nine out of ten scientific sources were published in English. (Dark) social media and

extremism is investigated in different fields but particularly in the social science (e.g., communication and criminology) and in the information sciences. Only three works could be traced back to other disciplines (business administration, mathematics, philosophy). The oldest text was published in the early days of digital media (Pfeiffer 2002) but did not deal with dark social media directly. Most publications date back to the years around the beginning of the second decade of the 21st century. Especially since 2019, dark social media raises increased interest.

Half of all texts reported content analyses, whereas studies using human subjects were seldom. In the following, we rely on the content analytical work to estimate the prevalence of dark participation. Both qualitative and quantitative as well as manual and computational content analyses were included. In the following, we present the combined findings of both studies in an integrated manner using different types of dark social media. For more detailed results on the individual studies, please refer to the Expert Report (Frischlich et al. 2022).

4 Results

4.1 Social Counter-Media

We examined three *social counter-media*: The microblogging service Gab, which mimics Twitter, the video streaming platform Bitchute, resembling YouTube, and the gaming-oriented livestreaming platform DLive, which mimics Twitch. Bitchute claims to use a decentralized system to serve videos to its users, which is based on WebTorrent, a peer-to-peer streaming client. Thereby, Bitchute promises users that de-platforming would be impossible, as no central actor has full control over the videos on the platform. Yet, the use of WebTorrent is disputed and as of 2021 this feature appears to have been deprecated[1]. However, DLive still uses blockchain for its donation systems, so users can monetize their content without using central actors like PayPal. Gab described itself as a "social network that promotes free expression, individual freedom and the free flow of information on the Internet" (via DuckDuckGo, August 21, 2021). BitChute stated that they would provide creators "with a service to unfold and freely express their ideas" (via DuckDuckgo.com, September 14, 2021). DLive described itself as a "live streaming community on the blockchain" and advertised "great games" (via DuckDuckGo, August 27, 2021).

Compared to mainstream social media of the same genre, social counter-media were not very successful as judged by the Similarweb statistics. In 2020, Gab had 1.82 million monthly users worldwide, compared to 901.80 million Twitter users. The difference is similar for Bitchute versus YouTube, even if videostreaming platforms generally outperformed

[1] For an examination of the decentralization claims, see https://www.dailydot.com/upstream/bitchute-de-centralization-claims/, https://arstechnica.com/tech-policy/2021/04/conspiracy-theorist-said-death-threats-were-jokes-but-jury-didnt-buy-it/

microblogging services. Globally, 161.58 times more people use YouTube (1.96 billion) than Bitchute (12.13 million). DLive reached only a minority. While a stunning 1.08 billion people worldwide used Twitch monthly, only 6,224 turned to DLive during the same period.

The three social counter-media differed in their injunctive norms. Gab did neither provide a netiquette nor state that it would cooperate with state authorities. In response to a subpoena from the U.S. Congress investigating the storming of the U.S. Capitol on January 6, 2021, the platform reportedly refused to cooperate, stating that it would not track content or investigate users (Kimball 2021). Bitchute demanded its users to refrain from discrimination based on ethnicity, gender, religion, politics, and the like and maintained a list of banned accounts. However, only three organizations were named on that list by January 2022. Finally, DLive called on its users to treat each other with respect in its ToS and states that it can remove content from its platform, ban users, or terminate accounts permanently.

On all three platforms, the true anonymity of the users vis-à-vis the platform and the pseudonymity on the platform itself was high. An email address was sufficient for registration, and self-portrayal took place with the help of a freely chosen nickname and profile picture. According to the SIDE model described above (Postmes et al. 1998), social counter-media thus allowed for deindividuation which makes it likely that the descriptive norms on the platform are particularly salient and influential on users' behavior.

To depict these descriptive norms, we reviewed the literature identified in the scoping review. These studies showed that social counter-media allowed for high levels of dark participation. In Gab's case, this was mainly due to right-wing extremist content (Jasser et al. 2021; Lima et al., 2018; Woolley et al., 2019) and conspiracy theories (Zeng & Schäfer 2021). Further, the amount of hate speech on Gab was higher than on Twitter (Zannettou, Bradlyn et al. 2018) and the level of hate seemed to be growing (Mathew, Illendula et al., 2019), likely because hateful postings were rewarded by other users. Accounts that frequently posted hate speech were more interconnected and had a higher reach than accounts that did not "deliver" such content (Mathew, Dutt et al. 2019). For BitChute, the proportion of Hate Speech even exceeded that observed for Gab in a comparative study (Trujillo et al. 2020). Moreover, only a small fraction of videos elicited user responses and almost all of them were hateful or conspiracy theorist. We did not find empirical work on DLive[2].

In terms of technological affordances, all three social counter-media were publication-oriented formats. Accordingly, the main functions provided were for publishing one's own content and reacting to the content of others. Users saw content that was uploaded by accounts they followed, that they searched for, or that was recommended to them algorithmically.

[2] But shortly after the data collection, the Institute for Strategic Dialogue published a report on DLive that can be downloaded under the following link: https://www.isdglobal.org/isd-publications/gaming-and-extremism-the-extreme-right-on-dlive/

Publication-oriented platforms fostered parasocial relationships with content providers but also allowed for exchange with unknown others for instance through commenting on certain posts or using certain hashtags. There were no pronounced group spaces. None of the social counter-media offered targeting. However, Gab planned options for advertisers, which, according to the platform, were intended to help freeing Gab from the "globalist system of enslavement, degeneration, and destruction" (Torba 2021).

Taken together, social counter-media offered new opportunity structures for the dissemination of extremist propaganda, conspiracy theories, and hate speech, as well as for the consumption or redistribution of corresponding content (dark reception, so to speak). Due to the high level of pseudonymity, conformity to descriptive norms on the platform was likely and due to the low reach, the NetzDG did not exert any moderation pressure. This might explain why an initial study shows that dark participation on these platforms is increasing over time (Mathew, Illendula et al. 2019).

4.2 Context-bound Alternative Social Media

Context-bound alternative social media resemble more prominent social media in the West in terms of their design and functionality. They do not position themselves as a "safe haven" for deviant positions, but rather address the entire population of their respective country of origin and beyond. We categorize them as alternative primarily through their orientation to specific, often non-U.S.-based, contexts of origin. In this study, two services were examined in greater detail: The social networking site *VKontakte*, which was founded in Russia and is replacing Facebook there, and the short-video platform *TikTok* from China, which enjoys great popularity especially among children and young people in Germany (Rathgeb & Schmid 2020) and is rapidly becoming a mainstream social medium across the world.

Context-bound alternative social media reach much more people than social counter-media, though still substantially less than the more prominent US-based applications. In 2020, VKontakte had 128.3 million monthly users following similarweb, 15-times less than Facebook's two billion. TikTok reached an impressive 1.01 billion users worldwide placing it between Facebook and YouTube, each with almost twice as many users, and Twitter with slightly less than one billion monthly users. Both context-bound alternative social media emphasized reach and functionality in their self-description, TikTok, for example, pronounced that their users can "watch and discover millions of personalized short videos" (via DuckDuckGo, August 24, 2021). VKontakte claims that it connects "millions of people" (via play.google.de, October 1, 2021). Both platforms are subject to the NetzDG in Germany.

Proclaimed injunctive norms were comparable to those of the well-known U.S. offerings. Extremism and terrorism are prohibited in both community standards. However, VKontakte explicitly refers to the Russian regulatory authorities, while TikTok states separate responsibilities for users in the US, UK, and EU vs. the rest of the world. Notable, TikTok has been

criticized obeying Chinese censorship also for Western audiences (for a media report, see Hern 2019).

Like Facebook, VKontakte requires a relatively large amount of personal data, including a phone number, a name, gender, and date of birth (even though the information is not necessarily verified). An email address is sufficient for using TikTok. Both platforms offer various options for self-expression, for example via nicknames and profile pictures, including the display of one's network of contacts on the platform (VKontakte) or the charitable organizations one supports (TikTok). Typical for TikTok are videos in which the user dances, sings, or otherwise performs. Consequently, it is likely that users' personal identity comparably salient on both platforms, increasing the likelihood that personal values and norms influence one's behavior (as opposed to a more deindividuated state in anonymous contexts).

Regarding descriptive norms, the (few) available studies on VKontakte show that the relation to the Russian legal space must be considered. According to media reports, VKontakte has recently been headed by a confidant of Russian president Putin (AFG & ARG 2021). To date, social movements positioning themselves against the Russian state have also found an important tool on the platform (Poupin 2021), and action has been taken against Russian ultranationalists through the platform (Kashpur et al. 2020). At the same time, VKontakte also provided a platform for extremist actors (Myagkov, Chudinov et al. 2020; Myagkov, Shchekotin et al. 2020). TikTok's parent company, ByteDance, also has ties to state institutions. Following news coverage, the Chinese government has held a stake in the company since April 2021 (Lang, 2021). Initial studies report extremist and hateful content on TikTok too (O'Connor, 2021; Weimann & Masri, 2020). Yet, comparative analyses of the prevalence of dark participation versus other content or other social media could not be identified. Clearly, dark participation is found on all major social media (for instance, on YouTube (e.g., O'Callaghan et al. 2013; Rauchfleisch & Kaiser 2020), Facebook (Farkas et al. 2018; S. Kim & Kim, 2021; Scrivens & Amarasingam 2020), Instagram (Bouko et al. 2021; Frischlich 2021), Twitter (Al-Rawi & Groshek 2018; Berger 2016; Bhat & Klein 2020). Thus, differences to context-bound alternative social media are not clear.

In terms of technological affordances, context-bound alternative social media did not differ from well-known mainstream social media. Content is shown based on the social network of followed accounts or algorithmically. Searching by keyword is also possible. TikTok influencers play an important role, especially among children and young people. Successful profiles can reach millions (Eisenbrand 2021). Users can also get in touch with unknown people. Both plattforms enable targeting. In sum, context-bound alternative social media offered similar opportunity structures as U.S.-based platforms but had closer links to governments.

4.3 Fringe Communities

We examined two image boards as cases for fringe communities: 4Chan and 8kun. 4Chan was founded in 2003 and has hardly changed its design since then. 8kun was founded in

2019 as a successor to 8Chan. 8Chan had gained global attention after the far-right terrorist who targeted and killed Muslims attending a religious service in Christchurch, New Zealand, posted his claim of responsibility there before he live-streamed the attack. An evaluation of the discussion around the attack showed that the proportion of pro-violence content on 4Chan was lower than on 8Kun, with most explicitly pro-violence content on the even smaller platform 16Chan (Comerford, 2021). 4Chan itself emphasizes true anonymity on the platform ("4chan is a simple, image-based bulletin board where any one can anonymously post comments and share images" (via DuckDuckgo.com, September 17, 2021), 8kun pronounced the usability ("On 8kun, you can create your own image board for free, without needing any experience or programming knowledge" (8kun, September 8, 2021)).

The reach of both chans lag far behind the popular discussion platform Reddit. While Reddit had 1.74 billion monthly users worldwide in 2020, 4Chan had just 4.82 million and 8Kun only 27.198. Fringe communities are thus the least used type of social media in our study.

Regarding injunctive norms, both chans emphasize that dark participation is welcome on their platforms. 4Chan does have terms of service, but it mainly points out that racist and pornographic content is allowed on certain boards only and that posting personal information ("doxing") and calling for invasion ("raiding") of boards is not allowed. In addition, it is noted that complaints about 4Chan may result in de-platforming. When entering boards, one is warned that the content is only suitable for adults. 8Kun states only one rule: no illegal content based on US-law. Both services claim to cooperate with law enforcement.

The chans are the only fully anonymous services in our study since they can be used without registration and most users do not have a profile. Even though individual users can start threads and others then respond to these posts, the relationships are not hierarchical in the sense of an influencer-follower relationship. By reading a thread on the board, single posts are often published by "anonymous." Therewith, the posts leave the impression of an overall chan-discourse more than a conversation between distinct individuals. References to other posts are made via post ids not via nicknames. Consequentially, the SIDE-Model would predict a high orientation towards the descriptive norms in the respective boards.

Studies emphasize that it is often difficult for outsider to distinguish between the (very) dark humor around which 4Chan was built (Phillips 2012) and terrorist content such as that of the Christchurch perpetrator still celebrated on the platform (Comerford 2021). It should also be emphasized that humor can be used in a strategic manner to "mainstream" hateful ideologies (Askanius 2021; Schmitt et al. 2020; Schwarzenegger & Wagner 2018). Clearly, hateful content is more prevalent on the chans than on other social media (Zannettou, Caulfield et al. 2018); moreover, image boards play a central role in the "birth" of political memes (Crawford et al. 2021). Studies report posts that encourage others to see themselves not only as digital activists, but as "survivors" and "soldiers" (Elley 2021). With roughly one quarter of

posts being hateful, 4Chan thereby was slightly more moderate than 8Chan, where one-third of all posts were classified as hate speech in a comparative study (Rieger et al. 2021).

Regarding the technological affordances, chan users see the content of the board they entered in chronological order. Within the boards, keyword search is enabled. There is no algorithmic sorting, but popular threads are recommended. The content of the boards cannot easily be deducted from its title. For example, the 4ChanBoard on the children's TV-series my little pony /mlp/ contained mainly pornographic references in December 2021.

Taken together, fringe communities provide the best opportunity structures for users' own dark participation. Through explicitly tolerating content otherwise considered deviant and their complete anonymity, it is likely that users orient themselves towards the (dark) participation rules on the platform. Noteworthy, hateful behavior in one community does not mean that corresponding behavior is found everywhere else too. Studies on Reddit show strong differences between individual subreddits, even the same users adapted to the prevailing discourse norms depending on the board they used (Gibson 2019).

4.4 Private-First Channels

Private-first channels such as chats, but especially instant messengers are enjoying great popularity. WhatsApp is now used by 80% of adults in Germany (Beisch & Koch 2021). Here, we examined three competitors to WhatsApp: (1) *Telegram*, an app founded by Russian siblings Pavel and Nikolai Durov during their time at VKontakte. (2) *Signal*, an open-source app which encrypts data end-to-end and does not store it on its own servers, designed by former Twitter security chef Moxie Marlinspike and WhatsApp co-founder Brian Acton, and funded by the nonprofit Signal Foundation. (3) The Swiss app Threema, which also places high value on encryption and data protection.

All instant messengers offer functionalities beyond chatting. Particularly Telegram promotes large-scale public channels. The messenger advertises "group chats with up to 200,000 people" and offers channels where thousands of people can be reached. Telegram combines these features in its self-presentation with other features that might be attractive for those fearing persecution: "end-to-end encryption" and possible "self-destruction" of the messages (via DuckDuckGo, September 22, 2021). Signal and Threema also enable group communication, but both services primarily emphasize the security of communication and seem to strive for smaller audiences. Signal, for example, describes itself as "a messaging app for simple private communication with friends. Signal uses your phone's data connection [...] to communicate securely [...] and can also encrypt the saved messages on your phone" (via https://github.com/signalapp, October 6, 2021). Threema promotes its business offering as a "secure and privacy-compliant messaging solution" (via DuckDuckgo.com, September 15, 2021).

Both Signal and Telegram are widely used in Germany: In 2021, 3.31 million people used Telegram and 3.41 million Signal daily. This is about 10 times less than WhatsApp users (32.14 million), but significantly more than Threema users (584.460).

Regarding injunctive norms, Threema states that it cooperates with law enforcement agencies. Signal does not make any announcements. Telegram states that it considers all private and group chats to be private and does not act against illegal content. It also explicitly denies cooperation with law enforcement. Yet, Telegram does allow its users to report channels, bots, and stickers as illegal or fraudulent. However, it remains unclear whether and how Telegram reacts to these reports. In 2021, the channels of infamous German conspiracy ideologue Attila Hildmann were blocked on Google and Apple-based Telegram apps. However, it remained unclear who was ultimately responsible for the blocking, Telegram itself or the larger stores (for media coverage, see sede & guth 2021).

Anonymity varied between the three instant messengers. Telegram required a phone number, whereas Threema could be used with the help of an anonymous so-called Threema ID. Signal required a phone number, but a (potentially inactive) landline number could also be entered. On the platforms themselves, users were identified via nickname, partially via profile image and their phone number.

Our scoping review identified several examinations of dark participation and extremism on Telegram. Different studies examined Islamic extremist activity on the platform, particularly during the most active period of the self-declared “Islamic State in Iraq and Syria” (ISIS or daesh) (Bloom et al. 2017; Clifford 2018; Inquirer 2016; Prucha 2015; Yayla & Speckhard 2017). When Telegram removed official ISIS-channels, it was considered to be a very successful disruption (Amarasingam et al. 2021). Yet, other extremists such as other Islamic extremists or right-wing extremists still find a “safe space to hate” (Guhl & Davey 2020, p. 1) and host multiple channels on the platform (Urman & Katz 2020; Walther & McCoy 2021). Since the outbreak of the Corona Crisis, extremist conspiracy believers have also been using Telegram increasingly (Holzer 2021; Hoseini et al. 2021). Telegram has also been found to enable drug trafficking and pornography (Jünger & Gärtner 2020). Studies on dark participation on Signal or Threema could not be identified.

All instant messengers displayed messages sorted by sender, group, or (Telegram) channel. There is no algorithmic recommendation. Getting in contact via instant messengers usually depends on knowing that persons contact details. On Telegram, users can search for groups or channels and join them with a single click. Telegram groups thus potentially allow for contact with so-far unknown others. Within the groups, messengers are particularly good in fostering interpersonal exchange and intragroup communication. None of the messengers offered targeting of specific user groups. Telegram also allowed parasocial relationships with channel owners.

Taken together, Telegram stood out among the instant messengers due to the high convergence between private and public communication channels and the explicit combination between privacy-oriented features with features for group and even mass communication. This raises the question how specific functionalities can be regulated in a more nuanced manner while preserving privacy protection in private first channels at scale.

5 Discussion

Social media have become intertwined with our everyday lives. This offers new opportunity structures for democratic participation, but also for "dark participation," the abuse of digital communication technologies for manipulative purposes (Quandt 2018), for instance through extremists (Rieger et al. 2020). Dark social media has been described as home base for dark participation. However, beyond the description of single cases it often remains unclear what exactly is meant by dark social media, what types of dark social media exist, how frequent dark participation is on which type of (dark) social media and what implications the surge of dark social has for extremism and thus for extremism prevention. The current paper contributes to filling this gap.

We discussed how regulatory attempts such as the German network enforcement act (NetzDG) motivated large-scale social media to moderate dark participation more harshly and therewith contributed to the success of dark social media with a similar variety of genres as far-reaching tech-giants. We defined dark social media as offerings that differ from mainstream social media either regarding the establishment of specific communication norms, what we labelled *alternative social media*, or by the original privacy of the content, what we termed *dark channels*. We analyzed four distinct types of dark social media: Three were alternative social media (1) *Social counter-media,* that orient themselves in functionality or design towards specific mainstream offerings while positioning themselves as a "safe haven" for those who feel ostracized by this mainstream. (2) *Context-bound alternative social media* which resemble the more familiar applications in functionality or design but often originate from non-US markets and thus orient towards other regulatory frameworks. (3) *Fringe communities* that do neither aim at imitating mainstream social media nor pleasing a mainstream audience but instead address specific subcultures. The last one were dark channels, namely (4) *private-first channels*, such as instant messengers, that were often founded with the attempt to mediate interpersonal, private communication.

Relying on the *theory of planned behavior* (Ajzen 1991), we suggested that the self-positioning of social media (e.g., as social counter-medium or fringe community) attracts users with matching attitudes. Further, we argued that injunctive norms postulated by the platforms but also the descriptive norms through the observation of others' behaviors on said platform (e.g., the share of dark participation) would shape users' behavior on that platform. Relying on the SIDE-Model (Postmes et al. 1998), we expected that particularly pseudonymity on the

platform would motivate an orientation on the descriptive norms on that platform. Finally, we suggested that the technological affordances of a given application (e.g., the use of recommendation algorithms) would shape concrete behavior in said context.

With the help of a detailed platform analysis of both mainstream and dark social media in Germany, as well as a scoping review on extremism and (dark) social media, we described the different types of dark social media as opportunity structures for extremism. The analyses showed that the social counter-media Gab, Bitchute, and DLive offered new opportunity structures for extremist ideologues to disseminate propaganda, conspiracy theories, or hate speech, and allowed their sympathizers to consume or redistribute corresponding content (dark reception, so to speak). Due to the high level of pseudonymity, conformity to the descriptive norms shaped by dark participation was likely to influence new users too. Fringe communities such as 4Chan and 8Kun were less suited to "preach" one's deviant worldview than social counter-media. However, through their high level of anonymity and explicit toleration of deviant content including gore violence, they provided ideal opportunity structures for own dark participation (and other behavior consistent with the displayed rules on the respective board). Noteworthy, the user base of fringe communities is small.

Slightly differentially, context-bound alternative social media do address a more mainstream audience and have a much larger reach than platforms staging themselves as "safe haven" for deviant positions. Although context-bound alternative social media also host dark participation and even extremism it remains unclear whether the share of problematic content differs from the share observed on more prominent US-based platforms. Noteworthy, both investigated platforms, VKontakte and TikTok, seem to have increasingly close ties to their respective governments in Russia and China that might influence their activities in the future even more. Overall, context-bound alternative news offered more-or-less the same opportunity structures for extremism as their more famous counterparts.

At the time of data collection, private-first channels such as instant messengers were the most prominent social media applications in Germany. Our examination of Telegram, Signal, and Threema showed that all three promoted privacy-oriented, secure interpersonal communication. However, particularly Telegram combined these features with a technical infrastructure clearly striving for large-scale group and even mass communication. Telegram furthermore declined cooperation with state-authorities and promoted a high number of features that might be of interest for extremists. Particularly public Telegram channels have been described as safe haven for extremists and terrorists since the most active period of the self-declares 'Islamic state' and although many official ISIS-channels have been banned since, extremists with various ideological stances still thrive on the platform. The convergence of public and private channels on Telegram is high, potentially allowing extremists to gain new followers and immediately lure them into group-conversations or intimate private

chats on the same platform. Overall, private-first channels are likely particularly suited to enable intragroup and interpersonal exchange.

Our study had several meaningful theoretical implications. First, the consideration of digital technologies as opportunity structures has been proven useful. Using a broad lens to examine the interplay of regulatory attempts, social media genre and type and how that interacts with factors shaping individual behavior provided a coherent picture of the changing digital realm. Second, the theory of planned behavior provided a helpful lens through which social media as socio-technological systems can be understood. Different social media invite different types of users through their self-positioning, they enforce different norms through their terms of services/netiquette and through their concrete moderation action and cooperation with law enforcement. Finally, different genres of social media, as blurry as the borders between these genres can be, provide distinct technological affordances that allow for distinct types of actions. Most plausible, considering the interplay between these different levels allows more precise estimates on how extremists might use a specific social medium, where dark participation might flourish and thus also provide meaningful starting points for prevention.

Notwithstanding, our study had several limitations that must be considered. First, we focused only on a small selection of dark social media. Although our platform analysis covered different theoretically meaningful types of dark social media, future research expanding our work and providing a more dynamic monitoring of the rapid changing social media landscape is needed. In a related vein, we took a very euro-centric perspective in this paper. What is considered alternative in Germany must by no means be alternative in another context such as the global east or south. Consequentially, comparative research that tries to understand how platforms emerge in different systems and carter to different audiences is needed to provide a more balanced picture of our globalized digital world.

We did not measure dark participation on the examined platforms directly but used a scoping review of prior work to gauge the prevalence of deviant content. Due to different definitions and methodological approaches this approach is inevitably very broad and limited. Although we did find some comparative work allowing us to induct the relative share of dark participation on different types and genres of dark social media, future research using transparent, uniform frameworks comparing multiple genres and types of social media is needed for reliable conclusions. Relatedly, we did not test the effects of using different platforms (neither did we find studies who did so). It is thus currently not empirically justified to draw conclusions about the effects dark social media use has on radicalization and extremism. However, there is work showing that the exposure to hate speech, particularly over time, can impair intergroup relations (Bilewicz & Soral 2020; Ziegele et al. 2018), and that consuming conspiracy theories can diminish trust in democratic institutions (Einstein & Glick 2015; Kim & Cao 2016; Pummerer et al. 2021), reduce the willingness for democratic participation (Jolley & Douglas 2014) and the acceptance of non-normative and violent collective action

(Lamberty & Leiser 2019). Here we summarized evidence showing that some types of alternative social media thrive on this kind of content.

Our study also had some meaningful practical implications. First, a successful strategy against extremism in the digital realm must take the constant shift in digital opportunity structures into account and ensure that regulations account for the emergence of new opportunity structures. Second, social norms play a crucial role for human behavior. Thus, socio-technological systems should provide ideal opportunity structures for democratic participation by design. Further, in times of our high-choice media environment (Van Aelst et al. 2017), regulations of single platforms likely motivate a "whack-a-mole" game in which media users seeking dark participation turn towards social counter-media and fringe communities because their needs are satisfied in these spaces. Finally, prevention actors targeting a mainstream audience likely find their audience on mainstream or context-bound alternative social media (for detailed discussion, see Frischlich et al. 2022). In contrast, those aiming at a monitoring of dark participants and extremists might also need to examine alternative social media closely.

6 References

AFG, & ARG. (2021, December 13). Kreml erhält Einfluss auf Online-Netzwerk. *www.t-online.de*. https://www.t-online.de/-/100003878

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179–211.

Al-Rawi, A., & Groshek, J. (2018). Jihadist propaganda on social media: An examination of ISIS-related content on twitter. *International Journal of Cyber Warfare and Terrorism, 8*(4), 1–15. https://doi.org/10/ghkm4q

Amarasingam, A., Maher, S., & Winter, C. (2021). *How Telegram disruption impacts Jihadist platform migration*. Lancaster: Centre for Research and Evidence on Security Threats. https://crestresearch.ac.uk/resources/how-telegram-disruption-impacts-jihadist-platform-migration/

Askanius, T. (2021). On frogs, monkeys, and execution memes: Exploring the humor-hate nexus at the intersection of neo-nazi and alt-right movements in Sweden. *Television & New Media, 22*(2), 147–165. https://doi.org/10.1177/1527476420982234

Baden, C., & Sharon, T. (2021). Blinded by the lies? Toward an integrated definition of conspiracy theories. *Communication Theory, 31*(1), 82–106. https://doi.org/10.1093/ct/qtaa023

Baele, S. J., Brace, L., & Coan, T. G. (2020). Uncovering the far-right online ecosystem: An analytical framework and research agenda. *Studies in Conflict & Terrorism*, 1–21. https://doi.org/10.1080/1057610X.2020.1862895

Bartlett, J., & Miller, C. (2010). *The power of unreason. Conspiracy theories, extremism and counter-terrorism*. London: Demos.

Beisch, V. N., & Koch, W. (2021). 25 Jahre ARD/ZDF-Onlinestudie: Unterwegsnutzung steigt wieder und Streaming/ Mediatheken sind weiterhin Treiber des medialen Internets [25years of the ARD/ZDF Study: Mobile use increasing again, streaming/media libraries are drivers of the mobile Internet]. *Media Perspektiven, 10*, 486–503.

Berger, J. M. (2016). Nazis vs. ISIS on Twitter: A comparative study of White nationalist and ISIS online social media networks (Occasional Papers, September). Washington D.C.: George Washington University: Program on Extremism.

Bhat P., & Klein O. (2020). Covert hate speech: White nationalists and dog whistle communication on Twitter. In G. Bouvier & J. Rosenbaum (eds.), *Twitter, the Public Sphere, and the Chaos of Online Deliberation* (S. 151–172). Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-41421-4_7

Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology, 41*(1), 3–33. https://doi.org/10.1111/pops.12670

Bloom, M., Tiflati, H., & Horgan, J. (2019). Navigating ISIS's preferred platform: Telegram. *Terrorism and Political Violence, 31*(6), 1242-1254. https://doi.org/10/gf3gp8

Bouko, C., Naderer, B., Rieger, D., Ostaeyen, P. V., & Voué, P. (2021). Discourse patterns used by extremist Salafists on Facebook: Identifying potential triggers to cognitive biases in radicalized content. *Critical Discourse Studies,* 1–22. https://doi.org/10.1080/17405904.2021.1879185

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210–230. https://doi.org/10/gzn

Breves, P., Amrehn, J., Heidenreich, A., Liebers, N., & Schramm, H. (2021). Blind trust? The importance and interplay of parasocial relationships and advertising disclosures in explaining influencers' persuasive effects on their followers. *International Journal of Advertising, 40*(7), 1–20. https://doi.org/10.1080/02650487.2021.1881237

Brosius, H. B., Haas, A., & Koschel, F. (2016). *Methoden der empirischen Kommunikationsforschung: Eine Einführung [Methods of empirical communication science: An introduction]* (7th edition). Wiesbaden: Springer Fachmedien VS.

Brown, W. J. (2021). Involvement with media personae and entertainment experiences. In P. Vorderer & C. Klimmt (Hg.), *The Oxford handbook of entertainment theory* (pp. 285–305). Oxford: Oxford University Press.

Chandrasekharan, E., Pavalanathan, U., Srinivasan, A., Glynn, A., Eisenstein, J., & Gilbert, E. (2017). You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. In Association for Computing Machinery (ed.), *Proceedings of the ACM on Human-Computer Interaction, Volume 1, Issue CSCW (pp.* 1–22). https://doi.org/10.1145/3134666

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology, 58*, 1015–1026.

Clifford, B. (2018). „Trucks, knives, bombs, whatever:" Exploring pro-Islamic state instructional material on Telegram. *CTC S, 11*(5), 23–29.

Comerford, M. (2021, March 24). Two years on: Understanding the resonance of the Christchurch attack on imageboard sites. *Global Network on Extremism & Technology*. https://gnet-research.org/2021/03/24/two-years-on-understanding-the-resonance-of-the-christchurch-attack-on-imageboard-sites/

Conway, M., Khawaja, M., Lakhani, S., Reffin, J., Robertson, A., & Weir, D. (2019). Disrupting daesh: Measuring takedown of online terrorist material and its impacts. *Studies in Conflict & Terrorism, 42*(1–2), 141–160. https://doi.org/10.1080/1057610X.2018.1513984

Costello, M., Barrett-Fox, R., Bernatzky, C., Hawdon, J., & Mendes, K. (2020). Predictors of viewing online extremism among america's youth. *Youth & Society, 52*(5), 710–727. https://doi.org/10.1177/0044118X18768115

Costello, M., Hawdon, J., Ratliff, T., & Grantham, T. (2016). Who views online extremism? Individual attributes leading to exposure. *Computers in Human Behavior, 63*, 311–320. https://doi.org/10.1016/j.chb.2016.05.033

Crawford, B., Keen, F., & Suarez-Tangil, G. (2021). Memes, radicalisation, and the promotion of violence on chan sites. In Association for the Advancement of Artificial Intelligence (ed.), *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 982-991), Palo Alto: AAAI Press. https://ojs.aaai.org/index.php/ICWSM/article/view/18121

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgement. *Journal of abnormal psychology, 51*(3), 629–636. https://doi.org/10/bdjjcf

Einstein, K. L., & Glick, D. M. (2015). Do I think BLS data are BS? The consequences of conspiracy theories. *Political Behavior, 37*(3), 679–701. https://doi.org/10.1007/s11109-014-9287-z

Eisenbrand, R. (2021, December 23). Ranking: Das sind die Influencer mit den meisten Followern und Views auf Tiktok [Ranking: these are the influences with the most followers and views on TikTok]. *Daily*. https://omr.com/de/tik-tok-top-20-influencer/

Elley, B. (2021). "The rebirth of the West begins with you!"—Self-improvement as radicalisation on 4Chan. *Humanities and Social Sciences Communications, 8*(1), 1–10. https://doi.org/10.1057/s41599-021-00732-x

Ellison, N., Lampe, C., & Steinfield, C. (2009). Social network sites and society: Current trends and future possibilities. *Interactions, 16*, 6–9. https://doi.org/10/dzz32k

Emmer, M., & Strippel, C. (2015). Stichprobenziehung für Online-Inhaltsanalysen: Suchmaschinen und Filter Bubbles [Sampling for online-content analyses: Search engines and filter bubbles]. In A. Maireder, J. Ausserhofer, C. Schumann, & M. Taddicken (Hg.), *Digitale Methoden in der Kommunikationswissenschaft* (S. 275–300). Berlin: Digital Communication Research. https://doi.org/10.17174/dcr.v2.12

Erjavec, K., & Kovačič, M. P. (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. https://doi.org/10/gfgnmm

Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication, 22*(1), 35–52. https://doi.org/10.1111/jcc4.12180

Faddoul, M., Chaslot, G., & Farid, H. (2020). *A longitudinal analysis of YouTube's promotion of conspiracy videos*. ArXiv:2003.03318 [Cs]. http://arxiv.org/abs/2003.03318

Farkas, J., Schou, J., & Neumayer, C. (2018). Cloaked facebook pages: Exploring fake Islamist propaganda in social media. *New Media & Society, 20*(5), 1850–1867. https://doi.org/10/gc92tw

Fischer, P., & Greitemeyer, T. (2010). A new look at selective-exposure effects: An Integrative Model. Current Directions in Psychological Science, 19(6), 384–389. https://doi.org/10.1177/0963721410391246

Frischlich, L. (2018). Propaganda3: Einblicke in die Inszenierung und Wirkung von Online-Propaganda auf der Makro-Meso-Mikro Ebene [Propaganda3: Insights into the staging and effects of online-propaganda on the macro-meso-micro level. In K. Sachs-Hombach & B. Zywietz (Hg.), *Fake-News, Hashtags & Social Bots: Neue Methoden der populistischen Propaganda* (S. 133–170). Wiesbaden: Springer Fachmedien VS. https://doi.org/10.1007/978-3-658-22118-8

Frischlich, L. (2021). #dark inspiration: Eudaimonic entertainment in extremist instagram posts. *New Media & Society, 23*(3), 554–577. https://doi.org/10/gghnhr

Frischlich, L., Schatto-Eckrodt, T., & Völker, J. (2022). *Rückzug in die Schatten? Die Verlagerung digitaler Foren zwischen Fringe Communities und „Dark Social" und ihre Implikationen für die Extremismusprävention* [*Withdrawal to the shadows? The shifting of digital forums between fringe communities and „dark social" and the implications for extremism prevention*] (CoRE-NRW Expert Report, No. 4). Bonn: CoRE-NRW. https://www.bicc.de/publications/publicationpage/publication/rueckzug-in-die-schatten-die-verlagerung-digitaler-foren-zwischen-fringe-communities-und-dark-so/

Gibson, A. (2019). Free speech and safe spaces: How moderation policies shape online discussion spaces. *Social Media + Society, 5*(1). https://doi.org/10.1177/2056305119832588

Gibson, J. J. (1979). *The ecological approach to visual perception: Classic edition* (2014 edition). New York: Psychology Press. https://doi.org/10.4324/9781315740218

Guhl, J., & Davey, J. (2020). *A safe space to hate: White supremacist mobilisation on telegram*. London: Institute for Strategic Dialouge. https://www.isdglobal.org/isd-publications/a-safe-space-to-hate-white-supremacist-mobilisation-on-telegram/

Guhl, J., Ebner, J., & Rau, J. (2018). *The online ecosystem of the German far-right*. London: Institute for Strategic Dialogue.

Hamid, N., Pretus, C., Atran, S., Crockett, M. J., Ginges, J., Sheikh, H., Tobeña, A., Carmona, S., Gómez, A., Davis, R., & Vilarroya, O. (2019). Neuroimaging 'will to fight' for sacred values: An empirical case study with supporters of an Al Qaeda associate. *Royal Society Open Science, 6*(6), 181585. https://doi.org/10.1098/rsos.181585

Heinderyckx, F. (2015). Gatekeeping theory redux. In T. P. Vos & F. Heinderyckx (ed.), *Gatekeeping in Transition* (pp. 253–268). London: Routledge.

Hern, A. (2019, September 25). Revealed: How TikTok censors videos that do not please Beijing. *The Guardian*. https://www.theguardian.com/technology/2019/sep/25/revealed-how-tiktok-censors-videos-that-do-not-please-beijing

Holt, K., Ustad Figenschou, T., & Frischlich, L. (2019). Key-dimensions of alternative news media. *Digital Journalism, 7*(7), 860–869. https://doi.org/10.1080/21670811.2019.1625715

Holzer, B. (2021). Zwischen Protest und Parodie: Strukturen der „Querdenken"-Kommunikation auf Telegram (und anderswo) [Between parody and protest: Structures of „Querdenken" Communication on Telegram (and beyond)]. In S. Reichardt (Ed.), *Die Misstrauensgemeinschaft der Querdenker. Die Corona-Proteste aus kultur- und sozialwissenschaftlicher Perspektive [The distrust community of the Cross-thinkers. Coronaprotests from a cultural and social scientific perspective]*. Frankfurt a. M.: Campus Verlag.

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry, 19*(3), 215–229. https://doi.org/10.1080/00332747.1956.11023049

Hoseini, M., Melo, P., Benevenuto, F., Feldmann, A., & Zannettou, S. (2021). *On the globalization of the Qanon conspiracy theory through telegram*. arXiv:2105.13020 [cs]. http://arxiv.org/abs/2105.13020

Hsueh, M., Yogeeswaran, K., & Malinen, S. (2015). "Leave your comment below": Can biased online comments influence our own prejudicial attitudes and behaviors? *Human Communication Research, 41*(4), 557–576. https://doi.org/10.1111/hcre.12059

Inquirer. (2016). Telegram's secret chat-bots boon for ISIS. http://technology.inquirer.net/46297/telegrams-secret-chats-bots-boon-for-isis

Jasser, G., McSwiney, J., Pertwee, E., & Zannettou, S. (2021). 'Welcome to #GabFam': Far-right virtual community on Gab. *New Media & Society*. https://doi.org/10.1177/14614448211024546

Jhaver, S., Boylston, C., Yang, D., & Bruckman, A. (2021). Evaluating the effectiveness of deplatforming as a moderation strategy on twitter. In Association for Computing Machinery *(ed.), Proceedings of the ACM on Human-Computer Interaction, Volume 5*, *Issue CSCW2 (pp.* 2-30). New York. https://dl.acm.org/doi/abs/10.1145/3479525

Johnson, N. F., Leahy, R., Restrepo, N. J., Velasquez, N., Zheng, M., Manrique, P., Devkota, P., & Wuchty, S. (2019). Hidden resilience and adaptive dynamics of the global online hate ecology. *Nature, 573*(7773), 261–265. https://doi.org/10/ghhwn2

Jolley, D., & Douglas, K. M. (2014). The social consequences of conspiracism: Exposure to conspiracy theories decreases intentions to engage in politics and to reduce one's carbon footprint. *British Journal of Psychology, 105*(1), 35–56. https://doi.org/10/ggvqh3

Jünger, J., & Gärtner, C. (2020). *Datenanalyse von rechts-verstoßenden Inhalten in Gruppen und Kanälen von Messengerdiensten am Beispiel Telegram [Data analysis of law-violating content in groups and channels of messengers using the example of Telegram].* Landesanstalt für Medien NRW.

Kashpur, V. V., Myagkov, M., Baryshev, A. A., Goiko, V. L., & Shchekotin, E. V. (2020). Where Russian online nationalists go when their communities are banned: A case study of Russian nationalism. *Nationalism & Ethnic Politics, 26*(2), 145–166. https://doi.org/10.1080/13537113.2020.1751921

Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and gratifications research. *The Public Opinion Quarterly, 37*(4), 509–523.

Kicinski, M., Springate, D. A., & Kontopantelis, E. (2015). Publication bias in meta-analyses from the Cochrane Database of Systematic Reviews. *Statistics in Medicine, 34*(20), 2781–2793. https://doi.org/10.1002/sim.6525

Kim, M., & Cao, X. (2016). The impact of exposure to media messages promoting government conspiracy theories on distrust in the government: Evidence from a two-stage randomized experiment. *International Journal of Communication, 10*, 3808–3827.

Kim, S., & Kim, J. (2021). *Propagation of the QANON conspiracy theory on Facebook*. OSF Preprints. https://doi.org/10.31219/osf.io/wku5b

Kimball, W. (2021, September 2). Gab to congress: You're gonna need a warrant for that or you know, talk to, or do something. *Gizmodo*. https://gizmodo.com/gab-to-congress-you-re-gonna-need-a-warrant-for-that-o-1847607826

Lamberty, P., & Leiser, D. (2019). *»Sometimes you just have to go in « – The link between conspiracy beliefs and political action* [Preprint]. https://doi.org/10.31234/osf.io/bdrxc

Lang, R. (2021, August 20). TikTok—Chinesische Regierung beteiligt sich an ByteDance [TikTok: Chinese government aquires a stake of ByteDance]. *netzpolitik.org*. https://netzpolitik.org/2021/tiktok-chinesische-regierung-beteiligt-sich-an-bytedance/

Lima, L., Reis, J. C. S., Melo, P., Murai, F., Araújo, L., Vikatos, P., & Benevenuto, F. (2018). Inside the right-leaning echo chambers: Characterizing gab, an unmoderated social system. arXiv:1807.03688 [cs]. http://arxiv.org/abs/1807.03688

Manakas, M. (2021, January 10). Sturm auf US-Kapitol: Parler, Gab, DLive – wo sich die Angreifer organisierten [Storm on the US-capitol: Parler, Gab, DLive- where the aggressors organized]. *der Standard*. https://www.derstandard.at/story/2000123143883/sturm-auf-das-kapitol-parler-gab-dlive-wo-sich-die

Manning, M. (2009). The effects of subjective norms on behaviour in the theory of planned behaviour: A meta-analysis. *British Journal of Social Psychology, 48*(4), 649–705. https://doi.org/10.1348/014466608X393136

Markert, R. (2020, March 12). Netz-DG: Vorbild für repressive Regierungen weltweit? [Netz-DG: Model for repressive governments worldwide?] *Süddeutsche.de*. https://www.sueddeutsche.de/digital/netz-dg-internetzensur-facebook-1.4840302

Mathew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2019). Spread of hate speech in online social media. In Association for Computing Machinery (ed.), *WebSci '19: Proceedings of the 10th ACM Conference on Web Science, June 2019* (pp. 173–182). https://doi.org/10.1145/3292522.3326034

Mathew, B., Illendula, A., Saha, P., Sarkar, S., Goyal, P., & Mukherjee, A. (2019). Temporal effects of unmoderated hate speech in Gab. arXiv:1909.10966 [cs]. http://arxiv.org/abs/1909.10966

Miller-Idriss, C. (2020). Hate in the homeland. Princeton: Princeton University Press. https://doi.org/10.1515/9780691205892

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC *Medical Research Methodology, 18*(1), 143. https://doi.org/10.1186/s12874-018-0611-x

Myagkov, M., Chudinov, S. I., Kashpur, V. V., Goiko, V. L., & Shchekotin, E. V. (2020). Islamist communities on VKontakte: identification mechanisms and network structure. *Europe-Asia Studies, 72*(5), 863–893. https://doi.org/10.1080/09668136.2019.1694645

Myagkov, M., Shchekotin, E. V., Chudinov, S. I., & Goiko, V. L. (2020). A comparative analysis of right-wing radical and Islamist communities' strategies for survival in social networks (evidence from the Russian social network VKontakte). Media, War & Conflict, 13(4), 425–447. https://doi.org/10.1177/1750635219846028

NetzDG, (2017). https://www.BMJV.de/DE/Themen/FokusThemen/NetzDG/NetzDG_node.html

Newman, N., Fletcher, R., Schulz, A., Simge, A., Robertson, C. T., & Kleis Nielsen, R. (2021). *Reuters digital news report 2021*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2021-06/Digital_News_Report_2021_FINAL.pdf

O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2013). Uncovering the wider structure of extreme right communities spanning popular online networks. In Association for Computing Machinery (ed.), WebSci '13: Proceedings of the 5th Annual ACM Web Science Conference May 2013 (pp. 276–285). https://doi.org/10/gf3hcz

O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2015). Down the (White) rabbit hole: The extreme right and online recommender systems. *Social Science Computer Review, 33(*4), 1–20. https://doi.org/10.1177/0894439314555329

O'Connor, C. (2021). *Hatescape: An in-depth analysis of extremism and hate speech on tiktok*. London: Institute for Strategic Dialogue.

Pfeiffer, T. (2002). „Das Internet ist billig, schnell und sauber. Wir lieben es", Rechtsextremisten entdecken den Computer ["The Internet is cheap, fast, and clean. We love it". Right-wing extremists discover the computer.]. In Bundeszentrale für politische Bildung (ed.), *Rechtsextremismus im Internet. Recherchen, Analysen, pädagogische Modelle zur Auseinandersetzung mit dem Rechtsextremismus*, Bonn.

Phillips, W. M. (2012). This is why we can't have nice things: The origins, evolution and cultural embeddedness of online trolling [Ph.D.]. https://search.proquest.com/docview/1237277556/abstract/E53C6686D73D419BPQ/1

Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries? SIDE-effects of computer-mediated communication. *Communication Research, 25*(6), 689–715. https://doi.org/10/ffsbdn

Poupin, P. (2021). Social media and state repression: The case of VKontakte and the anti-garbage protest in Shies, in Far Northern Russia. *First Monday, 26*(5). https://doi.org/10.5210/fm.v26i5.11711

Prucha, N. (2015). IS and the Jihadist information highway – Projecting influence and religious identity via telegram. *Perspectives on Terrorism, 10*(3), 48–58.

Pummerer, L., Böhm, R., Lilleholt, L., Winter, K., Zettler, I., & Sassenberg, K. (2021). Conspiracy theories and their societal effects during the COVID-19 pandemic. *Social Psychological and Personality Science*. https://doi.org/10.31234/osf.io/y5grn

Quandt, T. (2018). Dark participation: Manipulative user engagement in the news making process. *Media and Communication, 6*(4), 36–48. http://dx.doi.org/10.17645/mac.v6i4.1519

Quandt, T., Klapproth, J., & Frischlich, L. (2022). Dark social media participation and well-being. *Current Opinion in Psychology, 45*(June 2022, 101284). https://doi.org/10.1016/j.copsyc.2021.11.004

Rathgeb, T., & Schmid, T. (2020). *JIM-Studie 2020 Jugend, Information, Medien: Basisuntersuchung zum Medienumgang 12-bis 19-Jähriger* [*JIM-Study 2020: Youth, information, media. Basic data on the media use of 12 to 19 years olds*]. Medienpädagogischer Forschungsverbund Südwest. https://www.mpfs.de/fileadmin/files/Studien/JIM/2020/JIM-Studie-2020_Web_final.pdf

Rauchfleisch, A., & Kaiser, J. (2020). The German far-right on YouTube: An analysis of user overlap and user comments. *Journal of Broadcasting & Electronic Media, 64*(3), 373–396. https://doi.org/10.1080/08838151.2020.1799690

Rieger, D., Frischlich, L., Rack, S., & Bente, G. (2020). Digitaler Wandel, Radikalisierungsprozesse und Extremismusprävention im Internet [Digital change, radicalisation processes and extremism prevention in the Internet]. In B. Ben Slama & U. Kemmesies (ed.), Handbuch Extremismusprävention (pp. 351–388). Wiesbaden: Bundeskriminalamt.

Rieger, D., Kümpel, A. S., Wich, M., Kiening, T., & Groh, G. (2021). Assessing the extent and types of hate speech in fringe communities: A case study of alt-right communities on 8chan, 4chan, and reddit. *Social Media*, 1–14.

Rivis, A., & Sheeran, P. (2003). Descriptive norms as an additional predictor in the theory of planned behaviour: A meta-analysis. *Current Psychology, 22*(3), 218–233. https://doi.org/10.1007/s12144-003-1018-2

Rogers, R. (2020). Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. *European Journal of Communication, 35*(3), 213–229. https://doi.org/10/ghjcqz

Rottweiler, B., & Gill, P. (2020). Conspiracy beliefs and violent extremist intentions: The contingent effects of self-efficacy, self-control, and law-related morality. *Terrorism and Political Violence*, 1–20. https://doi.org/10.1080/09546553.2020.1803288

Schindler, M., & Domahidi, E. (2021). The growing field of interdisciplinary research on user comments: A computational scoping review. *New Media & Society, 23*(8), 2474–2492. https://doi.org/10.1177/1461444821994491

Schmitt, J. B., Harles, D., & Rieger, D. (2020). Themen, Motive und Mainstreaming in rechtsextremen Online-Memes [Themes, motives, and mainstreaming in right-wing extremist online-memes]. *M&K Medien & Kommunikationswissenschaft*, *68*(1-2), 73–93. https://doi.org/10.5771/1615-634X-2020-1-2-73

Schmitt, J. B., Rieger, D., Rutkowski, O., & Ernst, J. (2018). Counter-messages as prevention or promotion of extremism?! The potential role of YouTube: Recommendation algorithms. *Journal of Communication, 68*(4), 780–808. https://doi.org/10.1093/joc/jqy029

Schneider, J., Schmitt, J. B., Ernst, J., & Rieger, D. (2019). Verschwörungstheorien und Kriminalitätsfurcht in rechtsextremen und islamistischen YouTube Videos [Conspiracy theories and fear of crime in right-wing extremist and Islamist YouTube videos]. *Praxis der Rechtspsychologie, 29*(1), 41–66.

Schwarzenegger, C., & Wagner, A. (2018). Can it be hate if it is fun? Discursive ensembles of hatred and laughter in extreme right satire on Facebook. *SCM Studies in Communication and Media, 7*(4), 473–498. https://doi.org/10.5771/2192-4007-2018-4-473

Scrivens, R., & Amarasingam, A. (2020). Haters gonna "like": Exploring Canadian far-right extremism on Facebook. In M. Littler & B. Lee (ed.), *Digital Extremisms: Readings in Violence, Radicalisation and Extremism in the Online Space* (pp. 63–89). Springer International Publishing. https://doi.org/10.1007/978-3-030-30138-5_4

sede, & guth. (2021, June 9). Iphone- und Google-App: Zugang zu Telegram-Kanälen von Attila Hildmann gesperrt [IPhone and Google app: access to Telegram channels of Attila Hildmann blocked]. *FAZ.NET*. https://www.faz.net/aktuell/gesellschaft/kriminalitaet/zugang-zu-telegram-kanaelen-von-attila-hildmann-gesperrt-17380316.html

Shoemaker, P. J., & Vos, T. (2009). *Gatekeeping Theory*. London: Routledge.

Stern, J., & Berger, J. M. (2016). *ISIS: The State of Terror* (2nd ed.). New York: Ecco.

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication, 60*(3), 556–576. https://doi.org/10/d6p3kx

Sutton, A. J. (2000). Empirical assessment of effect of publication bias on meta-analyses. *BMJ, 320*(7249), 1574–1577. https://doi.org/10.1136/bmj.320.7249.1574

Swann, W. B., Buhrmester, M. D., Gómez, A., Jetten, J., Bastian, B., Vázquez, A., Ariyanto, A., Besta, T., Christ, O., Cui, L., Finchilescu, G., González, R., Goto, N., Hornsey, M., Sharma, S., Susianto, H., & Zhang, A. (2014). What makes a group worth dying for? Identity fusion fosters perception of familial ties, promoting self-sacrifice. *Journal of Personality and Social Psychology, 106*(6), 912–926. https://doi.org/10.1037/a0036089

Taddicken, M., & Schmidt, J.-H. (2017). Entwicklung und Verbreitung sozialer Medien [Development and dissemination of social media]. In J.-H. Schmidt & M. Taddicken (ed.), *Handbuch Soziale Medien* (pp. 3-22.). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-03765-9_1

Tarrow, S. (1988). National politics and collective action: Recent theory and research in Western Europe and the United States. *Annual Review of Sociology, 14*, 421–440.

Teare, G., & Taks, M. (2020). Extending the scoping review framework: A guide for interdisciplinary researchers. *International Journal of Social Research Methodology, 23*(3), 311–315. https://doi.org/10.1080/13645579.2019.1696092

Torba, A. (2021, August 5). Advertising on Gab. *Gab News*. https://news.gab.com/2021/08/05/advertising-on-gab/

Trujillo, M., Gruppi, M., Buntain, C., & Horne, B. D. (2020). What is BitChute? Characterizing the "free speech" alternative to YouTube. In Association for Computing Machinery (ed.), *Proceedings of the 31st ACM Conference on Hypertext and Social Media (pp.* 139–140). New York. https://doi.org/10.1145/3372923.3404833

Urman, A., & Katz, S. (2020). What they do in the shadows: Examining the far-right networks on Telegram. *Information, Communication & Society*, 1–20. https://doi.org/10.1080/1369118X.2020.1803946

Van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., de Vreese, C., Matthes, J., Hopmann, D., Salgado, S., Hubé, N., Stępińska, A., Papathanassopoulos, S., Berganza, R., Legnante, G., Reinemann, C., Sheafer, T., & Stanyer, J. (2017). Political communication in a high-choice media environment: A challenge for democracy? *Annals of the International Communication Association, 41*(1), 3–27. https://doi.org/10/f94bkm

van Prooijen, J.-W., Spardaro, J., & Wang, H. (2022). Suspicion of institutions: How distrust and conspiracy theories deteriorate social relationships | Elsevier Enhanced Reader. *Current Opinion in Psychology, 43*, 65–69. https://doi.org/10.1016/j.copsyc.2021.06.013

Walther, S., & McCoy, A. (2021). US extremism on telegram: Fueling disinformation, conspiracy theories, and accelerationism. *Perspectives on Terrorism, 15*(2), 100–124.

Weimann, G., & Masri, N. (2020). Research note: Spreading hate on TikTok. *Studies in Conflict & Terrorism*, 1–14. https://doi.org/10.1080/1057610X.2020.1780027

Wong, J. C. (2018, September 5). Don't give Facebook and YouTube credit for shrinking Alex Jones' audience | Julia Carrie Wong. *The Guardian*. http://www.theguardian.com/commentisfree/2018/sep/04/alex-jones-infowars-social-media-ban

Woolley, S. C., Pakzad, R., & Monaco, N. (2019). *Incubating Hate: Islamophobia and Gab.* German Marshall Fund of the United States. https://www.jstor.org/stable/resrep21229

Yayla, A. S., & Speckhard, A. (2017). *Telegram: the mighty application that ISIS Loves*. International Center for the Study of Violent Extremism. https://www.icsve.org/telegram-the-mighty-application-that-isis-loves/

Zannettou, S., Bradlyn, B., De Cristofaro, E., Kwak, H., Sirivianos, M., Stringhini, G., & Blackburn, J. (2018). What is Gab? A bastion of free speech or an alt-right echo chamber? In International World Wide Web Conferences Steering Committee (ed.), *WWW '18: Companion Proceedings of the The Web Conference 2018. Track: The Third International Workshop on Cybersafety, Online Harassment, and Misinformation* (pp. 1007–1014). Geneva. https://doi.org/10/gfzcbp

Zannettou, S., Caulfield, T., Blackburn, J., De Cristofaro, E., Sirivianos, M., Stringhini, G., & Suarez-Tangil, G. (2018). *On the origins of memes by means of fringe web communities*. ACM Internet Measurement Conference. http://arxiv.org/abs/1805.12512

Zeng, J., & Schäfer, M. S. (2021). Conceptualizing "dark platforms". Covid-19-related conspiracy theories on 8kun and gab. *Digital Journalism*, Ahead-of-print, 1–23. https://doi.org/10.1080/21670811.2021.1938165

Ziegele, M., Koehler, C., & Weber, M. (2018). Socially destructive? Effects of negative and hateful user comments on readers' donation behavior toward refugees and homeless persons. *Journal of Broadcasting & Electronic Media, 62*(4), 636–653. https://doi.org/10/gf8pn4

core-nrw

Netzwerk für Extremismusforschung
in Nordrhein-Westfalen

Connecting Research
on Extremism
in North Rhine-Westphalia

Impressum

Herausgeber und Kontakt
Maurice Döring
BICC · Pfarrer-Byns-Str. 1 · 53121 Bonn · Tel. +49 228.911 96-0
doering@core-nrw.de · www.core-nrw.de

Die Veröffentlichung erfolgt im Kontext des Netzwerkes CoRE-NRW, einem Verbund aus Wissenschaft und Praxis zur Erforschung des extremistischen Salafismus, des Rechtsextremismus und anderer Formen des Extremismus. Die Koordinierungsstelle am BICC arbeitet im Auftrag für das Ministerium für Kultur und Wissenschaft des Landes NRW. Die Inhalte der Publikation werden allein von den Autor:innen verantwortet. Sie spiegeln nicht die Position der Koordinierungsstelle oder des Ministeriums für Kultur und Wissenschaft des Landes NRW wider.

Forschungskontext
Dieses Forschungspapier steht in Verbindung mit dem CoRE-NRW Kurzgutachten 4, „Rückzug in die Schatten? Die Verlagerung digitaler Foren zwischen Fringe Communities und 'Dark Social' und ihre Implikationen für die Extremismusprävention", das von den Autor:innen im Auftrag von CoRE-NRW erstellt und im Januar 2022 veröffentlich wurde. Es präsentiert die dem Kurzgutachten zugrundeliegende Theorie und die dafür verwendete Methodik ausführlicher. Um den innovativen Ansatz in die international Fachdebatte einzuspeisen, erscheint dieses CoRE-NRW-Forschungspapier auf Englisch.

Autor:innen
Dr. Lena Frischlich leitet die Nachwuchsforschungsgruppe DemoRESILdigital: Demokratische Resilienz in Zeiten von Online-Propaganda, Fake News, Fear- und Hate Speech" am Institut für Kommunikationswissenschaft der Westfälischen Wilhelms-Universität Münster (WWU).
Tim Schatto-Eckrodt ist in der Nachwuchforschungsgruppe DemoRESILdigital tätig.
Julia Völker ist Kommunikationswissenschaftlerin und war bis 2021 am Institut für Kommunikationswissenschaft der WWU tätig.

Gestaltung
kippconcept gmbh, Bonn

Februar 2022