Adolescents and the dark side of social media—Law enforcement perspectives

Juho Äijälä, Reetta Riikonen, Aki-Mauri Huhtinen and
Teija Sederholm*

Department of Leadership and Military Pedagogy, Finnish National Defence University, Helsinki, Finland

Adolescents are the most active user group of social media sites. Due to being in a phase of both biological and psychological development, they may be particularly vulnerable to the darker side of social media, such as its illegal aspects or coordinated information influencing. With this research, we aimed to identify threats Finnish adolescents face on social media from a law-enforcement perspective. To reach this goal, we performed semi-structured interviews with police officers from Finnish preventive measures police units. To identify and structure threats that adolescents face, we employed a twofold analysis. In the first part, we conducted inductive content analysis, which revealed three primary threats: polarization, disinformation, and social media as a pathway to illegal activities. In the second part, we employed the Honeycomb-model of social media functionality as a classificatory device for structuring these threats. Our findings provide explorative insights into the threats social media might present to adolescents from the point of view of the Finnish law-enforcement system.

KEYWORDS

social media, adolescents, Honeycomb model, national security, threats, law enforcement

1 Introduction

The prevalence of different social networking sites (SNS) has risen significantly in the past decade. For example, in Finland, 69% of citizens aged 16–89 are registered on at least one social media platform [Official Statistics of Finland (OSF)., 2020]. Social media platforms are not merely places where users can freely connect, but multifaceted digital environments that possess various phenomena that can cause serious detrimental effects on the exposed user's personal life and the society they live in. The notable rise of SNS as a prominent media format and as a tool of interaction during the past decade has partly hindered the positivity that surrounded social media in its early years (Diamond and Plattner, 2012; Tucker et al., 2017). This has given rise to increasing concern about its adverse effects on the mental wellbeing of individuals, information space, and social stability (Loader and Mercea, 2011; Bossetta, 2018; Barberá, 2020). Negative effects of social media use and its malicious employment by users and organizations is well documented. Psychological literature has grown to include such concepts as "Facebook-depression" (Jelenchick et al., 2013; Yoon et al., 2019) with some research showcasing correlations between SNS use and depression, social anxiety, poor self-image, and eating disorders (Santarossa and Woodruff, 2017; Keles et al., 2020). Additionally, the excessive spread of misinformation and conspiracy theories (Allcott et al., 2019; Garry et al., 2021), information influencing (Ferrara, 2020;), the induction of filter bubbles (Bruns, 2019; Geschke et al., 2019), and the rise of political polarization (Howard et al., 2018; Tucker et al., 2018; Barberá, 2020) has sparked concern about the detrimental effects of SNS on western liberal democracies (Persily, 2017). Major international scandals

such as the involvement of the Russian Internet Research Agency (IRA) in the 2016 US Presidential Election (Bastos and Farkas, 2019) have raised concerns about the societal effects of SNS on various arenas, from academic research to political discussions. Recently, this has been particularly present in the discussion of COVID-19-related mis- and disinformation online (Pennycook et al., 2020). These factors can pose a significant threat to societal stability (Shu et al., 2020; McKay and Tenove, 2021) and individual wellbeing (Yoon et al., 2019; Keles et al., 2020). Thus, they have been characterized as the dark side of social media (Baccarella et al., 2018; Dhir et al., 2018; Demetis, 2020).

The current research aims to identify and structure some of the threats that the dark side of social media poses to Finnish adolescents from a law-enforcement perspective. We approached this issue by interviewing Finnish Preventive Measures Police Unit officers. Our analysis process is twofold. First, we employ data-driven content analysis (Krippendorff, 2018) to identify the most significant threats to adolescents on social media from the point of view of law-enforcement professionals. Secondly, we re-examine these threats by employing the Honeycomb model of social media functionality as a method of classification (Kietzmann et al., 2011; Baccarella et al., 2018). The model separates social media into seven functionalities, which we use to generate "threat profiles" for the identified threats. These profiles identify which functionalities of social media platforms help to give rise to which threats. Our research illustrates a method for breaking the complex threats posed by the dark side of social media into more manageable parts. Moreover, our approach provides novel insights about these threats due to our specific focus on police officers, who possess distinct professional knowledge to inspect social media from the point of view of national security.

2 Adolescent and the dark side of social media

Despite a significant rise of middle-aged and elderly users on social media, adolescents and young adults remain the largest and most active group, with three-quarters of Finns aged 16–24 using social media daily (Official Statistics of Finland (OSF)., 2020). Adolescence is a significant period of change and development both physically and psychologically. For instance, the dual-systems theory proposes that the adolescent brain matures sequentially, with the socioemotional system, which drives incentive processing and increases the tendency to seek exciting and risky behaviors, developing significantly before the prefrontal cortices of the brain driving cognitive control (Steinberg, 2010; Shulman et al., 2016). These distinct patterns of neural development render adolescents more likely to partake in risky behaviors and more vulnerable to emotional events via increased reward-seeking and a lack of cognitive control over their impulses (Steinberg, 2010). Moreover, adolescence is a period of identity formation (Klimstra et al., 2010) during which peer pressure has increased influence (Vargas, 2011). Hence, the dark side of social media may pose more significant threats to adolescents than to the adult population.

Willoughby (2019) has identified some of the main risks that adolescents face on social media: cyberbullying, online abuse, and exposure to negative content (i.e., violent or inappropriate content).

Moreover, many other researchers have raised concerns about the prevalence of disinformation, coordinated information influencing, and political polarization present on online platforms (Lazer et al., 2018; Tucker et al., 2018; Hills, 2019; Sheldon et al., 2019; Baccarella et al., 2020; Norri–Sederholm et al., 2020; Shu et al., 2020). Overall, the negative attributes of social media sites can generate an environment within the larger social media sphere that is capable of inducing various adverse effects on its users (Dhir et al., 2018; Keles et al., 2020). This can be regarded as the dark side of social media (Baccarella et al., 2018; Sheldon et al., 2019). Understanding the nature and prevalence of these malevolent aspects of social media is vital for understanding the threats it poses to adolescents and to society.

Next, we will outline and characterize some of the principal threats adolescents face on social media based on previous research. After this, we will describe the framework we will employ to analyse threats on social media. Then, the rationale and analysis of the current research will be presented.

2.1 Polarization

The polarization of digital environments and the possible presence of so-called 'filter bubbles' has received widespread attention both within academic circles and within public discourse (Howard et al., 2018; Tucker et al., 2018; Bruns, 2019; Geschke et al., 2019; Barberá, 2020). The studies posit that social media platforms fragment the information ecosystem causing different ideological groups to collectively get their information from different sources— a phenomenon often referred to as filter bubbles (Barberá et al., 2017; Barberá, 2020). This fragmentation of information is thought to be reflected in both built-in algorithms of social media (Levy, 2021) and innate human biases (Zollo, 2019; Moore-Berg et al., 2020; Rathje et al., 2021). Overall, these factors can cause the information environments of social media users to become homogenized significantly reducing the amount of information that counters or contradicts the users' pre-existing beliefs. This has been hypothesized to lead to extreme political polarization, radicalization, and belief in conspiracy theories (Tucker et al., 2018; Garcet, 2021; Garry et al., 2021).

However, some research has questioned the significance of filter bubbles (Zuiderveen Borgesius et al., 2016) suggesting that other factors could cause polarization on social media possibly due to human biases (Zollo, 2019; Moore-Berg et al., 2020). More importantly, while much effort has been focused on the effects and causes of political polarization (Barber and McCarty, 2015; McCarty, 2019; Moore-Berg et al., 2020), significantly less is known about how polarization spreads to the lives of adolescents.

2.2 Dis- and misinformation

This paper employs a representational definition of information (Floridi, 2013). According to this definition, information has semantic content about the world: "the sky is blue" represents a known fact about the world, whereas "the sky is red" contains information that is not in line with the physical reality, that

is misinformation. Although epistemologically simplistic, this definition is a pragmatic common-sense definition that can be easily conveyed to people not immersed in philosophical debate. False information can be divided into disinformation and misinformation. Disinformation is defined as intentionally shared false information, whereas misinformation is false information that is shared without an intention to deceive people (Fetzer, 2004; Lazer et al., 2018; Norri–Sederholm et al., 2020). Disinformation is often planned, and the target audience is well-defined and chosen to maximize the effect of the presented material. Dis- and misinformation are plentiful on social media (Tucker et al., 2018; Suarez-Lledo and Alvarez-Galvez, 2021) and are actively used to influence the behaviors, identities, and values of social media users (Bastick, 2021; Dobber et al., 2021).

The spread of disinformation is often coordinated by multiple operators that are difficult to track without technical tools to model pathways of information spread within the information ecosystem (Hussain et al., 2018; Bandeli and Agarwal, 2021). Thus, the presence and spread of disinformation on social media platforms can have severe detrimental effects on individual and societal levels (Wang et al., 2019; Shu et al., 2020). Firstly, consistent exposure to disinformation can reduce trust in institutions and the information ecosystem. Secondly, disinformation can reduce an individual's ability to distinguish between real and "fake news". Thirdly and most importantly, disinformation may lead to false beliefs which can significantly affect an individual's political and health-related behavior (Rapp and Salovich, 2018; Shu et al., 2020). Recently, this has been particularly prevalent as increased vaccine hesitancy due to COVID-19-related disinformation (Kricorian et al., 2021). Finally, given that disinformation often relies on emotionally manipulative information (Martel et al., 2020), adolescents may be especially vulnerable to it due to their underdeveloped systems of cognitive control (Shulman et al., 2016). However, the previous research is unclear about disinformation's effects on adolescents: while their brains are still developing, their experience of being indulged in the information ecosystem for most of their lives might give them unique skills and abilities in noticing and processing disinformation.

Past research has shown that Finnish adolescents are confident in their abilities to detect mis- and disinformation (Riikonen et al., 2020; Kaarkoski et al., 2021), but there is no research showing how much this corresponds to their actual ability. However, a recent large-scale nationally representative survey in the United States indicated that adolescents struggle to successfully evaluate the credibility of online information (measured by a selection of tasks requiring critical evaluation of online content). The authors found that 52% of high school students considered that an unofficial YouTube video showing CCTV footage from Russia constituted strong evidence for election fraud in the US 2016 Presidential Election (Breakstone et al., 2021). Similar worrying results underlining adolescents' poor social media literacy are reported by other studies (McGrew et al., 2018; Johnston, 2020) with a large literacy gap existing between different groups of adolescents (Nygren and Guath, 2021). However, in contrast to these findings, recent research has found a significant age effect of sharing "fake news" with users over the age of 65 sharing over seven times more news articles classified as fake than 18–28-year-olds

(Guess et al., 2019). This suggests that adolescents are not the most active sharers of misinformation and might be significantly better in detecting it that older generations.

2.3 Illegal content and activities

Adolescents are exposed to illegal activities and content on social media. This could be harmful in different ways. Firstly, adolescents might be the target of aggressive or illegal activities such as slander and cyberbullying. Secondly, they might be exposed to illegal or otherwise harmful contents. Thirdly, they might partake in these illegal and/or antisocial activities themselves through social media.

Adolescents can be targeted aggressively on social media, for example, via cyberbullying which can cross the line to illegal slander depending on the severity. The phenomenon of cyberbullying can be defined in various terms (Casas et al., 2020; Chun et al., 2020). It has been defined as "an aggressive, intentional act carried out by a group or individual, using electronic forms of contact, repeatedly and overtime against a victim who cannot easily defend him or herself" or "bullying through email and instant messaging, in a chat room, on a website…" (Kowalski and Limber, 2007; Chun et al., 2020). Oksanen et al. (2014) estimated that 67% of Finnish Facebook users aged 15–18 have been exposed to hateful online material, while 21% have been targeted by it. Näsi et al. (2014) performed a cross-national comparison of online abuse in the United States and Finland and reported that around 17% of Finnish respondents reported being subjected to online abuse or harassment (compared to 14% of American respondents). Notably, both studies found that being subjected to online abuse correlates with a significant reduction in subjective wellbeing (Näsi et al., 2014; Oksanen et al., 2014). While the authors note that no causal inferences can be drawn from this data, they also point out that some earlier research has established causal relationships between offline abuse and reduced wellbeing (Bucchianeri et al., 2014; Crowley and Cornell, 2020). Moreover, in a sizeable UK-cohort study, Kelly et al. (2018) found that being exposed to online harassment explained some of the variance in the depressive symptoms associated with social media use.

Overall, the literature suggests that large portions of the general population report being exposed to cyberbullying (Kowalski et al., 2019; Vidgen et al., 2019). Moreover, being targeted by online harassment and cyberbullying is associated with depressive symptoms (Keles et al., 2020; Kwan et al., 2020), relationship problems (Spears et al., 2009), and overall reduced wellbeing (Schenk and Fremouw, 2012; Maghsoudi et al., 2020). These adverse effects seem to be present regardless of cultural differences (Näsi et al., 2014), though slight differences between different cultural samples have been reported (Sorrentino et al., 2019).

Adolescents may also be exposed to harmful and illegal material on social media. This encompasses various types of content such as violent content (Friis, 2015; Boyd and Swanson, 2016), sexual material (Lewis et al., 2018; Smahel et al., 2020) and hate speech (Castano-Pulgarín et al., 2021). This inappropriate content is actively present in the digital lives of adolescents. An online survey of EU children (Smahel et al., 2020) found that on average 10% of

European adolescents had come across content showing ways to physically harm themselves, 11% had come across content featuring drug use, 17% had come across hate messages, and 13% had come across gory or violent images. In Finland, 18% had come across content showing ways to physically harm themselves, 10% had come across content featuring drug use, 17% had come across hate messages, and 11% had come across gory or violent images (Smahel et al., 2020). Exposure to violent, sexual or otherwise inappropriate content can present various effects on adolescents (Subrahmanyam and Smahel, 2010). Exposure to violent material can cause desensitization to real-world violence (Bushman and Anderson, 2009; Anderson et al., 2015) and have detrimental effects on mental health (Patton et al., 2014). In addition, Näsi et al. (2015) found that exposure to online hate material significantly reduced social trust among Finnish adolescents suggesting that it may threaten the feeling of societal security.

2.4 The Honeycomb model of social media functionality

The dark side of social media presents a wide field of scientific enquiry. Consequently, Baccarella et al. (2018) built on the so-called Honeycomb model of social media functionality (see Kietzmann et al., 2011) to formalize a framework for studying and structuring the negative aspects of social media platforms. While initially the model was primarily employed in marketing and managing research (Khan and Jan, 2015; Jayasuriya and Azam, 2017; Jayasuriya et al., 2018), recent proposals have suggested using it to study the dark side of social media (Baccarella et al., 2018). The logic behind this is sound; if different social media functionalities can be successfully employed for marketing purposes or to otherwise influence people, then the same modes can be used to influence people for more malicious purposes. The model identifies seven building blocks of functionality that can be malevolently employed to attack and manipulate users and spread disinformation: conversations, Sharing, Reputation, Groups, Identity, Relationships, and Presence. Noteworthily, the model does not claim to offer explanatory mechanisms for the phenomena present on social media, but instead provides a useful classification method.

Conversations is defined as the aspect by which users of social media platforms communicate with each other using, for example, built-in functions of the platforms such as "like", "comment", or "message". The dark side of these functionalities is the possibility of aggressive engagement, cyber-bullying, and spreading dis- and misinformation. Moreover, Baccarella et al. (2018) identifies the use of bots and other A.I. methods as a potential hazard and as a way of skewing and manipulating online conversations.

Sharing refers to the capability of users to share, distribute, and generate content, such as videos, memes, or pictures. A significant risk of this functionality is the availability of inappropriate content on social media platforms. This content includes violent, pornographic, illegal, or otherwise questionable material, fake news and disinformation.

Reputation reflects certain users' status within social media networks and how they can use their social standing to influence others. Moreover, it encompasses how users are aware of their status and the reputation to others. The status of users can be exhibited through the built-in functions of social media platforms, such as the number of friends and followers or the number of likes and shares their posts receive. According to Baccarella et al. (2018), reputation risks mainly stem from sharing false information that might damage or destroy the user's reputation—an event that can also have destructive effects on life outside social media.

Groups, in turn, relates to the degree to which users create groups and organize themselves around other users with similar interests, shared practices, and common worldviews. Polarization and the increased categorization of people as either in- or outgroup members are identified as significant risks of this functionality as social circles and groups define and narrow what kind of content the user consumes. Groups can also be pitched against each other. The groups are noted as creating social echo chambers, enhancing radicalization and polarization.

Identity captures how users display and develop their identities online. Users construct their identities by displaying their personal information, such as age, gender, or employment, participating in various groups, showcasing political opinions, and sharing content. Baccarella et al. (2018) note that other functionalities define the identities that users display online to some extent. They point out that this leads to a situation in which the user is not fully in control of their identity development online. This is problematic because the online identity can also influence the user's offline life. Major threats related to this functionality include shaming and the defamation of users' identities.

Relationships aims to encapsulate how users relate to other users on social media platforms and how this affects their behavior. For example, the relationship between two users affects the content they share with each other and the conversations they have. While social media can give rise to positive relationships, such as friendships and professional contacts, it can also induce negative relationships, including abuse, cyberbullying, stalking, scamming, or manipulation.

Presence considers how organizations and other users can track movement and presence online. This is present in both location-tracking and tracking when users use social media platforms. The negative aspects of this functionality revolve around the degree to which the presence and location of users can be tracked without their consent. For example, the recording of location data gives organizations the possibility to track users, but this extends to other parties as well.

While these functional blocks are not exhaustive, they can provide a prototypical model of the modes of influence that can be employed in social media (Baccarella et al., 2018). It is noteworthy that none of the functionalities fully encompass the threats but instead can be used to break down the phenomena present in social media into smaller parts.

3 Data and methods

3.1 Data collection

Our data were collected during December of 2019. Semi-structured interviews were employed to probe interviewees'

opinions on adolescents' social media use from the perspective of societal security. The interviewees took part in the interviews voluntarily. They were chosen based on their roles and professional expertise in Finnish Preventive Measures police units which focus on combating crime and improving public security. The interviewed police officers were recruited from three different police districts representing distinct geographical areas in Finland. Moreover, these officers had extensive experience in monitoring social media and in interacting with adolescents online. Three interviews were conducted, two of which were one-on-one interviews and one of which was a group interview. Before the interviews, the participants signed an informed consent form and were briefed about the purpose of the study. The interviews were audio-recorded (mean length 74 min) and transcribed to text.

3.2 Data-driven analysis

Our first aim was to identify the serious threats that adolescents encounter on social media. To reach this goal, the data was coded using Atlas.ti 8.4.22 software built for qualitative data analysis. The data was coded using inductive content analysis (Elo and Kyngäs, 2008; Krippendorff, 2018). Three primary code groups focusing on the threats adolescents face online were formulated of the quotations (total = 163): Polarization ($n = 68$ (56%)), Disinformation ($n = 28$ (23%)), and Social media as a pathway to illegal activities ($n = 25$ (21%)). In addition, as the officers discussed also other aspects of adolescents' online lives, two other code groups emerged: convergence of social media and real life ($n = 25$) and the Intergenerational gap in understanding social media ($n = 17$). However, as these two code groups did not primarily focus on the threats adolescents face, they were discarded from the current analysis.

The interviewees talked more about some themes than others. Hence, the size of the identified themes varied. For this reason, we employed the percentage of the coded content as a rough measure to quantify each theme's importance. It was decided to report the prevalence of each theme as a percentage of the coded content as the interviews also touched upon aspects not related to the current study (e.g., prevention strategies and the capabilities and resources of the police to monitor and respond to illegal content online). Measuring how much each of the identified themes (i.e., Polarization, Disinformation, Social media as a pathway to illegal activities) comprised of the whole interview would have underestimated the theme's importance. Instead, the approach we took measured how prevalent each theme was in the discussion concerning the threats that social media poses to adolescent users. This quantifies how much each theme was discussed in relation to other threats posed by social media leaving out the other topics discussed in the interviews which were beyond this article's scope.

3.3 Classificatory analysis

In the second phase, only the three identified main themes were analyzed. The Honeycomb model (Baccarella et al., 2018) was employed to re-analyse the coded content. Each identified theme

(Polarization, Disinformation, and Pathway to illegal activities) was re-coded with Atlas.ti using the functionalities of the Honeycomb model as pre-identified code groups. This was done to break down the identified code groups into more specific "threat profiles" which present the identified threats as functionalities of social media. Due to the scope of the article, we focus on the three most prominent functionalities of each identified threat when discussing the results.

4 Results

4.1 Data-driven analysis

The interviews were primarily dominated by the police officers discussing content relating to polarization online. This theme encompasses different behaviors and phenomena, including an "us vs. them" mentality, extremist groups, social divides and events and spontaneous swarming organized on social media. The police officers discussed, for example, the effect of polarization on adolescents' attitudes, radicalization and extremist propaganda. They also mentioned the strong ingroup/outgroup bias present on social media between different groups and sub-cultures. In addition, they raised concerns about things such as group fights and large-scale gatherings organized on social media and the ease with which a few influential users can coordinate such events.

Disinformation encompasses all material relating to the distribution of disinformation, such as fake news or other false information, and the effect this has on adolescents on social media. The identified threats in this theme are related to the prevalence of disinformation on social media platforms and the capabilities of adolescents to spot it. In addition, the officers emphasized the effect of confirmation bias and the long-term impact this could have on the worldviews of adolescents. They also raised concerns of adolescents as producers of disinformation and focused on their role in producing false profiles and information on social media platforms.

The third-most discussed theme was coined Social media as a pathway to illegal activities. This theme encompasses the parts of the interviews that concerned the aspects of social media that might function as pathways to various illegal activities. For example, the officers raised concerns about the ease of aggressively engaging with others on social media and committing offenses such as defamation or criminal threats. In addition, they pointed out the ease with which different groups could motivate adolescents to participate in illegal activities such as group fights or assaults. Moreover, the officers discussed the ease with which even younger people can be involved in either selling or using drugs via social media.

4.2 Classificatory analysis of the identified themes

After identifying threats that arose from the interviews (Polarization, Disinformation, Social media as a pathway to illegal activities), we employed the Honeycomb model of social media functionality as a method of classification. This approach aimed to provide a more nuanced picture of the identified threats by

examining the functional building blocks by which these threats are present on social media platforms.

4.2.1 Polarization

Topics concerning Polarization dominated the interviews. As seen in Table 1, Polarization is primarily generated by the functions of Groups ($n = 50$), Conversations ($n = 33$) and Identity ($n = 18$) with the most prominent being, unsurprisingly, Groups. Groups refers to the extent to which users of social media platforms create groups and societies online and how these groups generate behavior and group dynamics, such as ingroup/outgroup bias.

4.2.1.1 Groups

Based on their work observing and interacting with adolescents online, the police officers brought up the generally polarized nature of social media interactions. However, although they viewed political polarization online as a general threat to national security, they did not regard it as a threat related to most adolescents as adolescents' conversations typically revolve around everyday topics more than politics. Nevertheless, the officers discussed the general culture of polarization and aggressive engagement as they thought it served as an example for social interactions between adolescents and generated an unhealthy environment for adolescents on social media. For instance, they reported witnessing that even girls as young as 12 could threaten to kill their peers online. For example, one officer noted that: "... a child had written, and even girls write these things, on Snapchat that I will beat you into such a shape that even your parents won't recognize you". Another officer commented: "It happens a lot, the previous example was under 15 years old for example", and continued: "and they sometimes contact us, for example with a voice message, saying that they have been threatened, but they are too afraid to tell their parents or teachers but they well that they have to tell someone" In addition, they said that polarization online could provide ample opportunities for extremist groups to recruit or influence adolescents on social media:

4.2.1.2 Conversations

The police officers described that fights and gatherings were often organized on social media. These fights were often caused by personal arguments and disagreements which then became public and spread on social media gathering supporting groups for both sides of the fights. The officers noted that once the sides

had formed online, it was remarkably easy for a few individuals to organize gatherings and fights in real life: "According to the police officers, the invitations to these fights were sometimes reported to include requests for the fighters to take weapons, such as knives or clubs, with them Such invitations were reported to have circulated on social platforms such as Snapchat and WhatsApp. Moreover, the officers discussed how the 'facelessness' of social media conversations can encourage threats and extremely aggressive social interaction. These factors, in turn, were thought to increase polarization and group formation online escalating aggressive engagement and the generation of group fights". Overall, the spreading of information about fights and gatherings on social media more broadly was considered harmful as it could draw in curious individuals who were not originally involved in the argument. For example one officer noted about influential adolescents organizing gatherings on social media: "...if (s)he has enough followers. They can say on Snapchat, for example, and share the snap and for example it was on Jodel [anonymous social media app] as well that lets meet up at the [park name]. It was this one influencer we were following and when we had just emptied one park, he posts that this is the next place to be at and when we arrive it is full".

4.2.1.3 Identity

While the officers did not point out political identity as a significant source of polarization for adolescents, other identity displays were thought to be evident in group formation on social media. For instance, the officers recalled having witnessed that some of the real-life group fights that were organized online were branded as "immigrants vs. Finns": "for example, it was on the news as well, there was a confrontation between so-called native finns and immigrants, and the invitation to the fight was circulating online" and continued: "the example in its simplicity shows the power you can have if only two people make decide to machinate an group gathering and in a couple of days they can organize it so that tens or in this worst case we are talking about 300 adolescents who have been asked to carry weapons...". They raised worries about promoting such a discourse online as they thought it could provide opportunities for extremist groups to exploit such spontaneously generated polarization. One officer reported that: "I would say that even though it was not the case originally the situations [for the group fight], it did start to attract ideological people even though they were not originally involved. When the knowledge

TABLE 1 The frequencies and relative frequencies of each category (polarization, disinformation, and pathway to illegal activities) and functional building blocks of the Honeycomb-model (groups, conversations, identity, sharing, reputation, relationships).

	Polarization	Disinformation	Pathways to illegal activities	Total
Whole category	125 (63%)	49 (25%)	25 (12%)	199 (100%)
Groups	50 (88%)	5 (9%)	2 (3%)	57 (100%)
Conversations	33 (62%)	17 (32%)	3 (6%)	53 (100%)
Identity	18 (50%)	9 (25%)	9 (25%)	36 (100%)
Sharing	11 (33%)	14 (42%)	8 (24%)	33 (100%)
Reputation	8 (58%)	3 (21%)	3 (21%)	14 (100%)
Relationships	5 (83%)	1 (17%)	0 (0%)	6 (100%)

of such events with polarization spreads, it gives opportunities for extremists to join and promote their own activities" Moreover, the officers considered adolescence as a phase where young people are searching for their identities and were concerned that extremist groups could exploit this by providing harmful ideals and ways to develop identity. This influence could lead to adolescents basing their identities on extremist or illegal ideals. Thus, the inherent polarization also discussed in Groups could partly stem from the need of adolescents to construct and promote their identities on social media.

4.2.2 Disinformation

Disinformation as a category encompasses the extent to which false information and information influencing are purposefully present in social media. It is primarily generated by the functions of Conversations ($n = 17$), Sharing ($n = 14$), and Identity ($n = 9$). While Disinformation as a category did not dominate the interviews to the same extent as Polarization, the interviewed officers identified systematic disinformation and propaganda as one of the most significant concerns for national security overall. They saw the coordinated efforts to spread disinformation on social media, especially in disasters or emergencies, as a significant threat. The breakdown of the functionalities can be seen in Table 1.

4.2.2.1 Conversations

Generally, the officers speculated that adolescents' conversational culture renders them vulnerable to disinformation. They saw the conversations between adolescents on social media as "fractured" with little focus on the validity of the information. The officers considered that this fractured nature of conversations is beneficial for spreading disinformation: "… if you think about information influencing, it does not necessarily need good arguments, you just need to have a properly timed response in a conversation and a provocative take…" This was partly because the officers thought adolescents often take claims "at face value" as the conversational culture does not require factual justification. They reported that this might provide an opportunity for information influencing as a well-placed and timed piece of information could spread quickly and organically within conversational networks: "…there is a huge amount of information and there's false information of course, and that causes people to imagine situations that are perceived as threats. And then when we go to a local youth center and talk with young people, we notice that almost all of the rumors have started from social media" Moreover, the officers brought up that to them it seems that the source of information often dictates whether adolescents believe in the information they receive with information from a friend or a relative often judged to be trustworthy.

The interviewed officers also noted that knowledge about events spreads quickly among adolescents on social media. They described that in the absence of official information from authorities, someone will "fill the informational void" and then this information will often start spreading as the official version about the given event. This happens to such an extent that the police sometimes get contacted by newspapers and tabloids about whether the circulating information on social media is accurate. Furthermore, the police officers discussed the possibility that these rumors, often about violent events such as school shootings, can affect adolescents'

general sense of security: importantly, the officers emphasized that there is no clear way to determine whether such false information is produced on purpose or whether it is purely misinformation.

4.2.2.2 Sharing

The interviewees identified "fake news" and sharing of it as a major threat. They noted that during large-scale emergencies a coordinated flood of disinformation, appearing as tailored fake news articles and pictures, is often present and actively shared on social media. According to the interviewees, this suggests organized machinery that is aimed at producing and sharing disinformation to distort the general discourse: "…10 min after an event, or a bit more, and we can find disinformation or completely made-up news articles [about the event]…so there has to be an organized machinery behind it". The mentioned fake news were identified as false news articles or pictures posing as mainstream media. The officers were concerned about how adolescents might be especially vulnerable to this material due to possibly having lower critical thinking skills than adults.

4.2.2.3 Identity

In addition to being targets of disinformation, the officers identified adolescents as active producers of disinformation. They noted this as evident in fake profiles in which adolescents fake the identities of police officers on social media sites such as Instagram. Such fake profiles could be used to slander other adolescents or post false information regarding police records; "They had made an account with [the name of the officer] and used it to talk to people. Then when he [the officer] was visiting a school, a girl asked him do you really think that I'm this and that since you have denigrated me on Instagram. And then it became clear that it was not the police but a youth who had made fake profiles of officers and used them to slander teenagers". Moreover, fabricated "Most Wanted" list which included the names of local adolescents were found circulating online: "TIn addition, the officers thought that the vast quantity of differing information online allows adolescents to cherry-pick information based on their own preferences and that this is often used as a way to construct their own identities online.

4.2.3 Social media as a pathway to illegal activities

The interviewed officers emphasized the role of social media as a pathway to illegal activities. Social media as a pathway to illegal activities is primarily generated by the functions of Identity ($n = 9$), Sharing ($n = 8$), and Conversations ($n = 3$). The breakdown of the functionalities can be seen in Table 1. The officers were concerned that increasingly younger adolescents can routinely have access to illegal content on social media because of the prevalence of such content. The illegal content that the officers discussed included the selling and buying of drugs, threats made on social media, slander, and impersonation of police officers. Overall, the officers were concerned that the "real-life" background that used to dictate whether adolescents get involved with crime (i.e., upbringing, steady family life) would not matter anymore on social media.

4.2.3.1 Identity

According to the officers, one of the main aspects driving adolescents to illegal content and activities on social media is the idolisation of criminal culture and the attempts to use this culture

to build their online identity. The officers had observed adolescents both idolizing and emulating criminal behavior, for example, displaying weapons online, shoplifting, and engaging in acts of violence. For instance, an officer recalled how one adolescent had sworn to steal an expensive jacket from a convenience store if they got enough likes and how this was followed by other adolescents idolizing such acts. Moreover, while the officers reported that no proper gang culture exists at the moment, they were worried that some adolescents could aim to create such culture in the future:"…we are not talking about proper gangs. They are trying to but I wouldn't say that these groups fulfill the criteria for a gang, its mostly about building your own image. Of course we are investigating whether there's a leader for these kind of groups and we try to stay on top of things: however, officers based their concerns over their observations of adolescents having signs, such as number combinations, in social media usernames to signal that they belong to a particular group or "gang" as one officer put it.

4.2.3.2 Sharing

The officers reported that the glorification of criminal culture and illegal activities was a recurring theme in shared material. For instance, sharing pictures of guns, other weapons, or bulletproof vests was reportedly common within certain circles. However, the officers often did not view sharing of such material as a concrete threat as it was believed to be primarily used to boost online reputation and identity among adolescents. Such means of gaining reputation via sharing illegal material online is also highlighted by the already mentioned example of streaming the stealing of an expensive jacket if enough people like a certain post: "…one thing that was common knowledge and all the youth told as was this one guy who posted stuff online saying watch me stream while I steal this 1,200e jacket. Overall, the officers raised concerns about the availability of such material and its effects on driving adolescents to copy the behavior or to try something similar. However, in addition to this more "harmless" material, the officers reported that openly selling drugs on social media is a phenomenon among adolescents on social media that worries them and they had concerns about the increased availability of illegal substances on social media platforms.

4.2.3.3 Reputation

The officers noted that the idolisation of criminal culture was at least partly fuelled by a few well-known adolescent "influencers" on social media. These influencers were often known by the police and had past criminal records. In addition, they were well-known among local adolescents and could use their influence to organize the aforementioned group fights and gatherings, thereby comprising a potential threat. Moreover, copying these influential adolescents in the hope of gaining reputation on social media was also reported as a potential threat to both adolescents and civic order. One officer summarized: "for example these influencers we are following and we know… they have a criminal record and some have electronic angle monitors and might be 18–19 years old… but then there are 14-year olds who copy them with no criminal records and they had done the same thing of running and stealing [after seeing a video]".

5 Discussion

We aimed to map out the threats that Finnish adolescents face on social media from the perspective of law enforcement professionals and how these threats affect societal security. This was done based on concrete examples acquired via interviews with Finnish preventive measures police officers. Three primary threats were identified using data-driven content analysis: polarization, Disinformation, and Pathway to illegal activities. To generate more nuanced depictions of these complex threats, they were further analyzed by employing the Honeycomb model of social media functionality (Kietzmann et al., 2011; Baccarella et al., 2018) as a classificatory device. The threats were broken down into their most prevalent functions. By taking this approach, we were able to examine the functions that cause these threats instead of focusing on the threats themselves. The current approach allowed a more meticulous examination of the phenomena and provided a useful framework for better preventing or countering the threats posed by social media. Our approach expanded on previous research employing the Honeycomb model to study the dark side of social media, which has been largely quantitative (Sands et al., 2020; see Talwar et al., 2020). Similarly to Demetis (2020), who employed the Honeycomb model to qualitatively describe "ultra-dark" aspects of social media, the current research shows how the model can be employed via content analysis to break down large overarching themes into more manageable parts.

Overall, our results support previous findings about the societal threats related to social media use. The most significant threats identified in the police officers' interviews—Polarization and Disinformation—have been identified as significant societal threats also in previous research (Howard et al., 2018; Tucker et al., 2018; Barberá, 2020; Shu et al., 2020; McKay and Tenove, 2021). However, while worries about polarization as a threat to society have been widely discussed within the literature (Howard et al., 2018), this discussion has largely neglected adolescents. Our results suggest that the polarization present on social media also affects the lives of adolescents. The interviewees suggested that while political polarization as such does not play a pivotal role in the life of young people on social media, forms of polarization are still present. It is noteworthy that polarization is present as the formation of different groups based on strong group identities whose members use distinct symbols to represent themselves online. Thus, polarization presents threats on both individual and societal levels. According to the police officers, for example, the formation of such groups could be exploited by extremist groups or other outside actors via systematic information influencing. Such influencing could increase polarization and use existing tensions for further division or to even spark real-life violence. As the interviewees brought up having witnessed incidents of physical fights spreading from social media to real life, the online polarization of adolescents can be seen to already affect life outside social media. The officers were worried that this aspect could be fuelled by outside actors. The officers told that they could identify specific influential users who could effectively mobilize these groups. This suggests that possible information influencing coming from the right source could potentially effectively influence the behavior and actions of these groups. In addition to posing threats to societal security, these

factors can also have significant impacts on the individual wellbeing of young people joining such groups and fights.

Disinformation was proposed to cause threats due to its spread via the sharing of content and conversations between users. The interviewed officers speculated that the amount of false information present on social media and especially the speed with which it is produced is unlikely to be organically produced. This was reported to be especially true in the case of emergencies. Systematic information influencing has been employed widely on a global level (Bandeli and Agarwal, 2021; Ng and Taeihagh, 2021) and while it has also been recorded in Finland (Aro, 2019), research of its use within the country is scarce. The severity of the threat posed by disinformation is partly determined by the capabilities of citizens in detecting it. While Finnish adolescents are found to be confident in their abilities to spot false information (Riikonen et al., 2020; Kaarkoski et al., 2021), no extensive research has been conducted on the actual abilities of Finnish adolescents to detect false information online. The interviewed officers postulated that adolescents could be generally poor at detecting false information with this view reflecting the results of international studies on adolescents' social media literacy skills (McGrew et al., 2018; Johnston, 2020; Breakstone et al., 2021). If the officer's estimations of young people's capabilities are correct, adolescents' conversations on social media may pose a network for disinformation to quickly spread. As the officers proposed that disinformation is especially pronounced during emergencies, this could impede the functionality of emergency services and public authorities and thus pose a societal threat. Similar findings were reported by Papapicco et al. (2022), who performed focus group interviews with 41 Italian adolescents (aged 13–16). The study, focused particularly on racial hoaxes online, reported that Italian teenagers, as their Finnish peers, felt confident in recognizing racial hoaxes online. However, while the participants in this study first reported themselves to be "immune" to disinformation, after a second thought they could identify instances when they had believed fake news. This finding can be of particular interests in evaluating the confidence of Finnish adolescents in detecting misinformation; it might be that similarly to their Italian peers, the Finnish adolescents might report "immunity" to misinformation, but upon closer reflection might be able to identify cases when then have fallen for mis- or disinformation. The study by Papapicco et al. (2022) highly underlines the importance of conducting future studies looking at the actual abilities of adolescents in detecting false information online, instead of focusing on self-report.

Moreover, our analysis also revealed less discussed threats, mainly the worry that social media might act as an easy pathway for illegal activities among adolescents who would not normally engage in them. The prevalence of inappropriate content on social media and its effects on the wellbeing of adolescents has been extensively studied (Goodyear et al., 2018; Smahel et al., 2020) and it is known that social media plays a significant role in the radicalization of youth (Kardaş and Özdemir, 2018; Gaudette et al., 2021). This study extends prior research by focusing also on how social media promotes "lesser" crimes, such as shoplifting, which has received relatively little attention in prior research. If idolizing criminal culture on social media leads to adolescents actively pursuing criminal acts, this can have adverse effects for both the adolescents

themselves and the society in general. For instance, promoting criminal discourses on social media could provide new methods of influence for malevolent actors trying to reduce social trust. Such discourses could spread within the aforementioned conversational pathways between adolescents and escalate into actual criminal acts. These acts could harm the society they happen in and also the life of the adolescent performing them through legal consequences. Thus, more research into criminal discourses on social media and their effects on adolescents is required.

As the discussed factors can pose threats to national security, a key question is how to best counter them. Based on our analysis, disinformation could rapidly spread in conversational networks of adolescents (Conversations), especially during emergencies or otherwise distinct events. The rapid nature of disinformation spread in such situations renders traditional fact-checking services ineffective (Oeldorf-Hirsch et al., 2020; Zhang et al., 2021). Instead, more recent research has focused on psychological inoculation, that is, exposing adolescents to weakened disinformation in the form of a game (Van der Linden and Roozenbeek, 2021). Psychological inoculation has been shown to significantly increase recognition of false information in recent empirical studies (Basol et al., 2020; Roozenbeek et al., 2021; Van der Linden and Roozenbeek, 2021). Using a similar approach with Finnish adolescents could reduce the effect of disinformation spreading via sharing and within conversations. Importantly, psychological inoculation can help individuals recognize manipulative techniques often employed in disinformation. Thus, it does aim to guard or censor certain types of information, but instead aims to provide tools for better individual and critical evaluation of information. If the general ability of adolescents to detect such manipulative techniques and disinformation were to increase, the spread of disinformation within the conversational network would decrease as each "node" (i.e. adolescent) within the network would transmit less false information. Psychological inoculation could cause this increase in the detection of disinformation and thus reduce the overall flow of disinformation.

The approach of psychological inoculation assumes that the adolescents are motivated to detect false information. If identity formation also plays a role in the spread of disinformation, as suggested by our analysis, different additional approaches may be required. In identity formation, the truthfulness of information might not be the deciding factor, compared to how well it fits into the online persona the user wants to build. For example, if an adolescent that highly idolizes criminality comes across false information, such as the fake "Most Wanted" lists mentioned in the interviews, they are likely to pass the information on regardless of whether they perceive it to be true. Thus, simply training adolescents to better detect false information may not be enough.

Another approach could be to train adolescent to react and recognize to particular types of dis- and misinformation. For example, D'Errico et al. (2023) report a promising intervention aimed at reducing ethnic biases of moral disengagement. By promoting socio-analytical thinking via providing an alternate perspective of an immigrant, the intervention was shown to reduce racial biases when dealing with racially motivated misinformation. As mis- and disinformation online often relies on racial biases and prejudices (Papapicco et al., 2022; D'Errico et al., 2023), interventions such as this could provide vital methods for

countering the functional mechanisms (e.g., racial prejudice) that help false information spread online.

Of course, our research has limitations. First, by definition, our research excludes the voice of adolescents themselves, as the themes discussed represent the point of view of law enforcement officers. Consequently, our results about how adolescents' access and assess information online is based on the perspective of the police officers, which is unlikely to reflect how adolescents view the situation. For example, previous research indicates that although adolescents are confident in their ability to detect false information (Riikonen et al., 2020; Kaarkoski et al., 2021), adolescents' skills at detecting disinformation are generally poor (McGrew et al., 2018; Johnston, 2020; Breakstone et al., 2021; Nygren and Guath, 2021). This suggests a gap between adolescents' perceptions of their skills and their actual skills in dealing with the dark side of social media. Moreover, focusing on the police officers instead of the adolescents themselves provided a unique perspective for our study. Police officers possess a distinct skill set and professional expertise which makes them able to discuss the threats of social media from the point of national security in a way that adolescents could not. Future research is needed to examine adolescents' own perspectives of the threats of social media to find out whether they indeed perceive and experience the threats differently than police officers.

Secondly, interviewing police officers of course biased our results toward more extreme phenomena. It is important to note that the threats and phenomena discussed in our interviews are unlikely to reflect the experience of an average adolescent on social media or the most common threats they face. However, as the rationale of our research was to consider the most significant threats to societal security, we view this as a strength of our study. Indeed, one of the officers' concerns in the interviews was the reach of the extreme phenomena to all users of social media, causing adolescents who are generally distanced from violent, criminal, or otherwise illegal material to get exposed to it. The high exposure rates to inappropriate material presented in the EU online survey of children (Smahel et al., 2020) suggest that the officers' worries about the prevalence of such material are not unfounded. However, although our interviewees discussed this subject on a more general level, it is likely that the effect of such material on adolescents is related to various social factors such as social class, residential location, gender, education, and ethnicity. Thus, a task for future research is to examine whether the detrimental effects of social media are dependent on the everyday life and social context of the adolescents.

To conclude, the results presented here should be taken as a preliminary investigation of the threats faced by Finnish adolescents online, which can have an impact on national security. Noteworthily, the aim of our research was not to provide an exhaustive description of the threats present on social media, but to probe the opinions of professional police officers to provide a general overview of the threats from a law-enforcement perspective. Our explorative research can be expanded by more in-detail analysis of the identified threats and functionalities. Moreover, as our research has confirmed that from the point of view of law-enforcement adolescents face various types of threats on social media, research on combatting these threats is in the interest of both national security and the wellbeing of adolescents.

While our research focused on Finnish adolescents, it is highly likely that due to the global nature of social media, adolescents in other countries face similar threats.

Data availability statement

The datasets presented in this article are not readily available because the interviewees gave their consent to participate to the study on the condition that the data will be kept classified. Requests to access the datasets should be directed to teija.sederholm@mil.fi.

Ethics statement

Ethical approval was not required for the studies involving humans because the study followed the guidelines published by the Finnish National Board on Research Integrity: https://tenk.fi/sites/default/files/2023-05/RI_Guidelines_2023.pdf. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JÄ designed the concept of the study, wrote the manuscript, and conducted the data analysis. TS and RR conducted data collection. TS provided guidance on the methodology and supervised this research work. A-MH and RR provided assistance in commenting the text. All authors read and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Allcott, H., Gentzkow, M., and Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Res. Polit.* 6, 2. doi: 10.1177/2053168019848554

Anderson, C. A., Bushman, B. J., Donnerstein, E., Hummer, T. A., and Warburton, W. (2015). SPSSI research summary on media violence. *Anal. Soc. Issues Public Policy.* 15, 4–19. doi: 10.1111/asap.12093

Aro, J. (2019). *Putinin trollit: tositarinoita Venäjän infosodan rintamilta.* Helsinki: Werner Söderström Ltd.

Baccarella, C. V., Wagner, T. F., Kietzmann, J. H., and McCarthy, I. P. (2018). Social media? It's serious! Understanding the dark side of social media. *Eur. Manage. J.* 36, 431–438. doi: 10.1016/j.emj.2018.07.002

Baccarella, C. V., Wagner, T. F., Kietzmann, J. H., and McCarthy, I. P. (2020). Averting the rise of the dark side of social media: the role of sensitization and regulation. *Eur. Manag. J.* 38, 3–6. doi: 10.1016/j.emj.2019.12.011

Bandeli, K. K., and Agarwal, N. (2021). Analyzing the role of media orchestration in conducting disinformation campaigns on blogs. *Comput. Math. Organ. Theory.* 27, 134–160. doi: 10.1007/s10588-018-09288-9

Barber, M., and McCarty, N. (2015). "Causes and consequences of polarization", in *Political Negotiation: A Handbook*, eds. J. Mansbridge and C. J. Martin (Washington, D.C.: Brookings Institution Press), 37–90.

Barberá, P. (2020). "Social Media, Echo Chambers, and Political Polarization", in *Social Media and Democracy: The State of the Field, Prospects for Reform (SSRC Anxieties of Democracy)* eds. N. Persily and J. A. Tucker (Cambridge: Cambridge University Press), 34–55.

Barberá, P., Vaccari, C., and Valeriani, A. (2017). "Social media, personalisation of news reporting, and media systems' polarisation in Europe", in *Social Media and European Politics,* eds. M. Barisione and A. Michailidou (London: Palgrave Macmillan UK), 25–52.

Basol, M., Roozenbeek, J., and Van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *J. Cognition.* 3, 2. doi: 10.5334/joc.91

Bastick, Z. (2021). Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. *Comp. Human Behav.* 116, 106633. doi: 10.1016/j.chb.2020.106633

Bastos, M., and Farkas, J. (2019). "Donald trump is my president!": the internet research agency propaganda machine. *Social Media+ Soc.* 5, 3. doi: 10.1177/2056305119865466

Bossetta, M. (2018). The weaponization of social media: spear phishing and cyberattacks on democracy. *J. Int. Affairs.* 71, 97–106.

Boyd, R. W., and Swanson, W. S. (2016). The evolution of virtual violence: how mobile screens provide windows to real violence. *Pediatrics (Evanston).* 138, 2. doi: 10.1542/peds.2016-1358

Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., et al. (2021). Students' civic online reasoning: a national portrait. *Educ. Res.* 50, 505–515. doi: 10.3102/0013189X211017495

Bruns, A. (2019). *Are Filter Bubbles Real?* Cambridge, UK: Polity Press.

Bucchianeri, M. M., Eisenberg, M. E., Wall, M. M., Piran, N., and Neumark-Sztainer, D. (2014). Multiple types of harassment: associations with emotional well-being and unhealthy behaviors in adolescents. *J. Adolesc. Health.* 54, 724–729. doi: 10.1016/j.jadohealth.2013.10.205

Bushman, B. J., and Anderson, C. A. (2009). Comfortably numb: desensitizing effects of violent media on helping others. *Psychol. Sci.* 20, 273–277. doi: 10.1111/j.1467-9280.2009.02287.x

Casas, J. A., Ortega-Ruiz, R., and Monks, C. P. (2020). "Cyberbullying: A changing phenomenon", in *Online Peer Engagement in Adolescence*, eds. J. A. Casas, R. Ortega-Ruiz and C. P. Monks (London: Routledge), 71–84.

Castano-Pulgarín, S. A., Suárez-Betancur, N., Vega, L. M. T., and López, H. M. H. (2021). Internet, social media and online hate speech. Systematic review. *Aggress. Violent Behav.* 58, 101608. doi: 10.1016/j.avb.2021.101608

Chun, J., Lee, J., Kim, J., and Lee, S. (2020). An international systematic review of cyberbullying measurements. *Comput. Human Behav.* 113, 106485. doi: 10.1016/j.chb.2020.106485

Crowley, B. Z., and Cornell, D. (2020). Associations of bullying and sexual harassment with student well-being indicators. *Psychol. Violence.* 10, 615–625. doi: 10.1037/vio0000345

Demetis, D. S. (2020). Breaking bad online: a synthesis of the darker sides of social networking sites. *Eur. Manag. J.* 38, 33–44. doi: 10.1016/j.emj.2019.12.013

D'Errico, F., Cicirelli, P. G., Corbelli, G., and Paciello, M. (2023). Addressing racial misinformation at school: a psycho-social intervention aimed at reducing ethnic moral disengagement in adolescents. *Soc. Psychol. Educ.* 1–20. doi: 10.1007/s11218-023-09777-z

Dhir, A., Yossatorn, Y., Kaur, P., and Chen, S. (2018). Online social media fatigue and psychological wellbeing—a study of compulsive use, fear of missing out, fatigue, anxiety and depression. *Int. J. Inf. Manage.* 40, 141–152. doi: 10.1016/j.ijinfomgt.2018.01.012

Diamond, L., and Plattner, M. F. (2012). *Liberation technology: Social Media and the Struggle for Democracy.* Baltimore: JHU Press. doi: 10.1353/book.72079

Dobber, T., Metoui, N., Trilling, D., Helberger, N., and de Vreese, C. (2021). Do (Microtargeted) deepfakes have real effects on political attitudes? *Int. J. Press/Polit.* 26, 69–91. doi: 10.1177/1940161220944364

Elo, S., and Kyngäs, H. (2008). The qualitative content analysis process. *J. Adv. Nurs.* 62, 107–115. doi: 10.1111/j.1365-2648.2007.04569.x

Ferrara, E. (2020). "Bots, Elections, and Social Media: A Brief Overview" in *Disinformation, Misinformation, and Fake News in Social Media*, eds. K. Shu, S. Wang, D. Lee and H. Liu. (Cham: Springer Cham), 95–114. doi: 10.1007/978-3-030-42699-6_6

Fetzer, J. H. (2004). Information: does it have to be true? *Minds Mach.* 14, 223–229. doi: 10.1023/B:MIND.0000021682.61365.56

Floridi, L. (2013). *The Philosophy of Information.* Oxford: OUP Oxford.

Friis, S. M. (2015). "Beyond anything we have ever seen": beheading videos and the visibility of violence in the war against ISIS. *Int. Aff.* 91, 725–746. doi: 10.1111/1468-2346.12341

Garcet, S. (2021). Understanding the psychological aspects of the radicalisation process: a sociocognitive approach. *Forens. Sci. Res.* 6, 115–123. doi: 10.1080/20961790.2020.1869883

Garry, A., Walther, S., Rukaya, R., and Mohammed, A. (2021). QAnon conspiracy theory: examining its evolution and mechanisms of radicalization. *J. Deradicalizat.* 26, 152–216.

Gaudette, T., Scrivens, R., Davies, G., and Frank, R. (2021). Upvoting extremism: collective identity formation and the extreme right on Reddit. *New Media Society.* 23, 3491–3508. doi: 10.1177/1461444820958123

Geschke, D., Lorenz, J., and Holtz, P. (2019). The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *Br. J. Soc. Psychol.* 58, 129–149. doi: 10.1111/bjso.12286

Goodyear, V., Armour, K., and Wood, H. (2018). *The Impact of Social Media on Young People's Health and Wellbeing: Evidence, Guidelines and Actions.* Birmingham, UK: University of Birmingham. Available online at: http://epapers.bham.ac.uk/3070/1/The_Impact_of_Social_Media_on_Young_People%E2%80%99s_Health_and_Wellbeing_(FINAL_15TH_JAN).pdf (accessed October 11, 2022).

Guess, A., Nagler, J., and Tucker, J. (2019). Less than you think: prevalence and predictors of fake news dissemination on Facebook. *Sci. Adv.* 5, 1. doi: 10.1126/sciadv.aau4586

Hills, T. T. (2019). The dark side of information proliferation. *Persp. Psychol. Sci.* 14, 323–330. doi: 10.1177/1745691618803647

Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J., and François, C. (2018). *The IRA, Social Media and Political Polarization in the United States, 2012–2018.* Oxford: UK. Project on Computation Propaganda. Available online at: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1004&context=senatedocs (accessed October 11, 2022).

Hussain, M. N., Bandeli, K. K., Al-Khateeb, S., Tokdemir, S. S., and Agarwal, N. (2018). "Examining strategic integration of social media platforms in tracking disinformation campaign coordination," in *2018 International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction and Behavior Representation in Modeling and Simulation, BRiMS 2018.*

Jayasuriya, N., Azam, S. M., Khatibi, A., Atan, H., and Dharmaratne, I. R. (2018). The role of social media marketing on brand equity-a literature review. *Global J. Manage. Busi. Res.* 18, 31–39.

Jayasuriya, N. A., and Azam, S. F. (2017). The impact of social media marketing on brand equity: a study of fashion-wear retail in Sri Lanka. *Int. Rev. Manage. Market.* 7, 178–183. Available online at: https://econjournals.com/index.php/irmm/article/view/5663 (accessed October 11, 2022).

Jelenchick, L. A., Eickhoff, J. C., and Moreno, M. A. (2013). "Facebook depression?" social networking site use and depression in older adolescents. *J. Adolesc. Health.* 52, 128–130. doi: 10.1016/j.jadohealth.2012.05.008

Johnston, N. (2020). Living in the world of fake news: high school students' evaluation of information from social media sites. *J. Austral. Library Informat. Assoc.* 69, 430–450. doi: 10.1080/24750158.2020.1821146

Kaarkoski, M., Riikonen, R., Huhtinen, A.-M., Tillander, E., and Sederholm, T. (2021). Luotanko vai enkö luota? Nuorten luottamus sosiaalisessa mediassa leviävään informaatioon ja siihen liittyviä tekijöitä. *Tiede ja Ase* 79, 11–31.

Kardaş, T., and Özdemir, Ö. B. (2018). "The Making of European Foreign Fighters: Identity, Social Media and Virtual Radicalization", in *Non-State Armed Actors in*

the Middle East, eds. M. Yeşiltaş and T. Kardaş (Cham: Palgrave Macmillan Cham), 213–235. doi: 10.1007/978-3-319-55287-3_10

Keles, B., McCrae, N., and Grealish, A. (2020). A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. *Int. J. Adolesc. Youth*. 25, 79–93. doi: 10.1080/02673843.2019.1590851

Kelly, Y., Zilanawala, A., Booker, C., and Sacker, A. (2018). Social media use and adolescent mental health: findings from the UK Millennium Cohort Study. *EClinicalMed*. 6, 59–68. doi: 10.1016/j.eclinm.2018.12.005

Khan, M. F., and Jan, A. (2015). Social media and social media marketing: a literature review. *IOSR J. Busin. Manage*. 17, 12–15.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., and Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Busin. Horizons* 54, 241–251. doi: 10.1016/j.bushor.2011.01.005

Klimstra, T. A., Hale, I. I. I., Raaijmakers, Q. A. W., Branje, S. J. T., and Meeus, W. H. J. (2010). Identity formation in adolescence: change or stability? *J. Youth Adolesc*. 39, 150–162. doi: 10.1007/s10964-009-9401-4

Kowalski, R. M., and Limber, S. P. (2007). Electronic bullying among middle school students. *J. Adolesc. Health* 41, S22–S30. doi: 10.1016/j.jadohealth.2007.08.017

Kowalski, R. M., Limber, S. P., and McCord, A. (2019). A developmental approach to cyberbullying: prevalence and protective factors. *Aggress. Violent Behav*. 45, 20–32. doi: 10.1016/j.avb.2018.02.009

Kricorian, K., Civen, R., and Equils, O. (2021). COVID-19 vaccine hesitancy: misinformation and perceptions of vaccine safety. *Hum. Vaccin. Immunother*. 18, 1. doi: 10.1080/21645515.2021.1950504

Krippendorff, K. (2018). *Content Analysis: An Introduction to its Methodology*. New York, NY: Sage Publications.

Kwan, I., Dickson, K., Richardson, M., MacDowall, W., Burchett, H., Stansfield, C., et al. (2020). Cyberbullying and children and young people's mental health: a systematic map of systematic reviews. *Cyberpsychol. Behav. Soc. Netw*. 23, 72–82. doi: 10.1089/cyber.2019.0370

Lazer, D. M. J., Baum, M., Benkler, Y., Berinsky, A. J., Greenhill, K., Menczer, F., et al. (2018). The science of fake news. *Science (New York, N.Y.)*. 359, 1094–1096. doi: 10.1126/science.aao2998

Levy, R. E. (2021). Social media, news consumption, and polarization: evidence from a field experiment. *Am. Econ. Rev*. 111, 831–870. doi: 10.1257/aer.20191777

Lewis, L., Somers, J. M., Guy, R., Watchirs-Smith, L., and Skinner, S. R. (2018). "I see it everywhere": Uoung Australians unintended exposure to sexual content online. *Sex. Health* 15, 335–341. doi: 10.1071/SH17132

Loader, B. D., and Mercea, D. (2011). Networking democracy? Social media innovations and participatory politics. *Informat. Commun. Soc*. 14, 757–769. doi: 10.1080/1369118X.2011.592648

Maghsoudi, R., Shapka, J., and Wisniewski, P. (2020). Examining how online risk exposure and online social capital influence adolescent psychological stress. *Comput. Human Behav*. 113. doi: 10.1016/j.chb.2020.106488

Martel, C., Pennycook, G., and Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognit. Res.: Princ. Implicat*. 5, 47. doi: 10.1186/s41235-020-00252-3

McCarty, N. M. (2019). *Polarization: What Everyone Needs to Know®*. New York: Oxford University Press.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., and Wineburg, S. (2018). Can Students Evaluate Online Sources? Learning from assessments of civic online reasoning. *Theory Res. Soc. Educ*. 46, 165–193. doi: 10.1080/00933104.2017.1416320

McKay, S., and Tenove, C. (2021). Disinformation as a threat to deliberative democracy. *Polit. Res. Q*. 74, 703–717. doi: 10.1177/1065912920938143

Moore-Berg, S. L., Hameiri, B., and Bruneau, E. (2020). The prime psychological suspects of toxic political polarization. *Curr. Opin. Behav. Sci*. 34, 199–204. doi: 10.1016/j.cobeha.2020.05.001

Näsi, M., Räsänen, P., Hawdon, J., Holkeri, E., and Oksanen, A. (2015). Exposure to online hate material and social trust among Finnish youth. *Inform. Technol. People*. 28, 607–622. doi: 10.1108/ITP-09-2014-0198

Näsi, M., Räsänen, P., Oksanen, A., Hawdon, J., Keipi, T., and Holkeri, E. (2014). Association between online harassment and exposure to harmful online content: a cross-national comparison between the United States and Finland. *Comput. Human Behav*. 41, 137–145. doi: 10.1016/j.chb.2014.09.019

Ng, L. H., and Taeihagh, A. (2021). How does fake news spread? Understanding pathways of disinformation spread through APIs. *Policy Intern*. 13, 560–585. doi: 10.1002/poi3.268

Norri–Sederholm, T., Norvanto, E., Talvitie–Lamberg, K., and Huhtinen, A. M. (2020). "Misinformation and Disinformation in Social Media as the Pulse of Finnish National Security", in *Social Media and the Armed Forces*, eds. E. M. de Baseggio, O. Schneider and T. S. Tresch (Cham: Springer International Publishing), 207–225.

Nygren, T., and Guath, M. (2021). Students evaluating and corroborating digital news. *Scand. J. Educ. Res*. 66, 1–17. doi: 10.1080/00313831.2021.1897876

Oeldorf-Hirsch, A., Schmierbach, M., Appelman, A., and Boyle, M. P. (2020). The ineffectiveness of fact-checking labels on news memes and articles. *Mass Commun. Soc*. 23, 682–704. doi: 10.1080/15205436.2020.1733613

Official Statistics of Finland (OSF). (2020). *Use of Information and Communications Technology by Individuals*. Helsinki: Statistics Finland. Available online at: http://www.stat.fi/til/sutivi/2019/sutivi_2019_2019-11-07_tau_021_fi.html (accessed October 11, 2022).

Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M., and Räsänen, P. (2014). "Exposure to Online Hate among Young Social Media Users," in *Soul of Society: A Focus on the Lives of Children & Youth*, ed. M. N. Warehime (Bingley: Emerald Group Publishing Limited), 253–273. doi: 10.1108/S1537-466120140000018021

Papapicco, C., Lamanna, I., and D'Errico, F. (2022). Adolescents' vulnerability to fake news and to racial hoaxes: a qualitative analysis on italian sample. *Multimodal Technol. Interact*. 6, 20. doi: 10.3390/mti6030020

Patton, D. U., Hong, J. S., Ranney, M., Patel, S., Kelley, C., Eschmann, R., et al. (2014). Social media as a vector for youth violence: a review of the literature. *Comput. Human Behav*. 35, 548–553. doi: 10.1016/j.chb.2014.02.043

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., and Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. *Psychol. Sci*. 31, 770–780. doi: 10.1177/0956797620939054

Persily, N. (2017). The 2016 US election: can democracy survive the internet? *J. Democ*. 28, 63–76. doi: 10.1353/jod.2017.0025

Rapp, D. N., and Salovich, N. A. (2018). Can't we just disregard fake news? The consequences of exposure to inaccurate information. *Policy Insights Behav*. 5, 232–239. doi: 10.1177/2372732218785193

Rathje, S., Van Bavel, J. J., and Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proc. Natl. Acad. Sci*. 118:e2024292118.

Riikonen, R., Huhtinen, A. M., and Norri-Sederholm, T. (2020). "Not a problem for me: Young men's conceptions of their social media use and false information", in *7th European Conference on Social Media (ECSM 2020)*, eds. Karpasitis, C., et al. (New York: Curran Associates, Inc.), 240–245.

Roozenbeek, J., Maertens, R., McClanahan, W., and van der Linden, S. (2021). Disentangling item and testing effects in inoculation research on online misinformation: Solomon revisited. *Educ. Psychol. Meas*. 81, 340–362. doi: 10.1177/0013164420940378

Sands, S., Campbell, C., Ferraro, C., and Mavrommatis, A. (2020). Seeing light in the dark: Investigating the dark side of social media and user response strategies. *Eur. Manag. J*. 38, 45–53. doi: 10.1016/j.emj.2019.10.001

Santarossa, S., and Woodruff, S. J. (2017). #SocialMedia: exploring the relationship of social networking sites on body image, self-esteem, and eating disorders. *Social Media+ Soc*. 3, 2. doi: 10.1177/2056305117704407

Schenk, A. M., and Fremouw, W. J. (2012). Prevalence, psychological impact, and coping of cyberbully victims among college students. *J. Sch. Violence*. 11, 21–37. doi: 10.1080/15388220.2011.630310

Sheldon, P., Rauschnabel, P., and Honeycutt, J. M. (2019). *The Dark Side of Social Media: Psychological, Managerial, and Social Perspectives*. San Diego, CA: Academic Press.

Shu, K., Bhattacharjee, A., Alatawi, F., Nazer, T. H., Ding, K., Karami, M., et al. (2020). Combating disinformation in a social media age. *Wiley Interd. Rev: Data Mining and Knowl. Discov*. 10, 6. doi: 10.1002/widm.1385

Shulman, E. P., Smith, A. R., Silva, K., Icenogle, G., Duell, N., Chein, J., et al. (2016). The dual systems model: review, reappraisal, and reaffirmation. *Dev. Cogn. Neurosci*. 17, 103–117. doi: 10.1016/j.dcn.2015.12.010

Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., et al. (2020). *EU Kids Online 2020: Survey Results from 19 Countries*. Hamburg: EU Kids Online.

Sorrentino, A., Baldry, A. C., Farrington, D. P., and Blaya, C. (2019). Epidemiology of cyberbullying across Europe: Differences between countries and genders. *Educ. Sci.: Theory Pract*. 19, 74–91.

Spears, B., Slee, P., Owens, L., and Johnson, B. (2009). Behind the scenes and screens: Insights into the human dimension of covert and cyberbullying. *Zeitschrift für Psychologie* 217, 189–196. doi: 10.1027/0044-3409.217.4.189

Steinberg, L. (2010). A dual systems model of adolescent risk-taking. *Dev. Psychobiol*. 52, 216–224. doi: 10.1002/dev.20445

Suarez-Lledo, V., and Alvarez-Galvez, J. (2021). Prevalence of health misinformation on social media: systematic review. *J. Med. Internet Res*. 23, e17187. doi: 10.2196/17187

Subrahmanyam, K., and Smahel, D. (2010). *Digital Youth: The Role of Media in Development*. New York: Springer Science+Business Media.

Talwar, S., Dhir, A., Singh, D., Virk, G. S., and Salo, J. (2020). Sharing of fake news on social media: application of the honeycomb framework and the third-person effect hypothesis. *J. Retail. Consum. Serv*. 57, 102197. doi: 10.1016/j.jretconser.2020.102197

Tucker, J. A., Guess, A., Barber,á, P., Vaccari, C., Siegel, A., Sanovich, S., et al. (2018). Social media, political polarization, and political disinformation: a review of the scientific literature. *SSRN Electron. J.* doi: 10.2139/ssrn.3144139. [Epub ahead of print].

Tucker, J. A., Theocharis, Y., Roberts, M. E., and Barberá, P. (2017). From liberation to turmoil: social media and democracy. *J. Democ.* 28, 46–59. doi: 10.1353/jod.2017.0064

Van der Linden, S., and Roozenbeek, J. (2021). "Psychological inoculation against fake news", in *The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation,* eds. R. Greifeneder, M. E. Jaffé, E. J. Newman, and N. Schwarz (London: Routledge/Taylor & Francis Group), 147–169.

Vargas, R. (2011). Being in "bad" company: power dependence and status in adolescent susceptibility to peer influence. *Soc. Psychol. Q.* 74, 310–332. doi: 10.1177/0190272511414546

Vidgen, B., Margetts, H., and Harris, A. (2019). *How Much Online Abuse is There?* London: Alan Turing Institute. Available online at: https://www.turing.ac.uk/sites/default/files/2019-11/online_abuse_prevalence_full_24.11.2019_-_formatted_0.pdf (accessed October 11, 2022).

Wang, Y., McKee, M., Torbica, A., and Stuckler, D. (2019). Systematic literature review on the spread of health-related misinformation on social media. *Soc. Sci. Med.* 240, 112552. doi: 10.1016/j.socscimed.2019.112552

Willoughby, M. (2019). A review of the risks associated with children and young people's social media use and the implications for social work practice. *J. Soc. Work Pract.* 33, 127–140. doi: 10.1080/02650533.2018.1460587

Yoon, S., Kleinman, M., Mertz, J., and Brannick, M. (2019). Is social network site usage related to depression? A meta-analysis of Facebook–depression relations. *J. Affect. Disord.* 248, 65–72. doi: 10.1016/j.jad.2019.01.026

Zhang, J., Featherstone, J. D., Calabrese, C., and Wojcieszak, M. (2021). Effects of fact-checking social media vaccine misinformation on attitudes toward vaccines. *Prev. Med.* 145, 106408. doi: 10.1016/j.ypmed.2020.106408

Zollo, F. (2019). Dealing with digital misinformation: a polarised context of narratives and tribes. *EFSA J.* 17, e170720. doi: 10.2903/j.efsa.2019.e170720

Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., de Vreese, C. H., and Helberger, N. (2016). Should we worry about filter bubbles? Internet policy review. *J. Intern. Regulat.* 5, 1. doi: 10.14763/2016.1.401