MARKUS KAAKINEN

Disconnected Online

A social psychological examination of online hate

MARKUS KAAKINEN

Disconnected Online

A social psychological examination of online hate

■

UNIVERSITY OF TAMPERE

MARKUS KAAKINEN

Disconnected Online

A social psychological examination of online hate

UNIVERSITY
OF TAMPERE

The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of the University of Tampere.

Cover design by
Mikko Reinikka

ABSTRACT

Social media provides users with advanced opportunities for social interaction and extended social networks. Parallel to this, however, are emerging virtual forms of victimization and offending. One example of this is online hate, which refers to online communication that threatens or degrades an individual or a social group. This dissertation examines how online hate offending and victimization are related to social relations among adolescents and young adults, both online and offline. Moreover, it analyzes how online hate is associated with social capital embedded in online and offline social networks and online group behavior. This dissertation consists of five separate studies examining online hate offending, victimization, and exposure. Studies were conducted among Finnish young people but also in cross-national context involving Finland, Germany, the United Kingdom, and the United States. According to the results of these studies, online hate is related to online and offline social relations but in different ways. Both online hate offending and victimization were positively associated with social capital in online environment. In addition, online group behavior was associated with an increased likelihood of online hate offending. Offline social capital, in turn, associated with a lower risk of being an agent or victim of online hate offending. And furthermore, strong connection to offline social networks buffered the harmful consequences of victimization due to online offending. Online hate also reflects social tensions deriving from the wider societal condition. Thus, the results of this dissertation imply that social media has potential to both connect and disconnect individuals. Those online users with strongest connections to their online social networks are also most likely to be involved in online conflicts. Social media can also accentuate inter-group conflicts and distinctions in the society.

ABSTRAKTI

Sosiaalinen media tarjoaa käyttäjilleen edistyneitä mahdollisuuksia vuorovaikutukseen ja laajojen sosiaalisten verkostojen luomiseen. Näiden mahdollisuuksien myötä on kuitenkin syntynyt myös uusia aggression ja sosiaalisten konfliktien muotoja. Yksi esimerkki tästä on verkkoviha, joka viittaa internetissä tuotettuun tai jaettuun yksilöitä tai sosiaalisia ryhmiä loukkaavaan tai uhkaavaan materiaaliin. Vihasisältö verkossa on herättänyt huolta laajalti ja se on ollut viime vuosina tasaisesti yhteiskunnallisen keskustelun kohteena. Tässä väitöskirjassa tutkitaan, miten verkkovihan tuottaminen tai sen kohteeksi joutuminen ovat yhteydessä nuorten ja nuorten aikuisten sosiaalisiin suhteisiin verkossa ja sen ulkopuolella. Tarkemmin työssä tarkastellaan verkkovihan yhteyksiä sosiaaliseen pääomaan ja ryhmäkäyttäytymiseen. Väitöskirja koostuu viidestä osatutkimuksesta, joissa käsitellään sekä verkkovihan tuottamista että sen kohtaamista ja kohteeksi joutumista Suomessa, Saksassa, Iso-Britanniassa ja Yhdysvalloissa. Tutkimustulosten mukaan verkkoviha on yhteydessä sosiaalisiin suhteisiin, mutta nämä yhteydet ovat erilaisia verkossa ja sen ulkopuolella. Sosiaalinen pääoma verkossa ennusti sekä todennäköisempää verkkovihan tuottamista että sen uhriksi joutumista. Lisäksi verkkovihan tuottaminen oli yhteydessä ryhmäkäyttäytymiseen verkossa. Sen sijaan ne nuoret ja nuoret aikuiset, joilla oli paljon sosiaalista pääomaa verkon ulkopuolella, olivat muita harvemmin verkkovihan tuottajia tai sen uhreja. Vahvat sosiaaliset suhteet verkon ulkopuolella voivat myös suojata nuoria ja nuoria aikuisia verkossa tapahtuvien uhrikokemusten kielteisiltä vaikutuksilta. Lisäksi tutkimustulokset osoittavat, että verkossa leviävä vihasisältö heijastaa yhteiskunnan sosiaalisia jännitteitä. Tulosten perusteella voidaan päätellä, että verkkovuorovaikutus sekä yhdistää että erottaa ihmisiä. Ne käyttäjät, joilla on vahvimmat yhteydet sosiaalisen median sosiaalisiin verkostoihin, ovat myös muita todennäköisemmin osallisina virtuaalisissa konflikteissa. Tämän lisäksi sosiaalinen media voi kärjistää ennestään yhteiskunnallisia kahtiajakoja.

ACKNOWLEDGEMENTS

Doctoral studies and this doctoral dissertation as its main element have been a substantial but also inspiring and educative effort. And, without overlooking the enriching doctoral program itself, I have received most of the inspiration and lessons from the incredible people I have had the opportunity to work with during my PhD studies period. Even though I consider this dissertation as a beginning instead of an endpoint, it does offer me a possibility to unveil some of my indebtedness for the support I have gained during these years.

First of all, I would like to announce my inexpressible gratitude to my supervisors Professor Atte Oksanen and Dr., Docent Irmeli Järventie. Atte has been an incredible supervisor, research team leader, mentor, co-author, and a friend for all these years. Without all these elements, neither this dissertation or the research and teaching experience I have gained during my PhD period would have been possible. All the conversations concerning different aspects of academic work but also life in general (not forgetting good wine) have been invaluable. As the supervisor of my master's thesis, Irmeli was the one who encouraged me to proceed to doctoral studies and towards work within the academia. Thus, it is my pleasure to express her my deep gratitude for initiating my academic pursues and for the privilege of benefiting from her support and insightful comments for many years now. I would also like to thank the pre-examiners, professors Matthew Williams from the Cardiff University and Agneta Fischer from the University of Amsterdam, whose contribution truly helped me to reflect on my work and revise it to its final form.

My doctoral thesis has been a part of the research project Hate Communities: A Cross-National Comparison (2013-2016) and its sister project Web of Rage: Extreme online communities in the light of research and analytic journalism (2014–2015), both funded by the Kone Foundation. It has been a privilege to be a part of these projects and their incredible research group. Both projects were directed by Atte Oksanen and professor Pekka Räsänen from the

LIST OF ORIGINAL PUBLICATIONS

I. Kaakinen, M., Räsänen, P., Näsi, M., Minkkinen, J., Keipi, T., & Oksanen, A. (2018). Social capital and online hate production: A four country survey. *Crime, Law and Social Change, 69*(1), 25–39. doi:10.1007/s10611-017-9764-5.

II. Kaakinen, M., Keipi, T., Oksanen, A., & Räsänen, P. (in press). How does social capital associate with being a victim of online hate? Survey evidence from the US, UK, Germany and Finland. *Policy & Internet*. doi: 10.1002/poi3.173.

III. Kaakinen, M., Keipi, T., Räsänen, P., & Oksanen, A. (2018). Cybercrime victimization and subjective well-being: An examination of the buffering effect hypothesis among adolescents and young adults. *Cyberpsychology, Behavior, and Social Networking, 21*(2), 129–137. doi:10.1089/cyber.2016.0728.

IV. Kaakinen, M., Sirola, A., Savolainen, I., & Oksanen, A. (2018). Impulsivity, internalizing symptoms and online group behavior as determinants of online hate. Manuscript submitted for publication.

V. Kaakinen, M., Oksanen, A., & Räsänen, P. (2018). Did the risk of exposure to online hate increase after the November 2015 Paris attacks? A group relations approach. *Computers in Human Behavior, 78*, 90–97. doi:10.1016/j.chb.2017.09.02

CONTENTS

LIST OF FIGURES AND TABLES

1 INTRODUCTION

During this century, social media has significantly modified human communications and social networks, especially among young people. Social media users are becoming increasingly connected and aware of the social ties that originate from their offline lives (Hampton, 2016; Hampton, Goulet, Rainie, & Purcell, 2011), but online interaction also allows users to engage with virtual social networks (Abbas & Mesch, 2018; Cole, Nick, Zelkowitz, Roeder, & Spinelli, 2017; Wellman & Gulia, 1999; Williams, 2006). The large number of social media users, the multitude of interaction setups, and the existence of extended online social networks all enable new forms of abuse (Peterson & Densley, 2017). Online hate (or cyberhate) is a form of abuse that threatens or degrades individuals or social groups in the online space (Keipi, Näsi, Oksanen, & Räsänen, 2017; Oksanen, Hawdon, Holkeri, Näsi, & Räsänen, 2014). Online hate targets various social categories (e.g., religious groups or sexual minorities) as well as personal characteristics such as appearance (Costello, Hawdon, Ratliff, & Grantham, 2016; Keipi, Näsi, et al., 2017; Lunstrum, 2017). Concrete forms of online hate include terrorist organizations' hate propaganda (Benigni et al., 2017; Klausen, 2015), moral panic that targets social groups (Awan & Zempi, 2016; Lunstrum, 2017; Williams & Burnap, 2016), and racist campaigns that target individuals (Pew Research Center, 2017a; Pew Research Center, 2017b).

Since the 1990s, the Internet has served as an efficient medium for disseminating ideas and establishing social networks, and hate-based groups were among the first to realize this (Levin, 2002). In addition to these groups' websites, mainstream social media has grown to include clear examples of threatening and degrading communication (Costello et al., 2016; Hawdon, Oksanen, & Räsänen, 2017; Oksanen, Hawdon, Holkeri, et al., 2014). Several characteristics make social media a well-matched environment for hostile communication. First, social media users can share their thoughts with wide audiences relatively free of external control—anonymously in some cases (Barkun, 2017; Keipi, 2015). When online,

people can also network with those who share a hateful ideology; this facilitates the reinforcement and spread of hostile ideas (Douglas, 2007; Oksanen, Hawdon, & Räsänen, 2014). On the other hand, disagreements between social networks only tend to further polarize members of each group (Bakshy et al., 2015; Yardi & Boyd, 2010; Zollo et al., 2017). In addition, emotions connect online users, with negative emotions such as anger being especially likely to fuel online discussions (Song, Dai, & Wang, 2016).

Online hate is internationally recognized as a significant social problem (Council of Europe, 2015; Gagliardone et al., 2015). However, the legislation on online expression varies across nations. Most liberal countries such as the United States value freedom of speech even in case of hostile communication; other nations, however, are more willing to set limits on hostile expressions (Hawdon et al., 2017).

In recent years, researchers have expanded the body of research on online hate, using both survey results (Costello et al., 2016; Hawdon et al., 2017; Oksanen, Hawdon, Holkeri, et al., 2014) and data derived from social media settings (Burnap & Williams, 2015; Klaussen, 2015; Williams & Burnap, 2016). However, knowledge is still needed about how online and offline social dynamics relate to hostile online behaviors.

The aim of this dissertation is to contribute to the research by assessing how offending and victimization due to online hate are related to online and offline social networks and group behavior. This was done using a social psychological theoretical framework to combine two approaches: social identity (Tajfel & Turner, 1979; Turner & Oakes, 1986) and social capital (Coleman, 1988; Lin, 1999; Putnam, 1993). Together, these approaches provide a unique perspective on online hate, highlighting the ways in which both online and offline social networks contribute to online hate. In addition, this perspective acknowledges that both online-specific group behavior and the wider societal-group relationships are important in the online hate phenomenon.

This dissertation comprises five empirical studies: Study 1 analyzed how online hate offending is associated with the perceived quality of social relations (i.e., cognitive social capital) in both the offline and online environments. Study 2, in turn, analyzed the associations between the victimization to online hate and the quality of both online and offline social relations. Study 3 examined whether the victimization to offensive crime online is associated with lower subjective

well-being and whether either offline or online social belonging buffered this association. Study 4 analyzed the associations between online-hate-based offending and online group behavior (i.e., social homophily, social identification, and self-stereotyping in online interactions). Study 5 analyzed whether quantitative and qualitative changes in online hate exposure are associated with wider societal situations.

2 SOCIAL MEDIA AND ONLINE HATE

2.1 Social media and changing social networks

Internet use is increasing globally, especially for communicative purposes and particularly among adolescents and young adults (Organisation of Economic Co-operation and Development [OECD], 2016). In Finland, nearly 100% of people aged between 16 and 34 use the Internet (compared to 88% of the total population), and 95 to 96% use the Internet several times a day (Official Statistics of Finland, 2017). In addition, social media applications (e.g., Facebook, Instagram, and Twitter) have become key mediums for social interaction. The number of social-networking-site (SNS) users has been increasing (Official Statistics of Finland, 2014 & 2016; Pew Research Center, 2015a); SNS use is particularly common among teenagers and young adults (Official Statistics of Finland, 2016; OECD, 2016; Pew Research Center, 2015b).

Online communication has evolved relatively quickly. Domestic use of the World Wide Web increased during the 1990s, but the original online infrastructure (Web 1.0) was rather passive (Kaplan & Haenlein, 2010; O'Reilly, 2005). Those with the skills and opportunity to create and publish online content were able to reach ever-increasing audiences, but this infrastructure excluded the majority of users, who remained mere consumers of information (Keipi, Näsi, et al., 2017; Krämer, Neubaum, & Eimler, 2017). Toward the beginning of 2000s, Web 2.0 technology (O'Reilly, 2005) enabled participatory online use, thus making self-expression, information coproduction, and two-way communication more accessible; thus, the average online user became far more active (Keipi, Näsi, et al., 2017; Krämer et al., 2017; O'Reilly, 2005). The SNSs and other social media sites that leveraged this architecture emerged around the beginning of 2000s (Kaplan & Haenlein, 2010; Keipi, Näsi, et al., 2017). SNSs combine people, technology, and social practices into so-called networked publics, which have

become a major part of communication, especially for younger users (Boyd, 2011, 2014).

Van Dijck (2013a, 2013b) has used the concepts of connectedness and connectivity to describe the evolution of SNSs. The term *connectedness* refers to users' ability to establish and maintain social connections through social media platforms. The databaselike architecture allowed users to create personal profiles and to network and communicate with other users, mainly by exploiting social ties that originated from offline social networks (Boyd and Ellison, 2007; Van Dijck, 2013a). The term *connectivity* describes the change in SNS architecture from enabling connectedness between users to maximizing the number of connections and the amount of data flowing within SNSs; platform owners can then monetize these connections and generated data (Van Dijck, 2012, 2013b). Thus, users are not just connected to other users, as the platforms' algorithmic functions (e.g., Facebook's newsfeed or LinkedIn's network updates), for example, filter contacts and content or make suggestions to users, thus generating activity. Content that receives more views, comments, likes, or comments is more profitable for platform owners (Van Dijck, 2013a). Thus, even hostile or socially destructive online phenomena can be economically profitable for the platforms as long as they generate participation and increase data flow (Pew Research Center, 2017a).

Since their emergence, social media platforms have significantly shaped social networks. According to Hampton (2016) these platforms' ability to help users maintain social relations and engage in person-to-network (or one-to-many) communication has led to social networks and communities that are characterized by persistent contact and pervasive social awareness. Persistent contact refers to continuous social ties; such social relationships can more easily endure life events that used to disconnect people (e.g., moving to another city or changing workplaces). Due to the pervasive social awareness, in turn, users are constantly aware of the immediate activities, opinions, interests, and even locations of those in their social circle (Hampton, 2016). However, high awareness of and connectedness to social networks can also cause distress (Hampton, Rainie, Lu, Shin, & Purcell, 2015) and induce a spiral of silence, as people may choose to withhold thoughts that they think others in their social network would not approve of (Hampton et al., 2014).

Social media use has been mostly driven by offline lives and social connections (Boyd & Ellison, 2007; Hampton et al., 2011; van Dijck, 2013a); people with large

offline social networks tend to have large online social networks as well (Abbas & Mesch, 2018). However, from the beginning, online interaction has facilitated the development of new social ties independent of spatiality and of offline social networks (Abbas & Mesch, 2018; Williams, 2006), including through the formation of online communities that are based on shared interests (Wellman & Gulia, 1999). In this case, communities are conceived as social networks instead of as spatial entities (Wellman, 1979); more specifically, the online community is defined "as the network of personal relationships to which a given individual belongs and that he or she manages" (McEwen & Wellman, 2013, pp. 168–169). Thus, online communities include membership in a network of interconnected online users with mutual interests.

Online communities can form through various platforms, including Facebook (Bliuc, Best, Iqbal, & Upton, 2017; Chan & Fu, 2017), Twitter (Benigni, Joseph, & Carley, 2017; Komorowski, Huu, & Deligiannis, 2018), YouTube (Oksanen et al., 2015; Rotman & Preece, 2010), and discussion forums (Graham, Jackson, & Wright, 2016; Sowles, Krauss, Gebremedhn, & Cavazos-Rehg, 2017). Communities can also evolve around interconnected platforms that utilize various context-specific affordances for interaction and information sharing (Matamoros-Fernández, 2017). The motivation behind online community formation varies, and it includes (but is not limited to) content sharing (Kaplan & Haenlein, 2010; Mikal, Rice, Kent, & Uchino, 2015), professional cooperation (Komorowski et al., 2018; McLoughlin, Patel, O'Callaghan, & Reeves, 2018), peer support (Bliuc et al., 2017; Sowles et al., 2017), political participation (Chan & Fu, 2017; Wang & Shi, 2018), and radicalization or racism (Benigni et al., 2017; Matamoros-Fernández, 2017).

Online communities can serve an important function in individuals' social relatedness. Participation in online communities can foster engagement, social support, and a sense of belonging (Sun, Fang, & Lim, 2014; Walther & Jang, 2012), even for large-scale, anonymous, and restricted communication (Mikal et al., 2015). These online social ties are particularly important for individuals whose social relations in traditional social environments are weak (Cole et al., 2017; Leist, 2013; Mesch, 2012). Thus, online communities can fulfill the basic human need for belonging (Baumeister & Leary, 1995). In addition, online communities can offer alternative or complementary forms of social interaction. As SNS users are persistently and pervasively connected to their personal networks (Hampton,

2016), they may choose to refrain from expressing certain thoughts or aspects of their identities to avoid disagreements or social sanctions (Hampton et al., 2014). In this case, people with marginalized ideologies or identities are especially likely to engage with online communities so as to find others who will validate opinions and identities that are rejected elsewhere (Chang & Bazarova, 2016; Chun & Lee, 2017; Dengah, Snodgrass, Else, & Polzer, 2018; Haas, Irr, Jennings, & Wagner, 2011).

The self-selection of online social affiliations may lead to homophilic social relations. The tendency to network with people who are similar to oneself is of course not limited to online communication (McPherson, Smith-Lovin, & Cook, 2001). However, as homophily in social relations tends to increase in parallel with possibilities for social selectivity (Bahns, Pickett, & Crandal, 2011), the self-selective nature of social media interaction is particularly suitable for social homophily (Kang & Chung, 2017; Liang & Fu, 2017; Oksanen, Hawdon, & Räsänen, 2014). The combination of homophilic social networks and the need for a shared reality can lead to echo chambers in which users are only exposed to information from like-minded users (Bakshy, Messing, & Adamic, 2015; Stern & Ondish, 2018). Even though social network composition is the key determinant of online information exposure (Bakshy et al., 2015), other factors reduce information diversity as well, including personal preferences and the SNSs' algorithmic filtering technology (Helberger, Karppinen, & D'Acunto, 2016; Keipi, Näsi, et al., 2017; Pariser, 2011).

In addition to opinion congruence, shared emotional valence also encourages people to engage with online discussions (Himelboim, Smith, & Shneiderman, 2013; Himelboim et al., 2016; Song, Dai, & Wang, 2016). In other words, people are more likely to participate in discussions and social networks that share the direction of their stance (positive or negative). Perhaps surprisingly, expressing and sharing emotions with those who have similar emotional valence appears to be the main driver of online political discussion, and discussions that express anger are the most likely to encourage participation (Song et al., 2016).

Echo chambers reinforce cohesion within online social networks but also lead to polarization and conflicts between networks (Boutyline & Willer, 2017; Densley & Peterson, 2017; Stern & Ondish, 2018; Yardi & Boyd, 2010; Zollo et al., 2017). In some cases, online communities are explicitly formed as antagonistic responses to other groups (Lo, Surian, Prasetyo, Zhang, & Ee-Peng, 2013;

Zielinski, Nielek, Wierzbicki, & Jatowt, 2018). This tendency toward polarization and stratification in online social networks is referred to as "cyberbalkanization" (Van Alstyne & Brynjolffson, 2005), a concept that appears to also predict offline polarization on the societal level, especially among young people (Chan & Fu, 2017). However, polarization is not an automatic feature of online communication; it is related to users' personal characteristics and to specific forms of online use and social engagement (Williams, 2007).

2.2 Online aggression and online hate

Social media is an important source of interaction and offers possibilities for participation and social belonging. However, communication in social media is sometimes characterized by hostile behavior such as cyberbullying, harassment, flaming, trolling, and spreading of hateful content or misinformation (Hutchens, Cicchirillo, & Hmielowski, 2015; Keipi, Näsi, et al., 2017; Näsi, Räsänen, Kaakinen, Keipi, & Oksanen, 2017; Pew Research Center, 2017a, 2017b; Williams & Burnap, 2016).

Online hate is content that threatens or degrades individuals or social groups (Hawdon et al., 2017; Keipi, Näsi, et al., 2017; Lunstrum, 2017; Oksanen, Hawdon, Holkeri, et al., 2014; Perry & Olsson, 2009; Waldron, 2012). This content can be directed toward sexual minorities, political factions, or ethnic and religious groups, among others, but it can also target personal characteristics such as appearance. Thus, online hate is a heterogenic collection of hostile expressional phenomena; it can be motivated by emotions such as hatred or anger, but this is not an essential part of the definition (Brown, 2017).

Online hate is a distinct form of online abuse, but it shares some similarities with other types of online aggression such as cyberbullying, harassment, and flaming. Cyberbullying and harassment are forms of peer abuse that directly target a certain individual (including behaviors such as stalking and the spreading of misinformation), but online hate is defined as threatening or degrading actions that can target either an individual or an entire social category (Jones, Mitchell, & Finkelhor, 2013; Keipi, Kaakinen, Oksanen, & Räsänen, 2017). Cyberbullying and harassment can involve elements of online hate, as in some examples of racist or

political harassment (Pew Research Center, 2017a), but this is not essential. The difference between online hate and flaming, in turn, is that flaming refers to aggressive verbal outbursts due to emotional disinhibition (Voggeser, Singh, & Göritz, 2018); online hate, however, also includes more deliberate and ideologically motivated hostilities (Hawdon et al., 2017; Keipi, Näsi, et al., 2017; Waldron, 2012).

Online hate can have direct and indirect consequences. Researchers on other forms of online abuse have reported that online victimization can be hurtful and is associated with poorer mental health and increased distress (Fahy et al., 2016; Ybarra, Mitchell, Wolak, & Finkelhor, 2006). Being threatened or degraded online likely has parallel effects. Researchers have also expressed concerns about the indirect consequences of hateful online content, including its potential to induce offline violence against certain groups (Awan & Zempi, 2016; Douglas, 2007) or to endanger social inclusiveness for some groups (Waldron, 2012).

Online hate has become a part of national and international policy debates (Council of Europe, 2015; Gagliardone et al., 2015; U.S. National Intelligence Council, 2017). At the same time, however, authorities still have difficulty policing hostilities and offending actions in cyberspace (Wall & Williams, 2013; Williams et al., 2013). Germany's new online-hate-speech law tackles the problem by obligating SNS operators to remove reported hate material from their pages (British Broadcasting Corporation [BBC], 2018). It is worth noting that Germany already stands out for its strict legislation on hate speech (Allen & Norris, 2011; Hawdon et al., 2017). At the other end of the continuum are countries such as the United States that favor freedom of speech in their national legislation (Waldron, 2012). The drive to reduce hateful online content, which is widely shared but only partly legislation-driven has led SNS companies to develop protocols for identifying and managing such content (The Guardian, 2018; Pew Research Center, 2017b), as well as user-driven activism aimed at countering online hate (Farkas & Neumayer, 2017).

The spread of hostile content online is not a new phenomenon. Various hate groups have been active online since the start of the domestic Internet (Levin, 2002). White supremacy groups in United States were the first to take advantage of the Internet's expressional freedom to disseminate their ideologies and recruit members (Douglas, McGarty, Bliuc, & Lala, 2005; Gerstenfeld, Grant, & Chiang, 2003; Levin, 2002; M. A. Wong, Frank, & Allsup, 2015). A variety of hate groups,

including European far-right movements (Caiani & Parenti, 2013; Lucassen & Lubbers, 2012) and Islamist terrorist organizations (Benigni et al., 2017; Klausen, 2015), now use the Internet and social media. Online hate groups are part of wider hate-propagation ecosystem that includes online communities, offline activist groups, and political agents (George, 2017).

Hate communities' websites are not the only apparent source of hostile online content in the mainstream online experience (Costello et al., 2016; Foxman & Wolf, 2013; Hawdon et al., 2017; Oksanen, Hawdon, Holkeri, et al., 2014). Users encounter online hate on Facebook (Farkas & Neumayer, 2017; Lunstrum, 2017), Twitter (Burnap & Williams, 2015; Klaussen, 2015; Williams & Burnap, 2016), YouTube (Sureka, Kumaraguru, Goyal, & Chhabra, 2010), online blogs and forums (Cammaerts, 2009; Flores-Yeffal, Vidales, & Plemons, 2011; Sela, Kuflik, & Mesch, 2012), and news websites' comments and discussion sections (Erjavec & Kovačič, 2012; Rains, Kenski, Coe, & Harwood, 2017). However, it is worth noting that, although online hate content is highly visible, only a small proportion of all material shared on social media is hateful (Jakubowicz et al., 2017; Williams & Burnap, 2016).

Social media is a particularly suitable environment for online hate. For example, the physical disconnect between the perpetrator and victims (Vakhitova, Reynald, & Townsley, 2016) and the ability that users have (at least on some platforms) to threaten or degrade others while remaining anonymous (Black, Mezzina, & Thompson, 2016; Densley & Peterson, 2017; Keipi et al., 2014) can lower the threshold for hostile behavior. In social media, one can easily find like-minded social networks to welcome and verify even one's hostile thoughts (Barkun, 2017; Oksanen, Hawdon, Holkeri, et al., 2014). These networks' group processes are likely to further amplify the extreme attitudes of their members (Douglas, 2007; McGarty et al., 2011). Hate-based networks with rigid and clearly structured worldviews can be psychologically and socially rewarding for participants, as they offer social ties to similarly minded people and a shared sense of purpose and meaning (Jasko, LaFree, & Kruglanski, 2016; Simi, Blee, DeMichele, & Windisch, 2017; Stern & Ondish, 2018). Given that SNSs provide efficient means for information sharing and one-to-many communication, such hateful content can spread widely through social media (Barkun, 2017; Flores-Yeffal et al., 2011), especially in times of social tension or after triggering events (Awan & Zempi, 2016; Sela et al., 2012; Williams & Burnap, 2016).

Not all threats of degraded online communication reflect a hostile ideology per se, but toxic disinhibition may motivate some confrontational online interaction (Voggeser et al., 2018). In other words, threatening or degrading communication can result from a clash of views between users or between groups. Although SNSs allow users to engage in high selectivity regarding their social ties and content, these users are still exposed to contradictory information and interactions (Bakshy et al., 2015). In line with this, online confrontations usually develop around discussions of public issues (Cionea, Piercy, & Carpenter, 2017; Erjavec & Kovačič, 2012) and are triggered by interactions that cross the borders of online networks (Densley & Peterson, 2017; Hutchens et al., 2015; Zollo et al., 2017).

3 SOCIAL PSYCHOLOGICAL RISK FACTORS OF ONLINE HATE

Social psychology underlines that both personal and environmental factors are important in shaping human behavior (Crocker & Canevello, 2012; Lewin, 1936). This aim is manifested in prominent social psychological models regarding aggressive behavior, such as the general aggression model (Anderson & Bushman, 2002), the script-based information-processing model (Huesman, 1988), and the social information-processing model (Dodge & Crick, 1990). These models account for personal characteristics and the social environment, thus deriving a holistic understanding of the psychosocial process behind aggressive behavior.

Online communication and social media have formed a new vector for hostile behavior. According to a review by Peterson and Densley (2017), the emerging forms of aggressive behavior in online environments share many of the risk factors traditionally associated with aggression, but they also have their own particularities. Personal attributes (e.g., impulsivity and internalizing symptoms) and group processes are associated with offline aggression offline, but they also have context-specific roles in online aggression (Peterson & Densley, 2017).

Impulsivity is a multidimensional concept that refers to insufficient self-control and a personal propensity to engage in maladaptive behavior (Bettencourt, Talley, Benjamin, & Valentine, 2006; De Wit, 2009). Furthermore, impulsivity is related to sensation seeking, urgency, low perseverance, and low premeditation (Whiteside & Lynam, 2001). High impulsivity is an established risk factor for violence and criminal behavior (Krakowski & Czobor, 2013; Krueger, Markon, Patrick, Benning, & Kramer, 2007), and it has also been associated with cyberbullying (Vazsonyi, Machackova, Sevcikova, Smahel, & Cerna, 2012; R. Y. M. Wong, Cheung, & Xiao, 2018) and offensive online behavior (White, Cutello, Gummerum, & Hanoch, 2017). Social media users are able to significantly customize their communication environments according to their own preferences, but these users still encounter material that clashes with their

personal views (Bakshy et al., 2015). Thus, an insufficient capacity for behavioral disinhibition can lead to hostility in online confrontations. In addition, the lack of social presence and behavioral accountability in online communication could further lower impulsive persons' tendency to reflect on their behavior (Van Royen et al., 2017).

Internalizing symptoms are characteristic of depression, anxiety, and other negative-affect-laden disorders (Achenbach, 1966; Krueger & Markon, 2006). These symptoms can reduce individuals' capacity for emotional and behavioral regulation (Selby, Anestis, & Joiner, 2008), which may then be manifested in aggressive behavior (Krakowski & Czobor, 2013). Internalizing symptoms have also been linked to cyberbullying (Bonanno & Hymel 2013; Chen et al., 2017). In online interactions, such behavior can be perpetrated without making physical contact with the victims (Vakhitova et al., 2016); some online platforms even provide a veil of anonymity (Black, 2016). This can make hostile online behavior relatively safe for the offenders, thus making it more likely that those with internalizing symptoms experience such behavior (Peterson & Densley, 2017).

Group processes can induce online hate (Peterson & Densley, 2017). Online social interaction is characterized by homophilic social networks that evolve around shared interests and ideologies (Bakshy et al., 2015; Kang & Chung, 2017). Unlike with internal cohesion, online groups are often polarized in opposite directions, and border-crossing contacts between networks tend to be negative (Rains et al., 2017; Yardi & Boyd, 2010; Zollo et al., 2017). In this study, online hate's risk factors related to group processes are examined via a theoretical framework that consists of the social identity approach (SIA; Tajfel & Turner, 1979; Turner & Oakes, 1986) and social capital theory (Coleman, 1988; Lin, 1999; Putnam, 1993).

3.1 Social identity approach to online hate

Intergroup conflicts and prejudiced attitudes are major themes in the social psychological tradition (Allport, 1954; Brown 2010; Tajfel, 1970). These objectives are particularly salient within the SIA, which is arguably the most influential social psychological theory for explaining group behavior (Haslam,

Ellemers, Reicher, Reynolds, & Schmitt, 2010; Hogg et al., 2004). The SIA is based on components of social identity theory (SIT; Tajfel & Turner, 1979 & 1986) and self-categorization theory (Turner, 1985; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). Together, these theories form an interactionist framework that emphasizes how group behavior is produced in a reciprocal interaction between individual psychological processes and societal context (Hogg, Terry, & White, 1995; Turner & Oakes, 1986).

According to SIT, social affiliations partly define a person's self-concept (Tajfel & Turner, 1979; Turner & Reynolds, 2010). Thus, self-conception is dependent on which social groups a person belongs to (i.e., the in-groups) and on comparisons to groups that person does not belong to (i.e., out-groups). As people strive to achieve positive self-esteem and distinctiveness, they tend to favor in-groups over out-groups (Tajfel, 1970; Tajfel & Turner, 1979). The key evidence for this "groupness" of social behavior was derived from experiments using trivial groups that lacked a shared history of interactions between groups or group members (Tajfel, 1970; Tajfel, Billig, Bundy, & Flament, 1971). In addition to esteem and distinctiveness, possible motives for social identification with in-groups include uncertainty reduction (Hogg et al., 2004), social connection, meaningfulness, competence, and self-continuity (Thomas et al., 2017; Vignoles, 2011).

The mechanism of classifying people into in-group and out-group categories is universal, but the concrete forms of this behavior are dependent on the wider sociocultural context of intergroup relations, as well as on situational and personal factors (Tajfel, 1970; Tajfel et al., 1971; Tajfel & Turner, 1986). Tajfel and Turner (1979) proposed three conditions in which categories or attributes become means of social identification and intergroup comparison. Membership in a social group must be internalized for it to become relevant to the self-concept. In addition, social situations cause certain characteristics to be salient identity markers, and certain out-groups to be relevant points of reference. Political views and ethnicities, for example, can appear as relevant markers of social identities in some contexts but not in others.

Self-categorization theory supplements SIT by further elaborating on social categorization's role in self-conception (Turner, 1985; Turner et al., 1987). Social categorization is a cognitive process in which people are perceived as representatives of social categories (e.g., ethnic groups or genders) instead of as

unique individuals (Allport, 1954; Tajfel et al., 1971). In line with this, self-categorization theory suggests that self-categorization replaces the personal self-concept with a depersonalized self-conception on the social-category level (Turner, 1985; Turner et al., 1987). According to Turner and Oakes (1986), "a self-categorization is a cognitive grouping of the self as identical (similar, equivalent, interchangeable) to some class of stimuli in contrast to some other class of stimuli" (p. 241). For this reason, people tend to see themselves in terms of prototypical group attributes—a tendency known as self-stereotyping (Hogg et al., 2004; Leach, van Zomeren, Zebel, Vliek, & Ouwerkerk, 2008; Turner & Oakes, 1986).

Self-categorization also induces intergroup polarization, as categorization moves toward both optimal similarity within groups and maximal difference between groups (Turner & Oakes, 1986). Thus, people find opinions that are actually more radical than the group average—but also more clearly distinct from the out-group's views—to be prototypical of their in-groups. Furthermore, the social identity of deindividuation (SIDE) model predicts that, as the self-concept moves from the personal level to the social level, social control replaces internal control, thus encouraging people to follow perceived or expected group norms in their behavior (Reicher, Spears, & Postmes, 1995; Reicher, Spears, Postmes, & Kende, 2016).

As discussed above, the SIA predicts that people in general favor in-groups over out-groups (Tajfel, 1970; Tajfel & Turner, 1979). However, social identity's relationship to out-group discrimination is complex, and in-group identification does not automatically lead to out-group antipathies. In the minimal-group condition (Tajfel et al., 1971), social identification appears to be mainly related to an in-group-favoring allocation of positive outcomes (rewards); it is related only to lesser degree (or not at all) to the biased allocation of negative outcomes such as punishments (Brown, 2010; Mummendey et al., 1992). Instead of applying in-group favoritism, people tend to share negative outcomes equally between groups or attempt to minimize the total amount of punishment. Given this tendency, the SIA needs further elaboration to explain when exactly social identification is based on mere in-group liking and when it also induces out-group discrimination.

Jackson and Smith (1999) have suggested that, by its nature, social identification can be either insecure or secure. Insecure social identification is characterized by identification with the in-group, a depersonalized self-

conception, a perceived common fate between in-group members, and a conflictual relationship with out-groups. In secure social identification, on the other hand, the individual identifies with the in-group, but other elements are missing. Of these two forms of social identification, only the insecure form is related to intergroup bias or negative attitudes toward out-groups; secure social identity is related to positive perceptions of out-groups (Jackson & Smith, 1999).

Social identification is rooted in a sociocultural intergroup context (Tajfel et al., 1971; Tajfel & Turner, 1979, 1986; Turner & Oakes, 1986). This means that conflicts between social groups change in saliency depending on the societal situation. Staub and Bar-Tal (2003) proposed that societal contexts that threaten basic human needs are likely to motivate hostilities against the groups that are perceived to be responsible for the unsatisfactory circumstances (Staub & Bar-Tal, 2003). Thus, intergroup conflicts escalate in times of fear, economic recession, political polarization, or perceived intergroup threats, for example (Baumeister, 1997; Staub, 1989; Staub & Bar-Tal, 2003).

The societal conditions that motivate intergroup conflicts and out-group bigotry are key factors in several social-identity-based theories. Proponents of terror-management theory (Greenberg, Solomon, Pyszczynski, & Lyon 1990; Greenberg, Pyszczynski, & Solomon, 1986) have suggested that identification with the in-group and the resulting shared worldview function as a death-anxiety buffer. Hostilities are targeted toward those who threaten this buffer and people are willing to accept violent military operations (Burke, Martens, & Faucher, 2010) and martyrdom-based attacks (Pyszczynski et al., 2006) to enhance security. According to uncertainty-identity theory (Hogg, 2007), personal and social uncertainty can induce stronger identification with clearly bounded in-groups (Hogg, 2014; Hogg et al., 2013), accentuate perceived group differences (Federico, Hunt, & Fisher, 2013), and lead to the dehumanization of out-groups (Esses, Medianu, & Lawson, 2013). Group-threat theorists (Stephan & Renfro, 2002; Stephan, Ybarra, & Rios Morrison, 2009), in turn, have emphasized that a perceived threat from out-groups (whether realistic or symbolic) can motivate negative attitudes (Riek, Mania, & Gaertner, 2006) and discrimination (Kauff, Asbrock, Issmer, Thörner, & Wagner, 2015) toward those groups.

Social media interaction can generate new forms of social identification and intergroup relations. Certain online groups are increasingly important reference groups for identification (Howard & Magee, 2013; Lehdonvirta & Räsänen, 2011;

Tikal, Rice, Kent, & Uchino, 2016). In addition, online social-identity dynamics can induce intergroup conflicts and lead to discrimination against or even dehumanization of out-groups (Rains et al., 2017; Synnott, Coulias, & Ioannou, 2017). In line with the SIA (Turner and Oakes, 1986), attitudinal polarization is characteristic of online intergroup behavior (Lo et al., 2013; Yardi & Boyd, 2010; Zielinski et al., 2018; Zollo et al., 2017), and societal triggering events shape intergroup online processes just as they do for offline processes (Awan & Zempi, 2016; Williams & Burnap, 2016).

According to SIDE, social identification's deindividuation effect is particularly common in online interactions, which often facilitate anonymous communication and reinforce group-based categorizations instead of personal identities (Lea & Spears, 1991; Reicher, Spears, & Postmes, 1995). Deindividuation is related to aggression in general (Densley & Peterson, 2017), but especially so in an online context (Christopherson, 2007; Fox & Tang, 2014; Peterson & Densley, 2017; Rains et al., 2017; Tang & Fox, 2016). However, as the SIDE model indicates, deindividuation is related to increased aggression only when aggressive behavior is a group norm (Christopherson, 2007; Lea & Spears, 1991; Reicher, Spears, & Postmes, 1995).

3.2 Social capital approach to online hate

The meaning of social relationships and communities for human behavior and well-being has been an enduring theme in the social sciences (Bourdieu, 1984; Durkheim, 1879/2002; Tönnies, 1887/1988; Wellman, 1979). As a part of this continuum, the theory of social capital refers to the value of social networks to individuals and social collectives (De Silva, McKenzie, Harpham, & Huttly, 2005). The concept has been defined in different ways (Bourdieu, 1984; Coleman, 1988; Putnam, 1993, 2000), but the common ground for all social capital approaches (SCAs) is that social capital is conceived as an asset investment in social networks (Lin, 1999).

Here, the starting point is Putnam's (1993, 2000) theory that social capital consists of social networks and norms or reciprocity and trust embedded within them. Moreover, social capital has both internal and external value (Putnam &

Goss, 2002). For example, for social network members, internal value comes in the form of information or social support. The external or public value, in turn, refers to a positive spillover mechanism in which social capital invested in certain social networks facilitates a wider scale of social organizations.

Although Putnam's conception of social capital emphasizes the meaning of solidarity groupings in local communities (e.g., villages or neighborhoods), social capital can also be seen as an investment in egocentric personal networks (e.g., ties to groups of friends and colleagues) (Lin, 1999; Wellman, 1979; Wellman, 2001; Wellman et al., 2001). This view reflects the wider shift from locale-based communities to communities embedded in personal networks (Wellman, 1979 & 2001; Wellman et al., 2001). From this network perspective, social capital is defined as "investment in social relations by individuals through which they gain access to embedded resources to enhance expected returns of instrumental or expressive actions" (Lin, 1999, p. 39). In other words, social capital is generated and utilized within nonspatial personal networks, such as professional networks or online communities, instead of more easily observable civic participation in public places (Wellman et al., 2001).

The investment in social networks (i.e., social capital) can be operationalized from structural and cognitive perspectives and as an individual- or ecological-level resource (De Silva et al., 2005; Harpham, Grant, & Thomas, 2002; Lin, 1999; Yip et al., 2007). The structural view on social capital stresses network compositions and behavioral patterns, such as participation, but the cognitive approach emphasizes subjective evaluations of the quality of social relations (e.g., trust or a sense of belonging) (De Silva et al., 2005; Harpham et al., 2002; Wellman et al., 2011). Structural and cognitive social capital comes close to Putnam's (2000) division between bridging and bonding social capital (Islam, Merlo, Kawachi, Lindström, & Gerdtham, 2006; Murayama, Fujiwara, & Kawachi, 2012). In this division, bridging social capital refers to connections between individuals with diverse backgrounds, and bonding capital consists of strong social ties to dense social networks, such as family or close friends (Putnam, 2000). As an individual-level resource, social capital may refer to an individual's participation behavior or trust toward certain social networks. On the other hand, the ecological-level operationalization emphasizes the aggregated structural or cognitive investment within a certain collective (e.g., a neighborhood) (De Silva et al., 2005; Lin, 1999).

Social capital has been related to various positive outcomes on both collective and individual level and it has become a salient part of national and international public health policy making and planning (De Silva et al., 2005; Muntaner et al., 2000). In general, social capital facilitates the functioning of democracy and civil society (Putnam, 1993), and societies and communities with high social capital tend to show fewer social problems, such as violent crime (Kennedy et al., 1998; Rosenfeld et al., 2001). On the individual level, social capital associates with improved well-being (Elgar et al. 2011; Han 2013; Kawachi, Kennedy, & Glass, 1999), educational attainment (Dika and Singh, 2002), and lower substance misuse (Awgu, Magura, & Coryn, 2016), for example.

In addition to the direct positive effects, social networks may foster well-being indirectly by buffering the stress caused by negative life events (Cohen & Wills, 1985; Thoits, 2011). This *buffering hypothesis* (Cohen & Wills, 1985) means that when people face negative experiences, such as criminal victimization, their relationships to others offer them resources that facilitate reliance and recovery from negative events (Brooks, Lowe, Graham-Kevan, & Robinson, 2016; Schultz et al., 2013). These resources can be emotional, informational, or instrumental support or the sense of social belonging (Cohen & Wills, 1985). Even though the buffering hypothesis does not originate from the tradition of social capital, the premises of these concepts are compatible and the buffering effect of social capital has since been well documented (An & Jang, 2018; Frank, Davis, & Elgar, 2014; Lindström & Giordano, 2016).

There are, however, mixed findings concerning the buffering function of social networks originating from offline and online environments. Although offline and online social ties may serve as buffers against stressing events (Cole et al., 2017), there is some evidence suggesting that there are differences between these two forms of networks when it comes to protecting individuals against negative experiences (Minkkinen et al., 2016; Turja et al., 2017). Furthermore, in these studies, social ties to offline primary groups but not to online social networks were found to buffer young people against risky online experiences. This is in line with previous studies stressing that intimate offline connections are of high importance for well-being, and online social ties fail to reach this significance (Lee, Chung, & Park, 2018).

In addition to widely endorsed positive outcomes, social capital may have some less desirable consequences. Even though the public good perspective has

been widely endorsed, the agonistic tradition (Bourdieu 1984, 1986) has claimed that the elements of distinction, exclusion, and rivalry are integral aspects of social capital as well (Julien, 2014). Thus, social capital may strengthen the internal cohesion of certain social networks but also accentuate borders and conflicts between networks. In Putnam and Goss (2002) this would mean that social capital has the internal value for network members but not the external value that strengthens the overall social cohesion outside the networks. Indeed, social capital can foster cohesive yet uncivil social networks and exclusive solidarity, as in the examples of criminal coalitions (Ostrom, 1997; Putnam, 1993), civil wars (Pérez-Díaz, 2002), and even genocides (McDoom, 2014).

In the first years of computer-mediated communication, it was feared that online interaction would reduce the level of traditional interpersonal interaction and lead to a decline of individuals' and societies' social capital (Nie, 2001). However, this concern has since been questioned, and online communication has facilitated the creation of social capital (Bouchillon 2014; Hampton & Wellman 2003; Kim & Kim, 2017; Schrock, 2016; Wellman et al., 2001). One often-suggested explanation for this positive relationship is that SNS offers an efficient and accessible communication tool with one's personal networks, which facilitates the formation and utilization of social capital (Boase 2008; Boase and Wellman, 2004; Ellison, Gray, Lampe, & Fiore, 2014; Ellison, Lampe, Steinfield, & Vitak, 2011; Ellison, Steinfield, & Lampe, 2007).

Most studies on the subject have concentrated on the effects social media has on offline social capital (Abbas & Mesch, 2018; Williams, 2006). However, social capital is also generated within online communities based on virtual social ties (Oh, 2016; Park & Park, 2016; Perry et al., 2018; Williams, 2006). In these online communities, social capital is associated with higher user activity, as those members with the most social capital tend to be more likely to disseminate information and participate in community interactions (Chang & Chuang, 2011; Yen, 2016). In some cases, social capital generated online eventually spills over into the offline environment (Shen & Cage, 2015; Rosen et al., 2011).

According to Julien (2015), online social capital is predominantly seen as a public good fostering information sharing and other virtual civic behaviors. However, the distinction, exclusion, and conflict aspects of social capital is characteristic of online social networks, as online communities can use significant effort and creativity to separate themselves from other communities (Julien,

2015). This claim is supported by several other studies reporting polarization and conflicts between various online networks (Rains et al., 2017; Yardi & Boyd, 2010; Zollo et al., 2017). These notions are in line with Lin's (1999) early theorizing on the emerging phenomenon of online communities, which predicted that online social networks will promote social capital formation as well as "tensions, conflicts, violence, competition, and coordination issues among villages" (p. 47). As users with the most social capital embedded in online communities are the most active online communicators (Chang & Chuang, 2011; Yen, 2016), they are also more likely to expose themselves to risks and conflicts in online social networks (Green, 2007).

3.3 Combining the social capital and social identity approaches

In this study, the SIA and SCAs are combined to examine how group processes within online social networks form risk factors for online hate offending and victimization. There are some notable differences and similarities between these approaches. First, the SCA stresses investments in social networks that benefit individuals and social collectives (De Silva et al., 2005; Lin, 1999). The SIA, in turn, stresses the importance of social groups for self-construal (Tajfel & Turner, 1986). SCAs have focused on social relations within actual social networks, and SIT emphasizes subjective identification in actual social networks (e.g., work or online communities) and abstract social categories (e.g., ethnic groups; Hogg, Abrams, Otten, & Hinkle, 2004).

Social capital and social identification can be considered as theoretical explanations for the cohesiveness of social groups and the human being's ability to contribute to social collectives instead of just taking advantage of them (Baumeister, Ainsworth, & Vohs, 2016; Brewer & Kramer, 1986; Coleman, 1990; Putnam, 1993). Social capital theory suggests that capital embedded in certain social networks will benefit network members, but it may have a civic effect extending outside these networks as well (Putnam & Goss, 2002). The SIA does not propose that identification to a social group would enhance solidarity toward other groups as well. On the contrary, identification may induce in-group

favoritism and out-group discrimination (Tajfel & Turner, 1979). In addition to social integration, both approaches have been used to study conflictual group relations. Intergroup conflicts have been in the central focus of SIA from the beginning of the tradition (Tajfel, 1970; Tajfel & Turner, 1979). SCAs, on the other hand, have been mainly concerned about the benefits of social ties (Julien, 2015), but more agonistic accounts also exist (Bourdieu, 1984, 1986; Julien 2015; McDoom 2014; Pérez-Díaz 2002).

Social identification and social capital can reinforce each other. Social identification to a group motivates cooperation and collective action which, in turn, leads to accumulations of social capital (Brewer & Kramer, 1986; Kelly & Kelly, 1994; Kramer, 2006). This tendency has been identified in offline groups, such as organizations (Kramer, 2006), and within online communities (Teng, 2017; Yen, 2016). In addition, social identity may contribute to cognitive social capital by enhancing the sense of belonging to a certain social network (Baumeister & Leary, 1995; Steffens, Haslam, Schuh, Jetten, & van Dick, 2017). Furthermore, social identification is shaped by the interaction within groups (Jans, Leach, Garcia, & Postmes, 2014; Thomas et al., 2017). This indicates that social capital (e.g., the number and perceived quality of social ties) within social networks can induce social identification.

Together, the SIA and SCAs account for the group processes emerging from the social self-construal as well as the quantitative and qualitative aspects of social networks. The common factor between these approaches is the role of belonging. The knowledge of group membership is the starting point for social identification and self-categorization (Tajfel & Turner, 1979; Turner & Oakes, 1986). Furthermore, individuals generate social capital via investment in the social networks they belong to (Lin, 1999; Wellman et al., 2001). Thus, belonging to social networks, such as a group of friends or an online community, can induce social identification and self-categorization as well as cognitive (e.g., trust or sense of belonging) and structural (e.g., number of social ties) investment into social networks. The combined framework of the SIA and SCAs offers a useful analytical tool, and it has been used in earlier studies regarding group behavior in offline (Vahtera, Buckley, Aliyeva, Clegga, & Cross, 2017; Kramer, 2006) and online contexts (Kaye, Kowert, & Quinn, 2017; Teng, 2017; Yen, 2016). The combined theoretical framework of SIA and SCAs is shown in Figure 1.
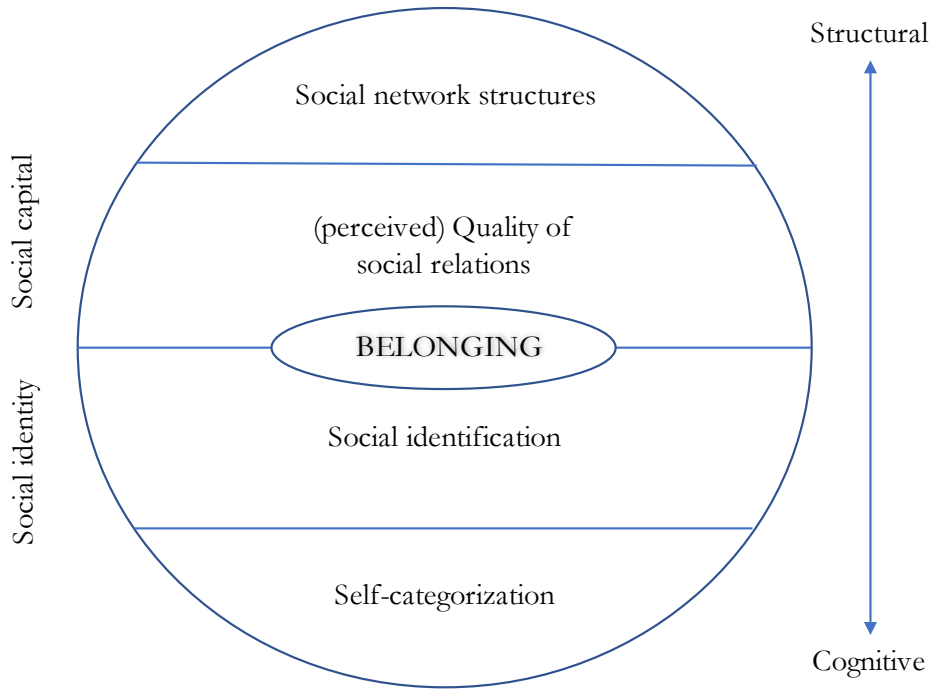
Figure 1. The theoretical framework of combined social identity approach and social capital approach.

4 STUDY OBJECTIVES AND HYPOTHESES

The objective of this dissertation is to study how online hate offending and victimization is affected by online and offline social relations. This is done by using a social psychological framework that combines social identity approach and social capital theory but also more personal risk factors. I concentrate on adolescents and young adults who are members of the age group that is most engaged with online social networking (Organization for Economic Cooperation and Development, 2016). The number of studies scrutinizing online hate has increased in recent years (Burnap & Williams, 2015; Costello et al., 2016; Hawdon et al., 2017; Keipi, Näsi, et al., 2017). However, there are still research gaps that this dissertation aims to contribute to. First, even though various studies have focused on the perpetrators of cyberbullying or harassment (Peterson & Densley, 2017) there is no research on determinants of offending due to online hate. Second, earlier research has analyzed the determinants of online hate victimization (Räsänen et al., 2016), but no studies have examined how offline and online social relations shape the likelihood and consequences of online hate victimization. Third, there is research on how triggering societal conditions contribute to online hate (Awan & Zempi, 2016; Burnap & Williams, 2015), but these studies have rather short time periods and concentrate on the occurrence of hate material in certain online platforms.

To contribute to the abovementioned research gaps, five studies were conducted. The first two studies analyzed how online hate offending (Study 1) and victimization (Study 2) associate with cognitive social capital in offline and online contexts. The third study examined whether the victimization of offending crime online is associated with lower subjective well-being and whether the offline and online social belonging buffered this association. The fourth study analyzed how online group behavior (social homophily, social identity dynamics), and personal risk factors (impulsivity and internalizing symptoms) are related online hate offending. The fifth study analyzed whether triggering societal

conditions (i.e., economic recession and terrorist attacks) manifested in the temporal change of online hate between 2013 and 2015.

4.1 Research hypotheses

4.1.1 How does cognitive social capital offline and online associate with offending due to online hate?

In the first study, we analyzed how cognitive social capital in personal networks offline (group of friends) and online (online community) associate with offending due to online hate among adolescents and young adults using data from Finland, Germany, the United Kingdom, and the United States. The cognitive social capital was conceptualized as trust and a sense of belonging to a personal network. Two competing hypotheses concerning the relationship between offending due to online hate and cognitive social capital were stated on the basis of social capital literature (see the Social capital approach to online hate section):

H1: Social capital embedded in offline and online networks will have public value in the sense that it will facilitate the social organization of online interaction. Therefore, trust and a sense of belongingness to a group of friends and an online community will have a negative association with conflict behavior (i.e., the production of online hate content).

H2: In online interaction, social capital embedded in offline and online networks will function as a resource in conflict. Given the distinction and exclusion between individuals and social groups, trust and a sense of belongingness to a group of friends and an online community will have a positive association with conflict-based online activity (i.e., online hate-content production).

4.1.2 How does social capital associate with being a victim of online hate?

In the second study, we examined the relationship between online hate victimization and cognitive social capital embedded in offline (group of friends) and online (online community) personal networks among adolescents and young adults. The cross-national comparison was based on four comparable data sets from Finland, Germany, the United Kingdom and the United States. We also examined the nature of association between online hate offending and online social capital (i.e., whether the association is linear or nonlinear). The cognitive social capital was measured as trust and a sense of belonging to a personal network. We stated two main hypotheses based on social capital literature (see the Social capital approach to online hate section):

H3: Cognitive social capital online will be linked to the likelihood of online hate victimization. As earlier research on the matter does not exist, the direction and the linearity of this hypothesized association are tested.

H4: Social capital in the offline context will be negatively associated with online hate victimization.

4.1.3 Is offensive cybercrime victimization associated with lower well-being, and does social belonging buffer the association?

In the third study, we analyzed the relationship between cybercrime victimization and subjective well-being among adolescents and young adults from Finland, Germany, the United Kingdom and the United States. We studied the association between crime victimization and subjective well-being and whether the association is moderated by a sense of belonging to offline primary groups (family and friends) or to an online community. Cybercrime victimization comprised general, offensive cyberfraud victimization. We stated the following hypotheses on the basis of cyber victimization and buffering effect literature (see the Online

aggression and online hate section and the Social capital approach to online hate section):

H5: Cybercrime victimization has a negative association with subjective well-being.

H6a: Social belongingness to offline primary groups buffers the negative association between cybercrime and subjective well-being.

H6b: Social belongingness to an online community buffers the negative association between cybercrime and subjective well-being.

4.1.4 Online group behavior and personal risk factors as determinants of online hate

In Study 4, we used an integrative approach including online group behavior and personal risk factors (impulsivity and internalizing symptoms) to analyze online hate offending among Finnish adolescent and young adults. Online group processes were conceptualized as social homophily, social identification, and self-stereotyping in a social media vignette experiment. On the basis of earlier studies concerning online interaction, online hate, and social identity (see the Social media and online hate and the Social identity approach to online hate section) we stated following hypotheses:

H7: Online hate offending will be positively associated with impulsivity.

H8: Online hate offending will be positively associated with internalizing symptoms.

H9: Online hate offending will be positively associated with social homophily in online networks.

H10: The association between online hate offending and social homophily online will be stronger among those with high impulsivity and internalizing symptoms.

H11: Online hate offending will be positively associated with online social identification.

H12: Online hate offending will have a direct positive association with self-categorization in online interaction and an indirect positive association with conformity to negative group norms.

H13: The saliency of the group identity will modify the association between online hate offending and self-stereotyping.

4.1.5 Did the risk of exposure to online hate increase after the November 2015 Paris attacks?

In Study 5, we analyzed the quantitative and qualitative changes in online hate exposure among Finnish adolescents and young adults between two time points ranging from spring 2013 to the end of 2015 (a few weeks after the Paris attacks). Our interest was whether the social conditions marked by fear, polarization, and uncertainty manifested in the temporal change of exposure to online hate. We stated three hypotheses on the basis of social identity literature (see the Social media and changing social networks section):

H14: As a consequence of the unsafety and uncertainty caused by events in 2015, people will witness more hostility expressed in online spaces; thus, our survey respondents will report more exposure to hateful material online in 2015 when compared to 2013.

H15: People will report more exposure to hate associated with categories relating to social uncertainty and perceived in-group threats (ethnicity or nationality, religion, political views, and terrorism).

H16: No similar effect will be found in terms of less relevant categories, such as gender, sexual orientation, disability, and appearance.

4.2 Contextualizing the research

In this dissertation, I study online hate among adolescents and young adults from Finland, Germany, the United Kingdom, and the United States (in Studies 1–3). In the following sections, I shall briefly discuss these factors that contextualize existing studies and the interpretations derived from them.

4.2.1 Adolescents and young adults and online behavior

The age group of adolescents and young adults is a distinctive population in terms of online behavior and aggression. Adolescents and young adults are the most active online users internationally (OECD, 2016; Official Statistics of Finland, 2016; Pew Research Center, 2015b), and the time spent online by young people is beyond comparison to any other daily activity (Rideout, Foehr, & Roberts, 2010; Yao and Zhong, 2013). For adolescents and young adults, online interaction can be an important medium for building and maintaining social relationships with peers and exploring identities (Boyd, 2014; Ellison, Blackwell, Lampe, & Trieu, 2016). However, the intensive use of social media may also endanger well-being and development, as it can relate to negative outcomes, such as body image concerns, internalizing symptoms (Marengo, Longobardi, Fabris, & Settanni, 2018), pathological use of SNS (Holmgren & Coyne, 2017; Yao and Zhong, 2013), and parasocial relations and social distrust (Baek, Bae, & Jang, 2013).

The neurological basis for emotions, motivation, and cognitive processing is still developing during adolescence and young adulthood, which manifests in relatively frequent risk-taking and immature decision-making at that age (Balogh, Mayes, & Potenza, 2013). Similarly, aggressive behavior tends to be most frequent during adolescence and then decreases toward young adulthood (Farrell, Sullivan, Esposito, Meyer, & Valois, 2005; Karriker-Jaffe, Foshee, Ennett, & Suchindran, 2008). Currently, online aggression has produced new forms of abuse among

young people (Kowalski, Giumetti, Schroeder, & Lattanner, 2014). As in offline aggression, cyberbullying and cyber victimization are most prevalent among younger people (Álvarez-García, Núñez, Barreiro-Collazo, García, 2017; Barlet & Coyne, 2014; Oksanen & Keipi, 2013).

4.2.2 Societal condition of Finland between 2013 and 2015

One aim of this dissertation is to examine changes in the online hate exposure among Finnish adolescents and young adults during 2013 and 2015. This requires contextualizing societal conditions that characterize the time period. Finland was hit by the global economic recession (i.e., "the Great Recession") in 2007, and the recovery was slow for years, as indicated by the prolonged youth unemployment (OECD, 2016) and record payment defaults (Oksanen, Aaltonen, & Rantala, 2015).

The stagnated economic situation preceded the so-called immigration crisis, caused by the conflicts in Syria and Iraq that extended throughout Europe. In Finland, there was a fivefold increase asylum seekers in 2015, reaching 32,000 that year. In Finnish society and around Europe, the increasing number of refugees further exacerbated the already existing ideological polarization between those opposing immigration and ethnic-cultural diversity and those taking a more liberal stance (Czymara & Schmidt-Catran, 2017; Haugerud, 2016; Kazharski, 2017; Wahlbeck, 2016). The increase in anti-immigration sentiments manifested in public demonstrations and in the increasing number of hate crimes, especially those targeting immigration centers (Finnish Ministry of the Interior, 2016; Tihveräinen, 2016). The theme of immigration was highly present in social media, which is part of a longer continuum in which social media and online social networks have offered the anti-immigration movement a medium for political mobilization and public agenda setting (Horsti, 2015).

The atmosphere in Europe was already insecure when 2015 culminated in the Paris terrorist attacks on November 13, as several strikes by international terrorist organizations had occurred around the world that year (Haugerud, 2016). Even though the direct consequences of the tragedy were in the French capital, the occurrence of such mass violence tend to cause the social climate of continuing

insecurity and a disproportionate fear of future attacks (Comer & Kendal, 2007). This secondhand terrorism can be even stronger in countries that have infrequent terrorist attacks, such as Finland (Comer, Bry, Poznanski, & Golik, 2016).

4.2.3 The cross-national context

In this dissertation, I study online hate exposure, offending, and victimization among adolescents and young adults in Finland and in a cross-national context, including Finland, Germany, the United Kingdom, and the United States (Studies 1–3). These nations are comparable regarding online risks, as they all are technologically advanced countries with high living standards (United Nations, 2013). However, there are sociodemographic differences between the four countries. The United States is the most populated of these countries with more than 300 million inhabitants, followed by Germany (82 million), the United Kingdom (63 million) and Finland (5.5 million). Finland has the lowest ethnic diversity with only 5.7% of the population being immigrants, and the proportion is significantly higher in other thee countries (13.2%, 14.4%, and 14.5% for the United Kingdom, Germany and the United States, respectively) (United Nations, 2015). In addition, Finland has the lowest income inequality, followed by the Germany, the United Kingdom and the United States (OECD, 2016).

According to the Organization for Economic Cooperation and Development (OECD, 2016), there are also differences between Finland, Germany, the United Kingdom, and the United States in the case of technology use and social cohesion indicators on the population level. Of these four nations, only Finland is above the OECD average regarding technology use skills. The general life satisfaction is highest in Finland, followed by the United States, Germany, and the United Kingdom. However, when it comes to trusting in other people, the United States has the lowest level of trust at the population level. Trust is highest in Finland followed by Germany and the United Kingdom, with all of these three countries having a trust level above the OECD average. Finnish people are the most likely to engage with online social networks, followed by the United Kingdom and Germany (no data available from the United States). The Americans and the British show the most prosocial behavior on the population level, followed by

Germans and Finns. However, there are no significant differences in cases of anti-social behavior between these countries (OECD, 2011).

There are also differences between these countries in the legislation regarding hate content online and offline (Hawdon et al., 2017; Waldron, 2012). According to Hawdon et al. (2017), the United States is at the liberal extreme, having traditionally favored freedom of speech even in cases of hatred-inciting material. On the other hand, Germany has more restricting legislation banning speech that incites hatred against certain social groups or attacks people's dignity. At this continuum, the United Kingdom can be placed closer to Germany, and Finland more closely resembles the United States, as Finnish legislation bans speech that threatens or insults certain social group, but in practice, these laws are rarely applied, and punishments are usually minor (Hawdon et al., 2017).

5 DATA AND METHODS

Next, I will discuss the data and methods utilized in this dissertation. The data, measures, and methods used for each research article included in the dissertation are also summarized in Table 1.

5.1 Data

5.1.1 YouNet survey

Studies 1 through 3 utilize *YouNet* survey data collected from Finnish ($n = 555$), German ($n = 978$), British ($n = 999$) and American ($n = 1,033$) adolescents and young adults between 2013 and 2014. The data were part of *Hate Communities: A Cross National Comparison Project* (funded by Kone Foundation, 2013–2016). Participants were aged 15 to 30 ($M_{FIN} = 22.59$, $SD_{FIN} = 4.21$; $M_{GER} = 23.21$, $SD_{GER} = 3.97$; $M_{UK} = 23.18$, $SD_{UK} = 4.14$; $M_{US} = 23.09$, $SD_{US} = 4.04$), and half of them were female (50% in Finland; 50.26% in Germany; 49.09% in the United Kingdom; and 50.34% in the United States). The survey was designed to investigate exposure and victimization to online hate and online hate offending among young people. The data were collected from the United States and Finland in spring 2013 and from the United Kingdom and Germany in spring 2014. In the online questionnaire, respondents were asked about their psychosocial and sociodemographic characteristics, online activities and online and offline social relations. For each country, respondents were recruited by Survey Sampling International (SSI), using panels that were balanced to demographically reflect adolescent and young adult populations in the country in terms of age, gender, and area of residence. In accordance with informed consent principles,

Table 1. Summary of research hypotheses, data, methods and measures by studies

	Study 1	Study 2	Study 3	Study 4	Study 5
Tested hypotheses (section 4)	H1, H2	H3, H4	H5, H6a, H6b	H7, H8, H9, H10, H11, H12, H13	H14, H15, H16
Data	YouNet (2013-2014)	YouNet (2013-2014)	YouNet (2013-2014)	YouGamble (2017), YouGamble-social media (2017)	YouNet (2013), SAMRISK Flash survey (2015)
Countries	Finland (n = 555), Germany (n=978), UK (n=999), U.S. (n=1,033)	Finland (n = 555), Germany (n=978), UK (n=999), U.S. (n=1,033)	Finland (n = 555), Germany (n=978), UK (n=999), U.S. (n=1,033)	Finland (N=1200, N=160)	Finland (N=555, N=192)
Respondents age	15-30	15-30	15-30	15-25	15-30
Statistical techniques	Logistic regression analysis	Logistic regression analysis	Linear regression analysis	Logistic regression analysis, structural equation modeling	Propensity score matching, logistic regression analysis
Dependent variables	Online hate offending	Online hate victimization	Subjective well-being	Online hate offending, self-stereotyping, group norm conformity	Online hate exposure
Independent variables	Offline trust, online trust, sense of belonging offline, sense of belonging online	Online social capital, offline social capital	Cybercrime victimization, victimization to offensive cybercrime, victimization to cyberfraud, sense of belonging offline, sense of belonging online	Impulsivity, internalizing symptoms, social identification online, social homophily online	Time
Covariates	Gender, age, online activity, living arrangements	Size of the online network, size of the offline network, gender, age, victimization offline, online activity	Country, gender, age, online activity	Age, gender, Facebook activity, YouTube activity	Age, gender, education, Living arrangements, Internet use

respondents were informed about the purpose and background of the study and their right to withdraw from the research at any point should they want to. In addition, participants were provided with contact information for possible further questions concerning the research project. After giving their consent, participants were provided a link to the online survey. The survey data were anonymous and no identifying information was collected from the respondents.

5.1.2 YouGamble and YouGamble Social Media surveys

Study 4 utilizes YouGamble ($N = 1200$, first substudy) and YouGamble Social Media ($N = 160$) data sets collected from Finnish adolescents and young adults in spring 2017 as part of *Problem Gambling and Social Media: Social Psychological Study on Youth Behavior in Online Gaming Communities* (funded by the Finnish Foundation for Alcohol Studies, 2017–2019). Both surveys were designed to investigate social media use, online risk behavior, addictive behaviors, and well-being among young people. For this dissertation, items concerning online hate offending and its hypothesized correlates were included in the analyses. The respondents of the YouGamble data set mirror the Finnish adolescent and young adult population regionally and in terms of age and gender distribution (Sirola, Kaakinen, & Oksanen, 2018). Participant recruitment utilized the Finnish respondent panels of SSI. Respondents were contacted via email and provided with information concerning the survey as well as a link to an online survey. The respondents were aged 15 to 25 ($M = 21.29$; $SD = 2.85$), and 50% were female. The respondents of the YouGamble social media survey were recruited via Finnish discussion forums and SNS sites, and they were provided with a link to an online survey. Respondents were aged 15 to 30 ($M = 22.48$; $SD = 2.58$), and 57% ($n = 91$) of them were females.

   As the surveys included items concerning gambling and other forms of harmful behavior, the Academic Ethics Committee of the Tampere Region reviewed the research proposal in December 2016. No ethical problems were raised by the committee. Participants in both surveys were informed about the aims of the study and provided contact information for possible further questions. Afterward, the participation respondents were offered additional information concerning the survey and research project and contact information for potential further questions. The survey data was anonymous, and no identifying information was collected from the respondents.

   The *YouGamble social media* survey included a vignette experiment concerning behavior and self-stereotyping in social media interaction. For the experiment, respondents were randomly assigned into a salient group identity condition and a control condition for the experiment. In the salient group identity condition, respondents were told that they had been assigned into group C, which consisted

of similar respondents (based on answers in previous questions). Those in the control condition were given no group information.

In the vignette experiment, respondents were shown vignette scenarios concerning gambling-related social media content and asked to indicate whether they would "like" (thumbs up) or "dislike" (thumbs down) the content or not react to the content at all in a real social media setting. In the vignettes, respondents were shown a manipulated distribution of other respondents´ earlier reactions. In half of the vignettes, a majority (about 90%) had disliked the content, but in the other half the majority had liked the content. For those in the salient identity condition, this distribution was framed by their in-group members' earlier responses. We also manipulated the stance toward gambling presented in the vignette content. In half of the vignettes, the content was pro-gambling-oriented (discussed the upside of gambling; e.g., entertainment), and the other half anti-gambling-oriented (discussed gambling-related harm; e.g., gambling problems). The third manipulated factor was narration of the content. Half of the vignettes were narrated as experience-based (first-person narration), and the other half was narrated as fact-based (third-person narration). For exact manipulations, see the vignettes translated into English in Appendix A

This 2 × 2 × 2 factorial design resulted in eight vignettes, in which each participant was shown four scenarios. Thus, vignettes were partitioned into two sets, and the factorial structure was designed in a manner where the majority disliked both pro- and anti-gambling content and experience-based and fact-based narration (Atzmüller & Steiner, 2010). Therefore, the group did not favor any form of gambling orientation or narration.

5.1.3 SAMRISK Flash survey

Study 5 utilizes the Finnish *YouNet* data set (*N* = 555) and the *SAMRISK Flash survey* data set (*N* = 193). The *SAMRISK Flash survey* was collected in the beginning of December 2015, approximately one month after the Paris attacks, as part of the *Disruption, Social Capital and Resilience: A Longitudinal and Comparative Approach* project (funded by The Research Council of Norway, 2014–2017). The original survey (*N* = 1,003) included respondents with ages ranging from 16 to

84 years, but only those respondents who were aged 16 to 30 were selected for this study (age, $M = 23.13$, $SD = 3.78$; 57.73% female). The data collection and respondent recruitment was administrated by TNS Finland, and the sample was stratified to mirror the Finnish population in terms of age, gender, and region. Quotas allowed small differences in the official population statistics. The survey was designed to examine psychosocial characteristics and offline and online activities after the Paris attacks. The survey also included questions concerning online hate exposure. The survey data was anonymous, and no identifying information was collected from the respondents.

5.2 Measures and methods

5.2.1 Study 1

In Study 1, it was tested whether online hate offending associated with cognitive social capital in offline and online environments. *Online hate offending* was measured with a survey item in which respondents were asked whether they had produced online material that someone had interpreted as threatening or degrading with either a yes or no response. A dummy variable indicating whether a respondent had engaged in online hate offending (0, *no*; 1, *yes*) was coded on the basis of this item. In line with earlier research, both offline and online cognitive social capital were assessed as trust and a sense of belonging (De Silva et al., 2006; Harpman et al., 2002; Yip et al., 2007). Trust was measured by survey items in which respondents were asked to rate whether they agreed with the following phrases: "Good friends can be trusted" (offline trust), and "Only people met on the Internet can be trusted" (online trust) on a scale of 1 (*you cannot be too careful*) to 10 (*can be trusted entirely*). The sense of belonging was measured by items in which respondents were asked to rate how closely they felt they belonged to their circle of friends (offline sense of belonging) and in an online community (online sense of belonging) on a scale from 1 (*not closely at all*) to 5 (*very closely*). In addition, gender, age, online activity (i.e., number of used online services), and living

arrangements (i.e., whether the respondent was living alone) were used as covariates in the analyses.

To assess the associations between online hate offending and social capital logistic regression analysis with a maximum likelihood estimation was conducted. Logistic regression analysis is a multivariate method used to predict the outcome of a dichotomous variable ($y = 1$) as a function of continuous and categorical dependent variables. In other words, logistic regression analysis estimates the (log of) odds of studied outcomes given the certain values of dependent variables (Keith, 2015; Tabachnik & Fidell, 2001). The modeled associations are often reported as regression coefficients or odds ratios. However, interpreting regression coefficients and odds rations between logistic regression models can be challenging, as they are affected by the unobserved heterogeneity present in logistic regression models (Mood, 2010). The reporting of average marginal effects (AME) is one way to adjust for this problem. AMEs indicate the change in the probability (versus the odds) of y = 1, resulting from a one-unit change in certain independent variables. In the case of regression models in Study 1, AMEs and their standard errors are reported along with the coefficient of model determination (McFadden's adjusted R^2).

5.2.2 Study 2

Study 2 examined the associations between online hate victimization and cognitive social capital.

Online hate victimization was measured with a survey item in which respondents were asked whether they agreed with the statement "I have personally been the target of hateful or degrading material online" (0, *no*; 1, *yes*). As in Study 1, online and offline cognitive social capital were measured as trust and as a sense of belonging to an online community or group of friends (see the Study 1 section). However, in this study, measures of trust and sense of belonging were combined into composite variables measuring both online and offline cognitive social capital. Due to the various scales for the indicator variables, we used standardized (*z*-score) variables to form composite variables for online and offline social capital. The scale reliabilities were .70 for online and .71 for offline social capital

50

(Cronbach's α calculated for aggregate data). Furthermore, the size of an online network (number of Facebook friends) and an offline network (number of good friends), gender, age, victimization experiences offline, and online activity were used as covariates.

As in Study 1, the logistic regression analysis with maximum likelihood estimation was used as a multivariate method for predicting the probability of online hate victimization given the values of independent variables. To test the relationship between online hate victimization and social capital and the linearity of the relationships, three logistic regression models with estimated regression coefficients, standard errors, and p values were conducted for the aggregate data as well as separately for each country. Model 1 estimated the linear relationship between online social capital and online hate victimization. In Model 2, the relationship between online social capital and online hate victimization was estimated by adding a squared online social capital term in the model. In Model 3, in turn, the third-order online social capital term was introduced for the purpose of estimating the cubic relationship. The model with the highest significant online capital term was reported for each country (Bliese & Ployhart, 2002; Keith, 2015). In addition to regression coefficients, we present margin plots to facilitate the interpretation of the relationship between online social capital and the change in the probability of online hate victimization. Plots were based on predictive margins and their confidence intervals that were calculated from the predictions of our logistic models. As in the case of AME coefficients, these estimates indicated the average probabilities of online hate victimization in terms of online social capital values. Plots are presented separately for each country sample.

5.2.3 Study 3

In Study 3, we analyzed the associations between online crime victimization and subjective well-being and whether offline and online belonging moderated this association. Cybercrime victimization, offensive cybercrime victimization, and cyberfraud victimization were measured with dichotomous variables (0, *no victimization experience during the past three years*; 1, *at least one victimization experience*

during the past three years). The variable measuring cybercrime victimization in general was based on a survey item in which the respondents were asked whether they had been subjected to cybercrime during the past three years. Those respondents answering yes to this question were then addressed with a follow-up question on whether the crime had been defamation, illegal threat, identity theft, fraud, sexual harassment, or something else. Two victimization variables were then coded on the basis of these answers. Offensive cybercrime victimization was coded as 1 if the respondent had experienced defamation, threat, or sexual harassment during the past three years. Meanwhile, cyberfraud victimization was coded as 1 if the responded had experienced identity theft or online fraud during the past three years. Both variables were coded as 0 for respondents with no victimization experiences. The country information, gender, age, and online activity were used as covariates in the multivariate analyses.

Subjective well-being was measured with a composite variable based on the following survey items: (a) "All things considered, how happy would you say you are?" and (b) "All things considered, how satisfied are you with your life as a whole nowadays?" For both items, the measurement scale ranged from 1 (*extremely unhappy or extremely unsatisfied*) to 10 (*extremely happy or extremely satisfied*). Both of these items are validated and widely used measures of subjective well-being (Abdel-Khalek, 2006; Diener, 2000; Jorm & Ryan, 2014). Responses to these questions were then combined into a sum variable with a scale from 2 to 20 and good internal consistency (Cronbach's α = .88). *Social belonging* in *offline and online* contexts was measured with variables based on survey items in which respondents were asked to indicate, on a scale from 1 to 5, how strongly they felt a sense of belonging to their families, circles of friends, school or work communities, and online communities. As such, the first three variables measured participants' social belonging to primary groups originating in the offline setting, whereas the last measures focused on social belonging to communities originating in, yet not necessarily restricted to, the online environment. The variable measuring social belonging in the offline context was formed by combining answers to the first three questions into a sum variable with a scale of 3 to 15 and sufficient internal consistency (Cronbach's α = 0.69). The variable measuring online social belonging was based on the question concerning participants' sense of belonging to an online community.

In the predictive analysis, the association between cybercrime victimization and SWB was assessed by using hierarchical linear regression models with ordinary least squares estimation. In the regression models, unstandardized and standardized (ß) regression coefficients and their standard deviations and statistical significances, coefficients of effect sizes (partial η^2), and model determination coefficients (McFadden's adjusted R^2) were estimated. The first model included SWB as a dependent variable and our covariates as predictors. In the second model (main effect model), cybercrime victimization was added as a predictor. The potential moderating role of social belonging, in turn, was analyzed with a third model including interaction terms between cybercrime victimization and offline and online social belonging (moderation effect model). Models 2 and 3 were estimated separately for each type of cybercrime victimization (cybercrime in general, offensive cybercrime, and cyberfraud). To elaborate on the hypothesized moderation effect, we present adjusted prediction plots to visualize how social belonging moderated the association between cybercrime victimization and SWB. Adjusted predictions reveal the effect of cybercrime victimization separately for low (one standard deviation below the mean), average (the mean), and high (one standard deviation above the mean) social belonging, with other covariates being kept fixed at their mean values.

5.2.4 Study 4

The Study 4 was divided two sub-studies to analyze how online hate offending is associated with online group processes (social homophily, social identification, and self-stereotyping) and individual risk factors (impulsivity and internalizing symptoms). For measuring *online hate offending*, respondents were asked how often they sent messages in social media that offended or threatened other users. Respondents were given the following reply options: 1 = "never," 2 = "less than once a year," 3 = "at least once a year," 4 = "at least once a month," 5 = "more than once a moth," 6 = "once a week," and 7 = "daily." The measure was dichotomized for further analyses to indicate whether the respondent had ever been engaged in online hate offending.

In the first sub-study, impulsivity, internalizing symptoms, online social identification, and online social homophily were used as independent variables predicting online-hate-based offending. Impulsivity was measured with the Eysenck Impulsiveness Scale (Dussault, Brendgen, Vitaro, Wanner, & Tremblay, 2011; Eysenck & Eysenck, 1978). The five-item scale showed acceptable internal consistency (Cronbach's α = .74), and thus, all items were added up to create a composite variable, with a higher figure indicating a higher level of impulsivity. Internalizing symptoms were measured with the General Health Questionnaire (Goldberg & Blackwell, 1970; Goldberg et al., 1997). The 12-item scale had good reliability and a Cronbach's α of .88. In accordance with the ordinal coding method for the General Health Questionnaire, the final variable for our analysis was conducted by adding up all 12 items (Pevalin, 2000). Online social identification was measured with two items in which respondents were asked to assess how well the following phrases described them: "In social media, I'm part of a community that is an important part of my identity" and "In social media, I'm part of a community that I'm proud of." For both items, the response scale ranged from 1 (*does not refer to me at all*) to 10 (*refers to me completely*). These items were modified from previous social psychological operationalizations of social identification (Leach et al., 2008). Items had good internal consistency (Cronbach's α = .83), and thus, they were added up to create a count variable for further analysis. Online social homophily was measured with two items: "In social media, I interact only with people who are similar to me" and "In social media, I interact only with people who share my interests." Here, respondents were again asked to assess how well the two phrases described them on a scale from 1 (*does not refer to me at all*) to 10 (*refers to me completely*). These items were modified from the attitude dimension of the perceived homophily scale (McCroskey, McCroskey, & Richmond, 2006; McCroskey, Richmond, & Daly, 1975). The two items were added together to create a scale with a Cronbach's α of .77. Age, gender, and Facebook and YouTube use were used as covariates in the analyses.

In the second substudy, we analyzed whether online hate offending was associated with self-stereotyping and group norm conformity in an online interaction experiment included in the *YouGamble* social media data set (see the YouGamble and YouGamble Social Media surveys section). The measures of online hate offending, group identity salience, self-stereotyping, and conformity to a negative group norm were included in the analyses. The same item was used

to measure online hate offending as in Study 1, and it was dichotomized in the same manner for the analyses. Information on whether the respondent was included in the salient group identity condition or the control condition (see the Data section) was included in the analysis as a dichotomous measure (0 = group identity condition; 1 = control condition). Self-stereotyping was measured with two items on a scale from 1 (*do not agree at all*) to 10 (*completely agree*): "I have a lot in common with other group members/respondents" and "I am similar to other group members/respondents" (adapted from Leach et al., 2008). These questions were presented to respondents after they had completed a social media vignette experiment (see the Data section). As the measure consisted of only two questions, the items were added up and used as an observed variable in our structural equation model. *Conformity to a negative group norm* was calculated as the number of dislikes in two vignettes with a majority disliking the content, minus the dislikes in the vignettes in which a minority disliked the content (Atzmüller & Steiner, 2010; Bergh et al., 2016).

In the first substudy, we conducted logistic regression analysis with maximum likelihood estimation to assess the associations between online hate offending and our predictors. In the models, odds ratios, standard errors, p values, AME coefficients, and pseudo-R^2 coefficients (McFadden's R^2) were estimated. Model 1 included all of our predictor variables and covariates. In Model 2, interactions between our personal risk factor variables, impulsivity and internalizing symptoms, and online social homophily were added. We used both the significance test of interaction terms and the likelihood ratio test to assess the statistical significance of the interaction effects. To elaborate on significant interactions in Model 2, we counted predictive margins, and based on these estimates, we present adjusted prediction plots to visualize interactions. In the second substudy, a structural equation model with maximum likelihood estimation was used to asses our theoretical model. For the modeled association in our structural equation model, β coefficients and their statistical significance are reported. We also report widely used fit indices, including χ^2 statistic along with degrees of freedom and the corresponding significance test, the root mean squared error of approximation with a 90% confidence interval, the standardized root mean squared residual, and the comparative fit index (Hu & Bentler, 1999).

5.2.5 Study 5

Study 5 analyzed the differences in the quantity and quality of online hate exposure before 2013 and after the 2015 Paris attacks (a few weeks after the assaults). Online hate exposure was measured in two data sets (see the Data section) by asking participants whether they had seen (in the past three months) online material that threatened or degraded individuals or social groups (0, *no*; 1, *yes*). The participants who had seen online hate content were then presented with a follow-up question concerning the characteristics to which the hateful material had related. Options included ethnicity or nationality, religious conviction, political views, sexual orientation, gender, disability, appearance, and terrorism. This measure of online hate and its subtypes was used in earlier research (Costello et al., 2016; Hawdon et al., 2015, 2016; Oksanen, Hawdon, Holkeri et al., 2014). Exposure to these hate content types was measured with dummy variables (0, *no exposure*; 1, *exposure to the particular type of online hate*). *The time variable* (i.e., the treatment variable) indicated whether the observation was made after the Paris attacks. The value of this variable was 0 for the respondents who had answered the survey before the attacks (the data set collected in 2013) and 1 for those included in the data set collected at the end of 2015. Age, gender, education, living arrangements (i.e., whether the respondent was living with his or her parents), and Internet use were used as covariates in the analyses.

These analyses used an analytical combination of the statistical technique of nearest-neighbor matching (Rubin, 1973; Stuart, 2010) and logistic regression analysis for estimating the effect of time on exposure to hateful online content. This chosen statistical technique allowed us to adjust for possible selection bias due to differences between the data sets in terms of sociodemographic characteristics and online use activity (for selection bias in observational data, see, e.g., Rickles, 2016; Stuart, 2010). With nearest-neighbor matching, we created two comparative sets of participants from the 2013 data set (the control group, or the group prior to the assault) and the 2015 data set (the treated group). In the process, every respondent in the treatment group was provided with a closest matching counterpart from the control group. The selection of control group respondents was based on the propensity score, a measure of the distance between observations based on the likelihood of belonging to the treatment

group. The likelihood was based on a logistic model using the selected covariates (i.e., age, gender, education, living arrangements, and Internet use) as predictors and membership in the treated group as an outcome variable.

In step two, we first combined the matched samples in one aggregate data set. Then, we estimated the effect of the Paris attacks (the treatment variable) as a change in the probability of online hate exposure by running a logistic regression model with the maximum likelihood estimation adjusted for the same covariates also used in the propensity score estimation. In our logistic regression models, AME coefficients, standard errors, z-values, and p-values were estimated. This two-step model combining the matching process and regression analysis is a recommended method for reducing selection bias and was used in earlier studies as well (Klement, 2016; Rubin & Thomas, 2000).

5.3 Research ethical reflection

Researching sensitive topics such as online hate content requires some ethical refection (Decker, Naugle, Carter-Visscher, Bell, & Seifert, 2011). Thus, ethical perspective has been considered when planning and conducting the research and when reporting the research findings. All the research surveys were formulated in manner that no actual hate contents were presented and no emotionally arousing tasks were involved. In case of a vignette experiment in Study 4 (see *YouGamble* and *YouGamble* Social Media surveys section), the research proposal was reviewed in advance by the Academic Ethics Committee of the Tampere Region.

All conducted surveys were computer administrated and participants could fill them without a presence of an interviewer. It is still possible, of course, that the questions concerning online hate or online crime would have caused distress among respondent with previous online victimization experiences, for example (See Decker et al., 2011). For that end, respondents were offered contact information for further questions or comments after filling the survey. All the surveys were anonymous and all respondents in the dataset were unidentifiable, which is of course particularly important when studying sensitive topics. To acquire informed consent, participants in all surveys were informed in advance

about the theme of the research and they were told that they are free to withdraw from the study at any point.

It is often discussed among criminological inquiry, that reporting research findings may cause stigma or deterministic attitudes towards certain groups of people (see e.g. Focquaer, 2018). To address this issue, all the analyses are conducted and results are reported in a manner that no specific individuals or social groups are being, or could be, referred to as perpetrators or victims of online hate. In addition, the limitations of causal inference are discussed.

6 OVERVIEW OF THE MAIN FINDINGS

6.1 Article I: Social capital and online hate production: A four-country survey article on crime law and social change

According to our results, online hate offending was rare among adolescent and young adults in all country samples, as the proportion of respondents reporting online hate offending ranged from 0.9% in Germany to 3.4% in the United Kingdom, 4.0% in Finland, and 4.1% in the United States. Online hate offending was associated with cognitive social capital indicators. An increase in trust toward a group of friends was associated with a decreased likelihood of online hate offending. A sense of belonging to a group of friends, however, was not significantly related to hate offending. Of the online cognitive social capital indicators, both trust and sense of belonging to an online community were significantly and positively associated with online hate offending. Thus, our results supported both the communitarian (H1) and agonistic (H2) hypotheses, as cognitive social capital invested in offline and online personal networks appears to have a different function when it comes to hostile online behavior.

Of our covariates, only gender was significantly associated with online hate offending, with males being the more likely offenders. In addition, young people who were living by themselves were more likely than others were to engage in online hate offending, yet this association was slightly nonsignificant.

6.2 Article II: How social capital associates with online hate victimization?

According to our results, the prevalence of online hate victimization experience varies between countries. Victimization was most prevalent in the U.S. data set, where 16% of young people had been the targets of threatening or degrading online material. The United States was followed by the United Kingdom, Finland, and Germany, with 12%, 10%, and 4%, respectively, reporting such an experience in the past three months.

Online cognitive social capital was associated with online hate victimization in all data sets. In the United States and Germany, the association was linear, with a higher amount of cognitive social capital being associated with an increased likelihood of online hate victimization. In the case of the United Kingdom, the relationship was quadratic indicating that online cognitive social capital was related to a higher probability of online hate victimization and the positive effect increased toward the highest values of social capital. In the Finnish data set, the association was cubic, that is, the likelihood of victimization experience first increased when cognitive social capital increased, but the probability then started to decrease toward the highest values (values of one standard deviation above the mean or more). Offline cognitive social capital, in turn, was associated with a decreased likelihood of online hate victimization in all country data sets but Finland's. Thus, our results supported both research hypotheses concerning the role of online (H3) and offline (H4) cognitive social capital.

Of our covariates, online activity and offline victimization experiences were associated with online hate victimization as the most active online users and those abused in the offline environment were more likely to become victims of online hate. The sizes of online or offline social networks were not associated with online hate victimization.

6.3 Article III: Cybercrime victimization and subjective well-being: An examination of the buffering effect hypothesis among adolescents and young adults

Only a relatively small proportion of respondents reported having been victims of cybercrime. The proportion was highest in the United Kingdom (7.4%) followed by the United States (6.1%), Germany (6.0%), and Finland (6.0%). Offensive cybercrime victimization was more prevalent compared with cyberfraud victimization in all country data sets.

As hypothesized, cybercrime victimization was associated with lower subjective well-being (SWB) (H5). The association was relatively weak, but it was strongest in the case of offensive cybercrime victimization. Our moderation analysis partly supported our hypotheses; belonging to offline social networks significantly buffered the negative association between cybercrime victimization and SWB (H6a), as the negative association mainly concerned those with weak belonging to offline networks. Online belonging did not have such a buffering effect (H6b). Of the included covariates, being Finnish and a possessing strong sense of belonging to offline networks predicted higher SWB.

6.4 Article IV: Impulsivity, internalizing symptoms, and online group behavior as determinants of online hate

According to our results, the majority of adolescents and young adults had never threatened or degraded others in their online interactions. 78% of the respondents in the first substudy and 75% in the second substudy reported having never been engaged in online hate offending whereas the remainder reported offending less than once a year or more often.

Both online group behavior and individual risk factors predicted online hate offending. As hypothesized, impulsivity (H7), internalizing symptoms (H8), and social homophily (H9) were all positively associated with online hate offending. In addition, internalizing symptoms and social homophily had an interaction

effect on hate offending, as offending was more likely to occur among those who highly preferred similar online contacts and had high internalizing symptoms (H10). Unlike hypothesized, however, online social identification was not significantly associated with online hate offending (H11).

The results based on survey experiment in second substudy also supported our hypotheses. As hypothesized, online hate offenders were more likely than others to self-stereotype in online vignette scenarios (H12) and this was true only in the case of salient group identity (H13). The indirect association between online hate offending and conformity to a negative group norm via self-stereotyping was almost significant in the salient group identity condition ($p = .074$) and non-significant in the control condition ($p = .537$). Of our covariates, younger age and male gender were associated with more likely online hate offending.

6.5 Article V: Did the risk of exposure to online hate increase after the November 2015 Paris attacks? A group relations approach

As hypothesized (H14), exposure to threatening or degrading material was substantially higher after the Paris attacks than it had been in 2013. The probability of online hate exposure was 27% higher among adolescents and young adults after the Paris attacks. The qualitative change in online hate exposure was also in line with our hypotheses (H15 and H16). Whereas the encountered online hate material in 2013 related equally often to sexual orientation and nationality or ethnicity followed by appearance, the three most frequently witnessed hate types in 2015 were nationality or ethnicity, religious conviction, and terrorism. The increase was strongest in the case of hate related to terrorism (42%), religious conviction (34%), ethnicity or nationality (28%), and political views (19%). As hypothesized, exposure to other types of online hate (i.e., gender, disability, sexual orientation, and appearance) did not increase at the same magnitude or did not increase at all during the time frame.

7 DISCUSSION

The main aim of this dissertation research was to examine online hate offending, victimization, and exposure among adolescents and young adults. The dissertation research included five studies that analyzed online hate from different perspectives. Studies 1 and 4 analyzed online hate offending and its associations with offline and online cognitive social capital (Study 1) and online group behavior (Study 4). Studies 2 and 3, in turn, examined victimization experiences. In Study 2, the associations between online hate victimization and online and offline cognitive social capital were analyzed. Study 3 concentrated on cybercrime victimization—specifically, whether offensive cybercrime victimization and cyberfraud victimization were associated with lower subjective well-being. It was also tested whether a sense of belonging to offline and online social networks buffered the potential negative association. In Study 5, it was analyzed whether the societal condition is manifested in the temporal change in online hate exposure. The studies were conducted among Finnish young people and in the cross-national context, including Germany, the United Kingdom, and the United States as well (Studies 1–3).

According to the results of this dissertation, the clear majority of young online users in Finland, Germany, the United Kingdom, and the United States does not threaten or degrade others in their online communication. In the international comparison between 2013 and 2014 (Studies 1 and 2), online hate victimization appeared to be more prevalent compared with online-hate-based offending in all of the countries. Given the even higher proportion of young people who had witnessed hate content online in 2013 (Study 5), it appears that hateful content is being amplified in online communication: a relatively small fraction of users are producing the hate content, but it can victimize many more and is being encountered among an even larger number of online users. These results are in line with earlier studies reporting that online hate is only a marginal part of online communication (Jakubowicz et al., 2017; Williams & Burnap, 2016) but remains

highly visible in mainstream online communication (Costello et al., 2016; Hawdon et al., 2017; Oksanen, Hawdon, & Räsänen, 2014).

The amplifier effect of online hate is likely due to the enhanced possibilities of one-to-many online communication that allows users to reach huge audiences. In addition, as people are persistently and pervasively connected to a wide network of various social ties in social media, they also encounter large amounts of shared information, which may include disturbing material (Hampton, 2016; Hampton et al., 2014), such as hostile ideas and opinions. Hostile content and discussion can also gain visibility due to SNSs' strategy to maximize the number of connections and the amount of data flowing within their platforms (Van Dijck, 2012, 2013b). This can lead to the promotion of such socially undesirable phenomena as harassment campaigns (Pew Research Center, 2017a).

In the case of online hate offending, however, it is worth noticing that the results at two time points using two measurements revealed significant variation in the proportion of Finnish adolescents and young adults who had engaged in online offending. In the survey data collected in 2013, 4% of respondents reported having produced threatening or degrading online content, whereas the proportion was 22 to 25% in 2017. This might indicate an increase in hate offending among young people during the four-year period, but drawing this conclusion would require more evidence, as the age ranges of the respondents (15–30 in 2013 and 15–25 in 2017) and the measurement methods differed between the two surveys.

Online hate is related to offline and online social relations. Studies 1 and 2 indicated that users who had the highest amounts of cognitive social capital embedded in their online social networks were also the most likely online hate offenders and victims. In the case of offline cognitive social capital, the association was reversed, as those with the highest amounts of social capital were less likely to offend others or to be victimized in their online interactions. Thus, social capital in the offline environment appears to have a public, socially integrative function (Putnam & Goss, 2002), whereas online social capital appears to function in a more antagonistic manner, making conflictual online behavior more likely (Julien, 2015).

In social media, people can execute selectivity over the social networks with which they engage (Bakshy et al., 2015). This may lead to the formation of similarly minded social networks but also antagonistic relations between various

online groups (Boutyline & Willer, 2017; Densley & Peterson, 2017; Stern & Ondish, 2018; Yardi & Boyd, 2010; Zollo et al., 2017). Thus, the users with the highest amounts of social capital embedded in online social networks become, more often than others, involved in the clashes that take place among various networks. This explanation received further support from the finding in Study 4 that online social homophily was associated with online hate offending, especially among those with low psychological well-being. Of course, in some cases, social capital may be invested in online hate groups that explicitly endorse and reinforce the expressing of hateful ideologies (Caiani & Parenti, 2013; Douglas et al., 2005; Gerstenfeld et al., 2003; Levin, 2002; Lucassen & Lubbers, 2012; Lunstrum, 2017; M. A. Wong et al., 2015). In addition, members with the highest amounts of social capital tend to be the most active participants and information disseminators (Chang & Chuang, 2011; Yen, 2016), which may increase the likelihood of victimization.

According to Kawachi et al. (1999), one mechanism through which social capital is likely to enhance safety and to inhibit harmful behaviors is that social capital fosters social control, which then makes deviant behavior less likely. However, in the case of online networks, people may be motivated to form online social networks that share and validate their behaviors that are considered to be deviant in other social contexts (Dengah et al., 2018). This could explain why social capital in the traditional offline context decreased the likelihood of being involved in hostile online communication (either as a victim or a perpetrator), whereas online social networks appear to have an opposite effect. This same mechanism of deviant peer norms has also been found in the offline environment in the case of criminal gangs, for example (Densley & Peterson, 2017; Peterson & Densley, 2017).

According to Study 3, being offended online is related to a weaker psychological well-being, at least in the case of occurrences identifiable as crimes (Study 3). Being a victim of online offensive crime seems to have a stronger negative effect compared with being a cyberfraud victim. Social relations can buffer the negative association, but, again, offline and online social networks appear to function differently. Social relations to offline primary groups protected adolescents and young adults from the harmful effects of cybercrime victimization as suggested by the buffering hypothesis (Cohen & Wills, 1985). This suggests that even though online communities may be important for young

people, and especially for those with weak social ties in the offline environment (Cole et al., 2017; Leist, 2013; Mesch, 2012), they do not function as equivalent psychosocial buffers to offline primary groups.

In addition to social capital, online hate offending appears to be related to social identity dynamics. Even though online hate offending was not, per se, associated with online social identification in Study 4, online hate offenders were more likely to self-stereotype and thus to follow negative group norms in their online behavior. This is in line with earlier research indicating that social identification is not necessarily related to out-group antagonism (Brown, 2010; Jackson & Smith 1999; Mummendey et al., 1992). Instead, social identification is more likely related to out-group antipathies when it leads to a depersonalized self-concept (e.g., self-stereotyping) or a salient rivalry between the in-group and out-group (Jackson & Smith 1999). Depersonalization was found to be particularly characteristic of online interaction in earlier studies (Lea & Spears, 1991; Reicher, Spears, & Postmes, 1995), and it has been reported to be related to online antisocial behavior (Christopherson, 2007; Fox & Tang, 2014; Peterson & Densley, 2017; Rains et al., 2017; Tang & Fox, 2016). However, as the social identity model of deindividuation effects model highlights, perceived in-group norms should play an important role in shaping depersonalized group behavior (Reicher et al. 1995; Reicher et al., 2016). Thus, it is likely that self-stereotyping would not be related to online hate offending if one's in-group norms endorse prosocial behavior.

The results of Study 5 indicated that online hate is also related to intergroup relations on a societal level. At times of public unsafety, uncertainty, and political polarization in the past, exposure to online hate became dramatically more frequent among Finnish adolescents and young adults. In addition to a quantitative increase in hate exposure, hate content was directed against certain social groups. Whereas hate material was increasingly related to terrorism, ethnicity or nationality, religious convictions, or political views, such less heated themes as gender, appearance, sexual orientation, and disability did not increase at the same magnitude, if at all. This is in line with earlier research reporting increased online hate content targeting ethnic and religious groups after triggering events (Awan & Zempi, 2016; Williams & Burnap, 2016). According to SIA, times of insecurity and uncertainty make certain social categories more relevant for making sense of social reality and thus for self-concept and social comparison

(Tajfel & Turner, 1979). Furthermore, a perceived threat and uncertainty are likely to reinforce identification with clearly bounded in-groups (such as nationalities) and antagonism toward out-groups (Greenberg et al., 1986; Hogg, 2007, 2014; Stephan et al., 2009). In online spaces, this is manifested in the increased presence of hate content. This trend may be self-reinforcing, as the increasing visibility of hateful opinions in social media can be found to be empowering among those holding similar views, which, in turn, lowers the threshold of expressing one's own ideas (Barkun, 2017; Chun & Lee, 2017). Thus, online hate trends can be seen as a reflection of wider societal group-relations (George, 2017). Thus, online hate may induce a vicious cycle in which strained group-relations are reflected in online hate and the hostilities expressed online, in turn, further polarize the offline social relations (see Chan & Fu, 2017).

7.1 Limitations

As this dissertation has explored online hate using cross-sectional data, all reported results must be interpreted as associations rather than causal relations. All directions in statistical models (i.e., dependent and independent variables) and their interpretations are based on a theoretical framework rather than on temporality or variable manipulation, for example. This limitation must also be acknowledged when interpreting the results of Studies 4 and 5, which used experimental design and propensity score matching to account for some of the problems in observational data but were still based on correlational analysis.

The measurement of online hate should also be acknowledged when interpreting the results. First, online hate was measured as the subjective experience of seeing (exposure), producing (offending), or being the target of (victimization) online material that threatens or degrades individuals or social groups. Thus, some idiosyncratic disagreement on the meaning of threatening or degrading content likely exists. As in the case of hate speech, online hate refers to the heterogeneous collection of expressive phenomena (see Brown, 2017), and the diversity present in this collection cannot be easily reduced to survey items. Thus, the personal interpretation of threatening or degrading material might be arbitrary to some degree, but it also captures the experience of being a witness,

offender, or victim of online hate. Second, questions concerning online hate can be affected by so-called social undesirability bias, which refers to the underreporting of some phenomena possibly considered to be sensitive among respondents (Tourangeau & Yan, 2007). However, this problem was addressed in the present studies, as all of the online surveys were self-administered, and as respondents were able to complete the surveys by using their own devices without the presence of an interviewer. These techniques have been found to reduce social undesirability bias (Tourangeau & Yan, 2007).

In addition to online hate measurement, the operationalization and measurement of social capital in this dissertation require some reflection. Social capital has been conceptualized and measured in various manners in past research (Bourdieu, 1984; Coleman, 1988; Putnam, 1993, 2000). In this dissertation, I approached social capital form the network capital perspective (Lin, 1999; Wellman, 1979; Wellman, 2001). In addition, the studies concentrated on the cognitive aspects of social capital, that is, the perceived quality of social relations (De Silva et al., 2005; Harpham et al., 2002). Of course, other possible approaches would have been available within the tradition. For example, concentrating on more structural elements or on bridging and bonding social capital (Putnam, 2000) probably would have provided unique insight into the phenomenon of online hate and its associations with social relations.

The dissertation used a cross-national data set in examining online hate in the international context. However, Finland, Germany, the United Kingdom, and the United States are all relatively similar Western countries with high living standards. Thus, an international comparison including a wider range of nations could offer more information on the cultural characteristics of online hate. On the other hand, the relative similarity among the chosen countries can also be seen as valuable for such a comparison, as the countries share many features concerning online communication and online risks, for example (United Nations, 2013).

7.2 Conclusion

Social media and SNSs are increasingly important forums of human interaction, entertainment, business, and learning, for example. Adolescents and young adults have been especially active in adopting new technology-driven platforms. However, social media does also have some less desirable social consequences. As these technologies allow for efficient connectivity to other users, they also facilitate forms of virtual abuse, of which online hate is an example.

According to the results of this dissertation, social media facilitates both connection and disconnection between users. The disconnect is manifested in the results indicating that a high amount of social capital invested in online communities is related to more probable online hate offending and victimization. In addition, homophilic social networks and perceiving oneself as an interchangeable part of an online group are associated with online hate offending. Thus, increased connection to online groups seems to come with increased conflicts and distinctions. In addition, online hate is connected to societal situation and social tensions associated with it. Thus, hostilities in social media are likely to further accentuate inter-group conflicts and distinctions.

REFERENCES

Abbas, R., & Mesch, G. (2018). Do rich teens get richer? Facebook use and the link between offline and online social capital among Palestinian youth in Israel. *Information, Communication & Society*, *21*(1), 63–79. doi:10.1080/1369118X.2016.1261168

Abdel-Khalek, A. M. (2006). Measuring happiness with a single-item scale. *Social Behavior and Personality*, *34*(2), 139–150. doi:10.2224/sbp.2006.34.2.139

Allen, J., & Norris, G. (2011). Is genocide different? Dealing with hate speech in a post-genocide society. *Journal of International Law and International Relations*, *7*(1), 146–174.

Allport, G. W. (1954). The nature of prejudice. Reading, MA: Addison-Wesley.

Álvarez-García, D., Núñez, J. C., Barreiro-Collazo, A., & García, T. (2017). Validation of the Cybervictimization Questionnaire (CYVIC) for adolescents. *Computers in Human Behavior*, *70*, 270–281. doi:10.1016/j.chb.2017.01.007

An, S., & Jang, Y. (2018). The role of social capital in the relationship between physical constraint and mental distress in older adults: A latent interaction model. *Aging & Mental Health*, *22*(2), 245–249. doi:10.1080/13607863.2016.1247431

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, *53*, 27–51. doi:10.1146/annurev.psych.53.100901.135231

Atzmüller, C., & Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology*, *6*(3), 128–138. doi:10.1027/1614-2241/a000014

Awan, I., & Zempi, I. (2016). The affinity between online and offline anti-Muslim hate crime: Dynamics and impacts. *Aggression and Violent Behavior*, *27*, 1–8. doi:10.1016/j.avb.2016.02.001

Awgu, E., Magura, S., & Coryn, C. (2016). Social capital, substance use disorder and depression among youths. *American Journal of Drug and Alcohol Abuse*, *42*(2), 213–221. doi:10.3109/00952990.2015.1111900

Baek, Y. M., Bae, Y., & Jang, H. (2013). Social and parasocial relationships on social network sites and their differential relationships with users' psychological well-being. *Cyberpsychology, Behavior, and Social Networking*, *16*(7), 512–517. doi:10.1089/cyber.2012.0510

Bahns, A. J., Pickett, K. M., & Crandal, C. S. (2011). Social ecology of similarity: Big schools, small schools and social relationships. *Group Processes and Intergroup Relations*, *15*(1), 119–131. doi:10.1177/1368430211410751

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. doi:10.1126/science.aaa1160

Balogh, K. N., Mayes, L. C., & Potenza, M. N. (2013). Risk-taking and decision-making in youth: Relationships to addiction vulnerability. *Journal of Behavioral Addictions*, *2*, 1–9. doi:10.1556/JBA.2.2013.1.1

Banks, J. (2010). Regulating hate speech online. *International Review of Law, Computers & Technology*, *24*(3), 233–239.

Barkun, M. (2017). President Trump and the "fringe." *Terrorism and Political Violence*, *29*, 437–443. doi:10.1080/09546553.2017.1313649

Barlett, C., & Coyne, S. M. (2014). A meta-analysis of sex differences in cyber-bullying behavior: The moderating role of age. *Aggressive Behavior*, *40*, 474–488. doi:10.1002/ab.21555

Baumeister, R. F. (1997). *Evil: Inside human cruelty and violence*. New York, NY: W. H. Freeman.

Baumeister, R. F., Ainsworth, S. E., & Vohs, K. D. (2016). Are groups more or less than the sum of their members? The moderating role of individual identification. *Behavioral and Brain Sciences*, *39*, 1–56. doi:10.1017/S0140525X15000618, e137

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529. doi:10.1037/0033-2909.117.3.497

Benigni, M. C., Joseph, K., & Carley, K. M. (2017). Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter. *PLoS ONE* *12*(12), e0181405. doi:10.1371/journal.pone.0181405

Bergh, R., Akrami, N., Sidanius, J., & Sibley, C. G. (2016). Is group membership necessary for understanding generalized prejudice? A re-evaluation of why prejudices are interrelated. *Journal of Personality and Social Psychology*, *111*(3), 367–395. doi:10.1037/pspi0000064

Billig, M. (2002). Henri Tajfel's "cognitive aspects of prejudice" and the psychology of bigotry. *British Journal of Social Psychology*, *41*(2), 171–188. doi:10.1348/014466602760060165

Black, E. W., Mezzina, K., & Thompson, L. A. (2016). Anonymous social media: Understanding the content and context of Yik Yak. *Computers in Human Behavior*, *57*, 17–22. doi:10.1016/j.chb.2015.11.043

Bliese, P. D., Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*, *5*(4), 362–387. doi:10.1177/109442802237116

Bliuc, A.-M., Best, D., Iqbal, M., & Upton, K. (2017). Building addiction recovery capital through online participation in a recovery community. *Social Science & Medicine*, *193*, 110–117. doi:10.1016/j.socscimed.2017.09.050

Boase, J. (2008). Personal networks and the personal communication system. *Information Communication and Society*, *11*(4), 490–508. doi:10.1080/13691180801999001

Boase, J., & Wellman, B. (2005). Personal relationships: On and off the Internet. In D. Perlman and A. L. Vangelisti (Eds.), *Handbook of Personal Relations* (pp. 709–723). Oxford, England: Blackwell.

Bouchillon, B. C. (2014). Social ties and generalized trust, online and in person: Contact or conflict—The mediating role of bonding social capital in America. *Social Science Computer Review*, *32*(4), 506–523.

Bourdieu, P. (1984). *Distinction: A social critique of the judgment of taste*. Cambridge, MA: Harvard University Press.

Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *The handbook of theory and research for sociology of education* (pp. 241–258). New York, NY: Greenwood.

Boutyline, A., & Willer, R. (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology*, *38*(3), 551–569. doi:10.1111/pops.12337

Boyd, D. (2011). Social network sites as networked publics: Affordances, dynamics, and implications. In Z. Papacharissi (Ed.), *Networked self: Identity, community, and culture on social network sites* (pp. 39–58). New York, NY: Routledge.

Boyd, D. (2014). *It's complicated: The social lives of networked teens*. New Haven, CT: Yale University Press.

Brooks, M., Lowe, M., Graham-Kevan, N., & Robinson, S. (2016). Posttraumatic growth in students, crime survivors and trauma workers exposed to adversity. *Personality and Individual Differences*, *98*, 199–207. doi:10.1016/j.paid.2016.04.051

Brown, A. (2017). What is hate speech? Part 2: Family resemblances. *Law and Philosophy*, *36*(5), 561–613. doi:10.1007/s1098

Brown, R. (2010). *Prejudice: Its social psychology.* Chichester, England: John Wiley & Sons.

Brewer, M. B., & Kramer, R. M. (1986). Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. *Journal of Personality and Social Psychology*, *50*, 543–549.

Burke, B. L., Martens, A., & Faucher, E. H. (2010). Two decades of terror management theory: A meta-analysis of mortality salience research. *Personality and Social Psychology Review* 14 (2), 155–195.

Burnap, P., & Williams, M. L. (2015). Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy & Internet*, *7*(2), 223–242. doi:10.1002/poi3.85

Caiani, M., & Parenti, L. (2013). *European and American extreme right groups and the Internet.* Farnham, England: Ashgate.

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

Chan, C. & Fu, K. (2017). The Relationship Between Cyberbalkanization and Opinion Polarization: Time-Series Analysis on Facebook Pages and Opinion Polls During the Hong Kong Occupy Movement and the Associated Debate on Political Reform. *Journal of Computer-Mediated Communication, 22,* 266–283. doi:10.1111/jcc4.12192.

Chang, P. F., & Bazarova, N. N. (2016). Managing stigma: Disclosure-response communication patterns in pro-anorexic websites. *Health Communication*, *31*(2), 217–229. doi:10.1080/10410236.2014.946218

Chang, H. H., & Chuang, S. S. (2016). Social capital and individual motivations on knowledge sharing: Participant involvement as a moderator. *Information & Management*, *48*(1), 9–18. doi:10.1016/j.im.2010.11.001

Christopherson, K. M. (2007). The positive and negative implications of anonymity in Internet social interactions: On the Internet, nobody knows

you're a dog. *Computers in Human Behavior*, *23*, 3038–3056. doi:10.1016/j.chb.2006.09.001

Chun, J. W., & Lee M. J.(2017). When does individuals' willingness to speak out increase on social media? Perceived social support and perceived power/control. *Computers in Human Behavior*, *74*, 120–129. doi:10.1016/j.chb.2017.04.010

Cionea, I. A., Piercy, C. W., & Carpenter, C. J. (2017). A profile of arguing behaviors on Facebook. *Computers in Human Behavior*, *76*, 438–449. doi:10.1016/j.chb.2017.08.009

Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin*, *98*(2), 310–357. doi:10.1037/0033-2909.98.2.310

Cole, D. A., Nick, E. A., Zelkowitz, R. L., Roeder, K. M., & Spinelli, T. (2017). Online social support for young people: Does it recapitulate in-person social support; can it help? *Computers in Human Behavior*, *68*, 456–464. doi:10.1016/j.chb.2016.11.058

Coleman, J. S. (1988). Social capital in the creation of human capital [Supplemental material]. *American Journal of Sociology*, *94*, 95–120.

Comer, J. S., & Kendall, P. C. (2007). Terrorism: The psychological impact on youth. *Clinical Psychology: Science and Practice*, *14*(3), 179–212. doi:10.1111/j.1468-2850.2007.00078.x

Comer, J. S., Bry, L. B., Poznanski, B., & Golik, A. M. (2016). Children's mental health in the context of terrorist attacks, ongoing threats, and possibilities of future terrorism. *Current Psychiatry Reports*, *18*(9), 79. doi:10.1007/s11920-016-0722-1

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. doi:10.1016/j.chb.2016.05.033

Council of Europe. (2017). No hate speech movement: Campaign for human rights online. Retrieved from http://www.nohatespeechmovement.org/

Crocker, J., & Canevello, A. (2012). Self and identity. Dynamics of person and their situations. In K. Deaux & M. Snyder (Eds.), *The Oxford handbook of personality and social psychology* (pp. 263–286). New York, NY: Oxford University Press.

Czymara, C. S., & Schmidt-Catran, A. W. (2017). Refugees unwelcome? Changes in the public acceptance of immigrants and refugees in Germany in the course of Europe's "immigration crisis." *European Sociological Review, 2017*, *33*(6), 735–751. doi:10.1093/esr/jcx071

Decker, S. E., Naugle, A. E., Carter-Visscher, R., Bell, K., & Seifert, A. (2011). Ethical issues in research on sensitive topics: participants' experiences of distress and benefit. *Journal of Empirical Research on Human Research Ethics,* *6*(3), 55–64. doi:10.1525/jer.2011.6.3.55.

Dengah, H. J. F., Snodgrass, J. G., Else, R. J., & Polzer, E. R. (2018). The social networks and distinctive experiences of intensively involved online gamers: A novel mixed methods approach. *Computers in Human Behavior*, *80*, 229–242. doi:10.1016/j.chb.2017.11.004

Densley, J., & Peterson, J. (2017). Groups aggression. *Current Opinion in Psychology*, *19*, 43–48.

De Silva, M. J., Harpham, T., Tuan, T., Bartolini, R., Penny, M., & Huttly, S. (2006). Psychometric and cognitive validation of a social capital measurement tool in Peru and Vietnam. *Social Science & Medicine*, *62*(4), 941–953. doi:10.1016/j.socscimed.2005.06.050

De Silva M. J., McKenzie, K., Harpham, T., & Huttly, S. (2005). Social capital and mental illness: A systematic review. *Journal of Epidemiology and Community Health*, *59*(8), 619–627.

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, *55*(1), 34–43.

Dika, S. L., & Singh, K. (2002). Applications of social capital in educational literature: A critical synthesis. *Review of Educational Research*, *72*(1), 31–60. doi:10.3102/00346543072001031

Dodge, K. A., & Crick, N. R. (1990). Social information-processing bases of aggressive behavior in children. *Personality and Social Psychology Bulletin*, *16*(1), 8–22. doi:10.1177/0146167290161002

Douglas, K. M. (2007). Psychology, discrimination, and hate groups online. In A. Joinson, K. McKenna, T. Postmes, U.-D. Reips (Eds.), *The Oxford handbook of Internet psychology* (pp. 155–164). New York, NY: Oxford University Press.

Douglas, K. M., McGarty, C., Bliuc, A.-M., & Lala, G. (2005). Understanding cyberhate: Social competition and social creativity in online White supremacist groups. *Social Science Computer Review*, *23*(1), 68–76. doi:10.1177/0894439304271538

Durkheim, E. (2002). *Suicide: A study in sociology*. New York, NY: Routledge. (Original work published 1879)

Dussault, F., Brendgen, M., Vitaro, F., Wanner, B., & Tremblay, R. E. (2011). Longitudinal links between impulsivity, gambling problems and depressive symptoms: A transactional model from adolescence to early adulthood. *Journal of Child Psychology and Psychiatry*, *52*(2), 130–138. doi:10.1111/j.1469-7610.2010.02313.x

Elgar, F., Davis, C. D., Wohl, M. J., Trites, S. J., Zelenski, J. M., & Martin, M. S. (2011). Social capital, health and life satisfaction in 50 countries. *Health & Place*, *17*(5), 1044–1053. doi:10.1016/j.healthplace.2011.06.010

Ellison, N. B, Blackwell, L., Lampe, C., & Trieu, P. (2016). "The question exists, but you don't exist with it": Strategic anonymity in the social lives of adolescents. *Social Media + Society*, *2*(4), 1–13. doi:10.1177/2056305116670673

Ellison, N. B, Gray, R., Lampe, C., & Fiore, A. T. (2014). Social capital and resource requests on Facebook. *New Media & Society*, *16*(7), 1104–1121. doi:10.1177/1461444814543998

Ellison, N. B., Lampe, C., Steinfield, C., & Vitak, J. (2011). With a little help from my friends. How social network sites affect social capital processes. In Z. Papacharissi (Ed.), *Networked self: Identity, community, and culture on social network sites* (pp. 124–145). New York, NY: Routledge.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends": Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, *12*(4), 1143–1168. doi:10.1111/j.1083-6101.2007.00367.x

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. doi:10.1080/15205436.2011.619679

Esses, V. M., Medianu, S., & Lawson, A. S. (2013). Uncertainty, threat, and the role of the media in promoting the dehumanization of immigrants and refugees. *Journal of Social Issues*, *69*(3), 518–536. doi:10.1111/josi.12027

Eysenck, S. B. G., & Eysenck, H. J. (1978). Impulsiveness and venturesomeness: Their position in a dimensional system of personality description. *Psychological Reports*, *43*, 1247–1255. doi:10.2466/pr0.1978.43.3f.1247

Fahy, A. E., Stansfeld, S. A., Smuk, M., Smith, N. R., Cummins, S., & Clark, C. (2016). Longitudinal associations between cyberbullying involvement and adolescent mental health. *Journal of Adolescent Health*, *59*, 502–509. doi:10.1016/j.jadohealth.2016.06.006

Farkas, J., & Neumayer, C. (2017). "Stop fake hate profiles on Facebook": Challenges for crowdsourced activism on social media. *First Monday*, *22*(9). doi:10.5210/fm.v22i9.8042

Farrell, A. D., Sullivan, T. N., Esposito, L. E., Meyer, A. L., & Valois, R. F. (2005). A latent growth curve analysis of the structure of aggression, drug use,

and delinquent behaviors and their interrelations over time in urban and rural adolescents. *Journal of Research on Adolescence*, *15*(2), 179–204. doi:10.1111/j.1532-7795.2005.00091.x

Federico, C. M., Hunt, C. V., & Fisher, E. L. (2013). Uncertainty and status-based asymmetries in the distinction between the "good" us and the "bad" them: Evidence that group status strengthens the relationship between the need for cognitive closure and extremity in intergroup differentiation. *Journal of Social Issues*, *69*(3), 473–494. doi:10.1111/josi.12025

Finnish Ministry of the Interior. (2016). Väkivaltaisen ekstremismin tilannekatsaus 1/2016. Sisäministeriön julkaisu 23/2016 [A report on Violent Radicalisation and Extremism 1/2016]. Helsinki, Finland: Sisäministeriö. Retrieved from http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/75390/Vaki valtaisen_ekstremismin_NETTI.pdf

Flores-Yeffal, N. Y., Vidales, G., & Plemons, A. (2011). The Latino cyber-moral panic process in the United States. *Information, Communication & Society*, *14*(4), 568–589. doi:10.1080/1369118X.2011.562222

Focquaert, F. (2018). Neurobiology and crime: A neuro-ethical perspective. *Journal of Criminal Justice*. Advance online publication. doi:10.1016/j.jcrimjus.2018.01.001

Fox, J., & Tang, W. Y. (2014). Sexism in online video games: The role of conformity to masculine norms and social dominance orientation. *Computers in Human Behavior*, *33*, 314–320. doi:10.1016/j.chb.2013.07.014

Frank, C., Davis, C. G., & Elgar, F. J. (2014). Financial strain, social capital, and perceived health during economic recession: a longitudinal survey in rural Canada. *Anxiety, Stress, & Coping*, *27*(4), 422–438. doi:10.1080/10615806.2013.864389

Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). Countering online hate speech. *UNESCO Series on Internet Freedom*. Retrieved from http://unesdoc.unesco.org/images /0023/002332/233231e.pdf

George, C. (2017). Hate spin: The twin political strategies of religious incitement and offense-taking. *Communication Theory*, *27*(2), 156–175. doi:10.1111/comt.12111

Germany starts enforcing hate speech law. (2018, January 1). *BBC*. http://www.bbc.com/news/technology-42510868

Gerstenfeld, P. B., Grant, D. R., & Chiang, C. (2003). Hate online: A content analysis of extremist Internet sites. *Analysis of Social Issues and Public Policy*, *3*(1), 29–44. doi:10.1111/j.1530-2415.2003.00013.x

Goldberg, D. P., & Blackwell, B. (1970). Psychiatric illness in general practice. A detailed study using a new method of case identification. *British Medical Journal*, *1*, 439–443. doi:10.1136/bmj.2.5707.439

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine*, *27*(1), 191–197. doi:10.1017/S0033291796004242

Graham, T., Jackson, D., & Wright, S. (2016). "We need to get together and make ourselves heard": Everyday online spaces as incubators of political action. *Information, Communication & Society*, *19*(10), 1373–1389. doi:10.1080/1369118X.2015.1094113

Green, M. C. (2007). Trust and social interaction on the Internet. In A. Joinson, K. McKenna, T. Postmes, & U.-D. Reips (Eds.), *The Oxford handbook of Internet psychology* (pp. 43–52). Oxford, England: Oxford University Press.

The Guardian. (2018). Facebook files. https://www.theguardian.com/news/series/facebook-files

Haas, S. M., Irr, M. E., Jennings, N. A., & Wagner, L. M. (2011). Communicating thin: A grounded model of online negative enabling support groups in

the pro-anorexia movement. *New Media and Society*, *13*, 40–57. doi:10.1177/1461444810363910

Hampton, K. N. (2016). Persistent and pervasive community: New communication technologies and the future of community. *American Behavioral Scientist*, *60*(1), 101–124. doi:10.1177/0002764215601714

Hampton, K. N., Goulet, L. S., Rainie, L., & Purcell, K. (2011). Social networking sites and our lives: How people's trust, personal relationships, and civic and political involvement are connected to their use of social networking sites and other technologies. Washington, DC: Pew Research Center. Retrieved from http://pewinternet.org/Reports/2011/Technology-and -social-networks.aspx

Hampton, K. N., Rainie, L., Lu, W., Dwyer, M., Shin, I., & Purcell, K. (2014). *Social media and the "spiral of silence."* Washington, DC: Pew Research Center. Retrieved from http://www.pewinternet.org/2014/08/26/social-media-and-the-spiral-of-silence/

Hampton, K. N., Rainie, L., Lu, W., Shin, I., & Purcell, K. (2015). *Social media and the cost of caring.* Washington, DC: Pew Research Center. Retrieved from http://www.pewinternet.org/2015/01/15/the-cost-of-caring/

Hampton, K. N., & Wellman, B. (2003). Neighboring in Netville: How the Internet supports community and social capital in a wired suburb. *City & Community*, *2*(4), 277–311. doi:10.1046/j.1535-6841.2003.00057.x

Han, S. (2013). Compositional and contextual associations of social capital and self-rated health in Seoul, South Korea: A multilevel analysis of longitudinal evidence. *Social Science & Medicine*, *80*, 113–120. doi:10.1016/j.socscimed.2012.12.005

Harpham, T., Grant, E., & Thomas, E. (2002). Measuring social capital within health surveys: Key issues. *Health Policy Plan*, *17*(1), 106–111. doi:10.1186/1743-8462-3-2

Haslam, S. A., Ellemers, N., Reicher, S. D., Reynolds, K. J., & Schmitt, M. T. (2010). The social identity perspective today: The impact of its defining ideas. In T. Postmes & N. Branscombe (Eds.), *Rediscovering social identity* (pp. 341–356). New York, NY: Psychology Press.

Haugerud, A. (2016). Public Anthropology in 2015: Charlie Hebdo, Black Lives Matter, migrants, and more. *American Anthropologist*, *118*(3), 585–601. doi:10.1111/aman.12606

Hawdon, J., Oksanen, A., & Räsänen, P. (2015). Online extremism and online hate: Exposure among adolescents and young adults in four nations. Nordicom Information, 37(3–4), 29–37.

Hawdon, J., Oksanen, A., & Räsänen, P. (2017). Exposure to online hate in four nations: A cross-national consideration. *Deviant Behavior*, *37*(3), 254–266. doi:10.1080/01639625.2016.1196985

Helberger, A., Karppinen, K., & D'Acunto, L. (2016). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, *21*(2), 191–207. doi:10.1080/1369118X.2016.1271900

Hogg, M. A. (2007). Uncertainty-identity theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 69–126). San Diego, CA: Academic Press. doi:10.1016/S0065-2601(06)39002-8

Hogg, M. A. (2014). From uncertainty to extremism social categorization and identity processes. *Current Directions in Psychological Science*, *23*(5), 338–342. doi:10.1177/0963721414540168

Hogg, M. A., Abrams, D., Otten, S., & Hinkle, S. (2004). The social identity perspective: Intergroup relations, self-conception, and small groups. *Small Group Research*, *35*, 246–276. doi:10.1177/1046496404263424

Hogg, M. A., & Adelman, J. (2013). Uncertainty–identity theory: Extreme groups, radical behavior, and authoritarian leadership. *Journal of Social Issues*, *69*(3), 436–454. doi:10.1111/josi.12023

Hogg, M. A., Terry, D. J., & White, K. M. (1995). A tale of two theories: A critical comparison of identity theory with social identity theory. *Social Psychology Quarterly*, *58*(4), 255–269. doi:10.2307/2787127

Holmgren, H. G., & Coyne, S. M. (2017). Can't stop scrolling! Pathological use of social networking sites in emerging adulthood. *Addiction Research & Theory*, *25*(5), 372–382. doi:10.1080/16066359.2017.1294164

Horsti, K. (2015). Techno-cultural opportunities: The anti-immigration movement in the Finnish mediascape. *Patterns of Prejudice*, *49*(4), 343–366. doi:10.1080/0031322X.2015.1074371

Howard, M. C., & Magee, S. M. (2013). To boldly go where no group has gone before: An analysis of online group identity and validation of a measure. *Computers in Human Behavior*, *29*, 2058–2071. doi:10.1016/j.chb.2013.04.009

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. doi:10.1080/10705519909540118

Huesmann, L. R. (1988). An information processing model for the development of aggression. *Aggressive Behavior*, *14*(1), 13–24. doi:10.1002/1098-2337(1988)14:1<13::AID-AB2480140104>3.0.CO;2-J

Hutchens, M. J., Cicchirillo, V. J., & Hmielowski, J. D. (2015). How could you think that?!?! Understanding intentions to engage in political flaming. *New Media & Society*, 17(8), 1201–1219. doi:10.1177/1461444814522947

Islam, M., Merlo, J., Kawachi, I., Lindstro€m, M., & Gerdtham, U. (2006). Social capital and health: Does egalitarianism matter? A literature review. *International Journal for Equity in Health*, *5*(3). doi:10.1186/1475-9276-5-3

Jackson, J. W., & Smith, E. R. (1999). Conceptualizing social identity: A new framework and evidence for the impact of different dimensions. *Personality and Social Psychology Bulletin*, *25*(1), 120–135. doi:10.1177/0146167299025001010

Jakubowicz, A., Dunn, K., Mason, G., Paradies, Y., Bliuc, A.-M., Bahfen, N, . . . Connelly, K. (2017). *Cyber racism and community resilience: Strategies for combating online race hate.* Cham, Switzerland: Palgrave Macmillan.

Jans, L., Leach, C. W., Garcia, R. L., & Postmes, T. (2014). The development of group influence on in-group identification: A multilevel approach. *Group Processes & Intergroup Relations*, 18(2), 190–209. doi:10.1177/1368430214540757

Jasko, K., LaFree, G., & Kruglanski, A. (2016). Quest for significance and violent extremism: The case of domestic radicalization. *Political Psychology*, *38*(5), 815–831. doi:10.1111/pops.12376

Jones, L., Mitchell, K., & Finkelhor, D. (2013). Online harassment in context: Trends from three youth Internet safety surveys (2000, 2005, 2010). *Psychology of Violence*, *3*(1), 53–69. doi:10.1037/a0030309

Jorm, A., & Ryan, S. M. (2014). Cross-national and historical differences in subjective well-being. *International Journal of Epidemiology*, *43*(2), 330–340. doi:10.1093/ije/dyt188

Julien, C. (2015). Bourdieu, social capital and online interaction. *Sociology*, *49*(2), 356–373. doi:10.1177/0038038514535862

Kang, J. H., & Chung, D. Y. (2017). Homophily in anonymous online community: Sociodemographic versus personality traits. *Cyberpsychology, Behavior, and Social Networking.* Advance online publication. doi:10.1089/cyber.2016.0227

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, *53*, 59–68. doi:10.1016/j.bushor.2009.09.003

Kappas, A. (2017). The psychology of (cyber)emotions. In J. A. Holyst (Ed.), *Cyberemotions: Collective emotions in cyberspace* (pp. 37–52). Cham, Switzerland: Springer International.

Karriker-Jaffe, K. J., Foshee, V. A., Ennett, S. T., & Suchindran, C. (2008). The development of aggression during adolescence: Sex differences in trajectories of physical and social aggression among youth in rural areas. *The Journal of Abnormal Child Psychology*, *36*(8), 1227–1236. doi:10.1007/s10802-008-9245-5

Kauff, M., Asbrock, F., Issmer, C., Thörner, S., & Wagner, U. (2015). When immigrant groups "misbehave": The influence of perceived deviant behavior on increased threat and discriminatory intentions and the moderating role of right-wing authoritarianism. *European Journal of Social Psychology 45*, 641–652. doi:10.1002/ejsp.2116

Kawachi, I., Kennedy, B. P., & Glass, R. (1999). Social capital and self-rated health: A contextual analysis. *American Journal of Public Health*, *89*(8), 1187–1193.

Kaye, L.K., Kowert, R., & Quinn, S. (2017). The role of social identity and online social capital on psychosocial outcomes in MMO players. *Computers in Human Behavior*, *74*, 215–223. doi:10.1016/j.chb.2017.04.030

Kazharski, A. (2017). The end of 'Central Europe'? The rise of the radical right and the contestation of identities in Slovakia and the Visegrad Four. *Geopolitics*. Advance online publication. doi:10.1080/14650045.2017.1389720

Keipi, T. (2015). *Now you see me, now you don't: A study of the relationship between Internet anonymity and Finnish young people.* Doctoral dissertation in economic sociology. Turku: Faculty of Social Sciences, University of Turku. Retrieved from Annales Universitatis Turkuensis database (B 405) http://urn.fi/URN:ISBN:978-951-29-6178-8

Keipi, T., Oksanen, A., Räsänen, P., Näsi, M., Holkeri, E., & Hawdon, J. (2014). Who prefers anonymous self expression online? A survey-based study of Finns aged 15 to 30. *Information, Communication & Society*, *18*, 717–732. doi:10.1080/1369118X.2014.991342

Keipi, T., Kaakinen, M., Oksanen, A., & Räsänen, P. (2017). Social tie strength and online victimization: An analysis of young people aged 15–30 years in four nations. *Social Media + Society*, *3*(1), 1–12. doi:10.1177/2056305117690013

Keipi, T., Näsi, M. J., Oksanen, A., & Räsänen, P. (2017). *Online hate and harmful content: Cross-national perspectives*. Abingdon, England & New York, US: Routledge.

Keith, T. Z. (2015). Multiple regression and beyond: An introduction to multiple regression and structural equation modeling (2nd ed.). London, England: Routledge.

Kelly, C., & Kelly, J. E. (1994). Who gets involved in collective action? Social psychological determinants of individual participation in trade unions. *Human Relations*, *47*, 63–87. doi:10.1177/001872679404700104

Kennedy, B. P., Kawachi, I., Prothrow-Stith, D., Lochner, K., & Gupta, V. (1998). Social capital, income inequality, and firearm violent crime. *Social Science and Medicine*, *47*(1), 7–17. doi:10.1016/S0277-9536(98)00097-5

Kim, B., & Kim, Y. (2017). College students' social media use and communication network heterogeneity: Implications for social capital and subjective well-being. *Computers in Human Behavior*, *73*, 620–628. doi:10.1016/j.chb.2017.03.033

Klausen, J. (2015). Tweeting the Jihad: Social media networks of Western foreign fighters in Syria and Iraq. *Studies in Conflict & Terrorism*, *38*(1), 1–22. doi:10.1080/1057610X.2014.974948

Klement, C. (2016). Outlaw biker affiliations and criminal involvement. *European Journal of Criminology*, *13*(4), 453–472. doi:10.1177/1477370815626460

Komorowski, M., Huu, T. D., & Deligiannis, N. (2018). Twitter data analysis for studying communities of practice in the media industry. *Telematics and Informatics*, *35*(1), 195–212. doi:10.1016/j.tele.2017.11.001

Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, *140*(4), 1073–1137. doi:10.1037/a0035618

Krämer, N. C., Neubaum, G., & Eimler, S. C. (2017). A brief history of (social) cyberspace. In J. A. Holyst (Ed.), *Cyberemotions: Collective emotions in cyberspace* (pp. 11–36). Cham, Switzerland: Springer International.

Kramer, R. M. (2006). Social capital and cooperative behavior in the workplace: A social identity perspective. *Advances in Group Processes*, *23*, 1–30.

Krosch, A. R., Tyler, T. R., & Amodio, D. M. (2017). Race and recession: Effects of economic scarcity on racial discrimination. *Journal of Personality and Social Psychology*, *113*(6), 892–909. doi:10.1037/pspi0000112

Lea, M., & Spears, R. (1991). Computer-mediated communication, de-individuation and group decision-making. *International Journal of Man-Machine Studies*, *34*(2), 283–301. doi:10.1016/0020-7373(91)90045-9

Leach, C. W., van Zomeren, M., Zebel, S., Vliek, M. L. W., & Ouwerkerk, J. W. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology*, *95*(1), 144–165. doi:10.1037/0022-3514.95.1.144

Lee, M. J., & Chun, J. W. (2016). Reading others' comments and public opinion poll results on social media: Social judgment and spiral of empowerment. *Computers in Human Behavior*, *65*, 479–487. doi:10.1016/j.chb.2016.09.007

Lee, S., Chung, J. E., & Park, N. (2018). Network environments and well-being: An examination of personal network structure, social capital, and perceived social support. *Health Communication*, *33*(1), 22–31. doi:10.1080/10410236.2016.1242032

Lehdonvirta, V., & Räsänen, P. (2011). How do young people identify with online and offline peer groups? A comparison between UK, Spain and Japan.

Journal of Youth Studies, *14*(1), 91–108. doi:10.1080/13676261.2010.506530

Leist, A. K. (2013). Social media use of older adults: A mini-review. *Gerontology*, *59*(4), 378–384. doi:10.1159/000346818

Levin, B. (2002). Cyberhate: A legal and historical analysis of extremists' use of computer networks in America. *American Behavioral Scientist*, *45*(6), 958–986. doi:10.1177/0002764202045006004

Lewin, K. (1936). *Principles of topological psychology*. New York, NY: McGraw-Hill.

Liang, H., & Fu, K. (2017). Information overload, similarity, and redundancy: Unsubscribing information sources on Twitter. *Journal of Computer-Mediated Communication*, *22*, 1–17. doi:10.1111/jcc4.12178

Lin, N. (1999). Building a network theory of social capital. *Connections*, *22*(1), 28–51.

Lindström, M., & Giordano, G. N. (2016). The 2008 financial crisis: Changes in social capital and its association with psychological wellbeing in the United Kingdom: A panel study. *Social Science & Medicine*, *153*, 71–80. doi:10.1016/j.socscimed.2016.02.008

Lo, D., Surian, D., Prasetyo, P. K., Zhang, K., & Ee-Peng, L. (2013). Mining direct antagonistic communities in signed social networks. *Information Processing & Management*, *49*(4), 773–791. doi:10.1016/j.ipm.2012.12.009

Lucassen, G., & Lubbers, M. (2012). Who fears what? Explaining far-right-wing preference in Europe by distinguishing perceived cultural and economic ethnic threats. *Comparative Political Studies*, *45*(5), 547–574. doi:10.1177/0010414011427851

Lunstrum, E. (2017). Feed them to the lions: Conservation violence goes online. Geoforum, 79, 134–143. doi:10.1016/j.geoforum.2016.04.009

Marengo, D., Longobardi, C., Fabris, M. A., & Settanni, M. (2018). Highly-visual social media and internalizing symptoms in adolescence: The mediating

role of body image concerns. *Computers in Human Behavior*, *82*, 63–69. doi:10.1016/j.chb.2018.01.003

Matamoros-Fernández, A. (2017). Platformed racism: The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society*, *20*(6), 930–946. doi:10.1080/1369118X.2017.1293130

McCroskey, L. L., McCroskey, J. C., & Richmond, V. P. (2006). Analysis and improvement of the measurement of interpersonal attraction and homophily. *Communication Quarterly*, *54*(1), 1–31. doi:10.1080/01463370500270322

McCroskey, J. C., Richmond, V. P., & Daly, J. A. (1975). The development of a measure of perceived homophily in interpersonal communication. *Human Communication Research*, *1*, 323–332. doi:10.1111/j.1468-2958.1975.tb00281.x

McDoom, O. S. (2014). Antisocial capital: A profile of Rwandan genocide perpetrators' social networks. *Journal of Conflict Resolution*, *58*(5), 865–893. doi:10.1177/0022002713484282

McEwen, R., & Wellman, B. (2013). Relationships, community, and networked individuals. In R. Teigland & D. Power (Eds.), *The immersive Internet* (pp. 168–179). London, England: Palgrave-Macmillan.

McGarty, C., Lala, G., Douglas, K. (2011). Opinion-based groups: (Racist) talk and (collective) action on the Internet. In Z. Birchmeier, B. Deitz-Uhler, G. Stasser, (Eds.), *Strategic uses of social technology: An interactive perspective of social psychology* (pp. 145–171). Cambridge, England: Cambridge University Press.

McLoughlin, C., Patel, K. D., O'Callaghan, T., & Reeves, S. (2018). The use of virtual communities of practice to improve interprofessional collaboration and education: Findings from an integrated review. *Journal*

of Interprofessional Care *32*(2), 136–142. doi:10.1080/13561820.2017.1377692

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415–444. doi:10.3410/f.725356294.793504070

Mesch, G. S. (2012). Minority status and the use of computer-mediated communication a test of the social diversification hypothesis. *Communication Research*, *39*(3), 317–337. doi:10.1177/0093650211398865

Minkkinen, J., Oksanen, A., Näsi, M., Keipi, T., Kaakinen, M., Keipi, T., & Räsänen, P. (2016). Does social belonging to primary groups protect young people from the effects of pro-suicide sites? A comparative study of four countries. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, *37*(1), 31–41, doi:10.1027/0227-5910/a000356

Mood C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, *26*, 67–82. doi:10.1093/esr/jcp006.

Mummendey, A., Simon, B., Dietze, C., Grünert, M., Haeger, G., Kessler, S., Lettgen, S., & Schäferhoff, S. (1992). Categorization is not enough: Intergroup discrimination in negative outcome allocation. *Journal of Experimental Social Psychology*, *28*(2), 125–144. doi:10.1016/0022-1031(92)90035-I

Muntaner, C., Lynch, J., & Smith, G. D. (2000). Social capital and the third way in public health. *Critical Public Health*, *10*(2), 107–124. doi:10.1080/713658240

Murayama, H., Fujiwara, Y., & Kawachi, I. (2012). Social capital and health: A review of prospective multilevel studies. *Journal of Epidemiology*, *22*(3), 179–187. doi:10.2188/jea.JE20110128

Näsi, M., Räsänen, P., Hawdon, J., Holkeri, E., & Oksanen, A. (2015). Exposure to online hate material and social trust among Finnish youth. *Information, Technology, and People*, *28*(3), 607–622. doi:10.1108/ITP-09-2014-0198

Näsi, M., Räsänen, P., Kaakinen, M., Keipi, T., & Oksanen, A. (2017). Do routine activities help predict young adults' online harassment: A multi-nation study. *Criminology & Criminal Justice*, *17*(4), 418–432. doi:10.1177/1748895816679866

Nie, N. H. (2001). Sociability, interpersonal relations, and the Internet: Reconciling conflicting findings. *American Behavioral Scientist*, *45*(3), 420–435. doi:10.1177/00027640121957277

Official Statistics of Finland. (2014). *One half of Finnish residents participate in social network services*. Helsinki, Finland: Author. Retrieved from http://www.stat.fi/til/sutivi/2014/sutivi_2014_2014-11-06_tie_001_en.html

Official Statistics of Finland. (2016). *Finnish residents use the Internet more and more often*. Helsinki, Finland: Author. Retrieved from http://www.stat.fi/til/sutivi/2016/sutivi_2016_2016-12-09_tie_001_en.html

Official Statistics of Finland. (2017). *Mobile phone is becoming increasingly popular for using the Internet—use purposes are diversifying*. Helsinki, Finland: Author. Retrieved from http://www.stat.fi/til/sutivi/2017/13/sutivi_2017_13_2017-11-22_tie_001_en.html

Oh, J.-H. (2016). Immigration and social capital in a Korean-American women's online community: Supporting acculturation, cultural pluralism, and transnationalism. *New Media & Society*, *18*(10), 2224–2241. doi:10.1177/1461444816655627

Oksanen, A., Aaltonen, M., & Rantala, K. (2015). Social determinants of debt problems in a Nordic welfare state: A Finnish register-based study. *Journal of Consumer Policy*, *38*(3), 229–246. doi:10.1007/s10603-015-9294-4

Oksanen, A., Garcia, D., Sirola, A., Näsi, M., Kaakinen, M., Keipi, T., & Räsänen, P. (2015). Pro-anorexia and anti-pro-anorexia videos on YouTube: Sentiment analysis of user responses. *Journal of Medical Internet Research*, *17*(11), e256. doi:10.2196/jmir.5007

Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M., & Räsänen, P. (2014). Exposure to online hate among young social media users. *Sociological Studies of Children & Youth*, *18*, 253–273. doi:10.1108/S1537-466120140000018021

Oksanen, A., Hawdon, J., & Räsänen, P. (2014). Glamorizing rampage online: School shooting fan communities on YouTube. *Technology in Society*, *39*, 55–67. doi:10.1016/j.techsoc.2014.08.001

Oksanen, A., & Keipi, T. (2013). Young people as victims of crime on the Internet: A population-based study in Finland. *Vulnerable Children & Youth Studies*, *8*(4), 298–309. doi:10.1080/17450128.2012.752119

O'Reilly, T. (2005. What is Web 2.0—Design patterns and business models for the next generation of software. Retrieved from http://oreilly.com/pub/a/web2/archive/what-is-web-20.html ?page=1

Organisation for Economic Co-operation and Development. (2016). Society at a glance 2016: OECD social indicators. Paris, France: Author. Retrieved from http://www.oecd-ilibrary.org/docserver/download/8116131e.pdf?expires=1517138386&id=id&accname=guest&checksum=7EE605B22487AFE1367661DCE87D0534

Organisation for Economic Co-operation and Development. (2011). Society at a Glance 2011: OECD Social Indicators. Paris, France: Author. http://www.oecd-

ilibrary.org/docserver/download/8111041e.pdf?expires=1517140609&
id=id&accname=guest&checksum=48EA100B254F038400DE0E1E3
A8869AF

Ostrom, E. (1997). Investing in capital, institutions, and incentives. In C. Clague (Ed.), *Institutions and economic development: Growth and governance in less-developed and post-socialist countries* (pp. 153–181). Baltimore, MD: Johns Hopkins University Press.

Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. London, England: Penguin.

Park, H., & Park, S. J. (2016). Communication behavior and online knowledge collaboration: Evidence from Wikipedia. *Journal of Knowledge Management*, *20*(4), 769–792. doi:10.1108/JKM-08-2015-0312

Pérez-Díaz, V. (2002). From civil war to civil society: Social capital in Spain from the 1930s to the 1990s. In R. Putnam (Ed.), *Democracies in flux: The Evolution of social capital in contemporary society* (pp. 245–288). New York, NY: Oxford University Press.

Perry, R., Drachen, A., Kearney, A., Kriglstein, S., Nacke, L. E., Sifa, R., Wallner, G., & Johnson, D. (2018). Online-only friends, real-life friends or strangers? Differential associations with passion and social capital in video game play. *Computers in Human Behavior*, *79*, 202–210. doi:10.1016/j.chb.2017.10.032

Perry, B., & Olsson, P. (2009). Cyberhate: The globalization of hate. *Information & Communications Technology Law*, *18*(2), 185–199. doi:10.1080/13600830902814984

Peterson, J., & Densley, J. (2017). Cyber violence: What do we know and where do we go from here? *Aggression and Violent Behavior*, *34*, 193–200. doi:10.1016/j.avb.2017.01.012

Pevalin, D. J. (2000). Multiple applications of the GHQ-12 in a general population sample: An investigation of long-term retest effects. *Social Psychiatry and Psychiatric Epidemiology*, *35*, 508–512. doi:10.1007/s001270050272

Pew Research Center (2015a). Social media usage: 2005–2015. Retrieved from http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/

Pew Research Center. (2015b). Teens, social media & technology overview 2015. Retrieved from http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/

Pew Research Center. (2017a). Online harassment 2017. Retrieved from http://www.pewinternet.org/2017/07/11/online-harassment-2017/

Pew Research Center. (2017b). The future of free speech, trolls, anonymity and fake news online. Retrieved from http://www.pewinternet.org/2017/03/29/the-future-of-free-speech-trolls-anonymity-and-fake-news-online/

Postmes, T., Spears, R., Sakhel, K., & De Groot, D. (2001). Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin 27*, 1242–1254. doi:10.1177/01461672012710001

Putnam, R. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton, NJ: Princeton University Press.

Putnam, R. (2000). Bowling alone: The collapse and revival of American community. New York, NY: Simon and Schuster.

Putnam, R., & Goss, K. A. (2002). Introduction. In R. Putnam (Ed.), *Democracies in flux: The evolution of social capital in contemporary society* (pp. 3–20). New York, NY: Oxford University Press.

Pyszczynski, T., Abdollahi, A., Solomon, S., Greenberg, J., Cohen, F., & Weise, D. (2006). Mortality salience, martyrdom, and military might: The Great

Satan versus the Axis of Evil. *Personality and Social Psychology Bulletin*, *32*(4), 525–537. doi:10.1177/0146167205282157

Rains, S. A., Kenski, K., Coe, K., & Harwood, J. (2017). Incivility and political identity on the Internet: Intergroup factors as predictors of incivility in discussions of news online. *Journal of Computer-Mediated Communication*, *22*, 163–178. doi:10.1111/jcc4.12191

Rickles, J. (2016). A review of propensity score analysis: Fundamentals and developments. *Journal of Educational and Behavioral Statistics*, *41*(1), 109–114. doi:10.3102/1076998615621303

Rideout, V. J., Foehr, U. G., & Roberts, D. F. (2010). Generation M²: Media in the lives of 8- to 18-year-olds. A Kaiser Family Foundation study, January 2010. *Kaiser Family Foundation*. Retrieved from https://kaiserfamilyfoundation.files.wordpress.com/2013/04/8010.pdf

Tihveräinen, T. (2016). Poliisin tietoon tullut viharikollisuus Suomessa 2015. Poliisiammattikorkeakoulun katsauksia 10/2016 [Report on suspected hate crimes reported to the police in Finland in 2015]. Retrieved from http://www.theseus.fi/bitstream/handle/10024/116581/Katsauksia_10.pdf

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883. doi:10.1037/0033-2909.133.5.859

Reicher, S. D., Spears, R., & Postmes, T. (1995). Social identity model of deindividuation phenomena. *European Review of Social Psychology*, *6*, 161–198. doi:10.1080/14792779443000049

Reicher, S.D., Spears, R., Postmes, T., & Kende, A. (2016). Disputing deindividuation: Why negative group behaviours derive from group norms, not group immersion. *Behavioral and Brain Sciences*, *39*, e161. doi:10.1017/S0140525X15001491

Riek, B.M., Mania, E.W., & Gaertner, S. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review*, *10*(4), 336–353. doi:10.1207/s15327957pspr1004_4

Rosen, D., Lafontaine, P. R., & Hendrickson, B. (2011). CouchSurfing: Belonging and trust in a globally cooperative online social network. *New Media & Society*, *13*(6), 981–998. doi:10.1177/1461444810390341

Rosenfeld, R., Messner, S. F., & Baumer, E. P. (2001). Social capital and homicide. *Social Forces*, *80*(1), 283–309. doi:10.1353/sof.2001.0086

Rotman, D., & Preece, J. (2010). The "WeTube" in YouTube—Creating an online community through video sharing. *International Journal of Web Based Communities*, *6*(3), 317–333.

Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics*, *29*(1), 159–184. doi:10.2307/2529684

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. Journal of the American Statistical Association, 95(450), 573–585. doi:10.2307/2669400

Schrock, A. R. (2016). Exploring the Relationship Between Mobile Facebook and Social Capital: What Is the "Mobile Difference" for Parents of Young Children? Social Media + Society, 2(3), 1–11. doi:10.1177/2056305116662163

Schultz, D., Jaycox, L. H., Hickman, L. J., Setodji, C., Kofner, A., Harris, R., & Barnes, D. (2013). The relationship between protective factors and outcomes for children exposed to violence. *Violence and Victims*, *28*(4), 697–714.

Sela, S., Kuflik, T., & Mesch, G. S. (2012). Changes in the discourse of online hate blogs: The effect of Barack Obama's election in 2008. *First Monday*, *17*(11).

Shen, C., & Cage, C. (2015). Exodus to the real world? Assessing the impact of offline meetups on community participation and social capital. *New Media & Society*, *17*(3), 394–414. doi.org/10.1177/1461444813504275

Simi, P., Blee, K., DeMichele, M., & Windisch, S. (2017). Addicted to hate: Identity residual among former White supremacists. *American Sociological Review 2017*, *82*(6), 1167–1187. doi:10.1177/0003122417728719

Sirola, A., Kaakinen, M., & Oksanen, A. (2018). Excessive gambling and online gambling communities. *Journal of Gambling Studies,* DOI: 10.1007/s10899-018-9772-0

Sowles, S. J., Krauss, M. J., Gebremedhn, L., & Cavazos-Rehg, P. A. (2017). "I feel like I've hit the bottom and have no idea what to do": Supportive social networking on Reddit for individuals with a desire to quit cannabis use. *Substance Abuse*, 38:4, 477–482. doi:10.1080/08897077.2017.1354956

Spears, R., Postmes, T., Lea, M., & Wolbert, A. (2002). When are net effects gross products? The power of influence and the influence of power in computer-mediated communication. *Journal of Social Issues*, *58*, 91–107. doi:10.1111/1540-4560.00250

Staub, E. (1989). The roots of evil: The psychological and cultural origins of genocide and other forms of group violence. New York, NY: Cambridge University Press.

Staub, E., & Bar-Tal, D. (2003). Genocide, mass killing and intractable conflict: Roots, evolution, prevention, and reconciliation. In D. O. Sears, L. Huddy, & R. Jervis (Eds.), *Oxford handbook of political psychology* (pp. 710–751). New York, NY: Oxford University Press.

Steffens, N. K., Haslam, S. A., Schuh, S. C., Jetten, J., & van Dick, R. (2017). A meta-analytic review of social identification and health in organizational contexts. *Personality and Social Psychology Review*, *21*(4), 303–335. doi:10.1177/1088868316656701

Stephan, W. G., & Renfro, C. L. (2002). The role of threats in intergroup relations. In D. Mackie, & E. R. Smith (Eds.), *From prejudice to intergroup emotions* (pp. 191–208). New York, NT: Psychology Press.

Stephan, W. G., Ybarra, O., & Rios Morrison, K. (2009). Intergroup threat theory. In T. Nelson (Ed.), *Handbook of prejudice* (pp. 43–59). Mahwah, NJ: Lawrence Erlbaum Associates.

Stern, C., & Ondish, P. (2018). Political aspects of shared reality. *Current Opinion in Psychology, 23*, 11–14.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1–21. doi:10.1214/09-STS313

Sun, Y., Fang, Y., & Lim, K. H. (2014). Understanding knowledge contributors' satisfaction in transactional virtual communities: A cost–benefit trade-off perspective. *Information & Management*, *51*(4), 441–450. doi:10.1016/j.im.2014.02.008

Sureka, A., Kumaraguru, P., Goyal, A., & Chhabra, S. (2010). Mining YouTube to discover extremist videos, users, and hidden communities. In P.-J. Cheng, M.-Y. Kan, W. Lam, & P. Nakov (Eds.), *Information retrieval technology, volume 6458 of lecture notes in computer science* (pp. 13–24). Berlin, Germany: Springer.

Synnott, J., Coulias, A., & Ioannou, M. (2017). Online trolling: The case of Madeleine McCann. *Computers in Human Behavior*, *71*, 70–78. doi:10.1016/j.chb.2017.01.053

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn & Bacon.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223, 96–102.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178. doi:10.1002/ejsp.2420010202

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

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel, & W. G. Austin (Eds.), *The psychology of intergroup relations* (pp. 7–24). Chicago, IL: Nelson-Hall.

Tang, W. Y., & Fox, J. (2016). Men's harassment behavior in online video games: Personality traits and game factors. *Aggressive Behavior*, *42*, 513–521. doi:10.1002/ab.21646

Teng, C.-I. (2017). Impact of avatar identification on online gamer loyalty: Perspectives of social identity and social capital theories. *International Journal of Information Management*, *37*(6), 601–610. doi:10.1016/j.ijinfomgt.2017.06.006

Thoits, P. (2011). Mechanisms linking social ties and support to physical and mental health. *Journal of Health and Social Behavior*, *52*(2), 145–161. doi:10.1177/0022146510395592

Thomas, W. E., Brown, R., Easterbrook, M. J., Vignoles, V. L., Manzi, C., D'Angelo, C., & Holt, J. J. (2017). Social identification in sports teams: The role of personal, social, and collective identity motives. *Personality and Social Psychology Bulletin*, *43*(4), 508–523. doi:10.1177/0146167216689051

Tikal, J. P., Rice, R. E., Kent, R. G., & Uchino, B. N. (2016). 100 million strong: A case study of group identification and deindividuation on Imgur.com. *New Media & Society*, *18*(11), 2485–2506. doi:10.1177/1461444815588766

Turja, T., Oksanen, A., Kaakinen, M., Sirola, A., Kaltiala-Heino, R., & Räsänen, P. (2017). Proeating disorder websites and subjective well-being: A four-country study on young people. *International Journal of Eating Disorders*, *50*(1), 50–57. doi:10.1002/eat.22589

Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behavior. In E. J. Lawler (Ed.), *Advances in group processes: Theory and research* (pp. 77–122). Greenwich, CT: JAI.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory.* Oxford, England: Blackwell.

Turner, J. C., & Oakes, P. J. (1986). The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology*, *25*, 231–252. doi:10.1111/j.2044-8309.1986.tb00732.x

Turner, J. C., & Reynolds, K. J. (2010). The story of social identity. In T. Postmes & N. R. Branscombe (Eds.), *Rediscovering social identity: Key readings* (pp. 13–32). New York, NY: Psychology Press.

Tönnies, F. (1988). *Community & society* [1887]. East Lansing: Michigan State University Press.

United Nations. (2013). *Comprehensive study on cybercrime.* Vienna, Austria: United Nations Office on Drugs and Crime. Retrieved from https://www.unodc.org/documents/organized-crime/UNODC_CCPCJ_EG.4_2013/CYBERCRIME_STUDY_210213.pdf

United Nations. (2015). International migrant stock 2015. Retrieved from http://www.un.org/en /development/desa/population/migration/data/estimates2/estimates15.shtml

United States National Intelligence Council. (2017). Global trends: Paradox of progress. January 2017. Retrieved from https://www.dni.gov/files/documents/nic/GT-Full-Report.pdf

Vahtera, P., Buckley, P. J., Aliyeva, M., Clegg, J., & Cross, A. R. (2017). Influence of social identity on negative perceptions in global virtual teams. *Journal*

of International Management, *23*, 367–381. doi:10.1016/j.intman.2017.04.002

Vakhitova, Z. I., Reynald, D. M., & Townsley, M. (2016). Toward the adaptation of routine activity and lifestyle exposure theories to account for cyber abuse victimization. *Journal of Contemporary Criminal Justice*, *32*(2), 169–188. doi:10.1177/1043986215621379

Van Alstyne, M., & Brynjolffson, E. (2005). Global village or cyber-Balkans? Modeling and measuring the integration of electronic communities. *Management Science*, *51*, 851–868.

Van Dijck, J. (2012). Facebook as a tool for producing sociality and connectivity. *Television & New Media*, *13*(2), 160–176. doi:10.1177/1527476411415291

Van Dijck, J. (2013a). "You have one identity": Performing the self on Facebook and LinkedIn. *Media, Culture & Society*, *35*(2), 199–215. doi:10.1177/0163443712468605

Van Dijck, J. (2013b). *The culture of connectivity. A critical history of social media*. New York, NY: Oxford University Press.

Vignoles, V. L. (2011). Identity motives. In S. J. Schwartz, K. Luyckx, & V. L. Vignoles (Eds.), *Handbook of identity theory and research* (pp. 403–432). New York, NY: Springer. doi:10.1007/978-1-4419-7988-9_18

Voggeser, B. J., Singh, R. K., & Göritz, A. S. (2018). Self-control in online discussions: Disinhibited online behavior as a failure to recognize social cues. *Frontiers in Psychology*, *8*, 2372. doi:10.3389/fpsyg.2017.02372

Wahlbeck, Ö. (2016). True Finns and non-true Finns: The minority rights discourse of populist politics in Finland. *Journal of Intercultural Studies*, *37*(6), 574–588. doi:10.1080/07256868.2016.1235020

Waldron, J. (2012). *The harm in the hate speech*. Cambridge, MA: Harvard University Press.

Wall, D. S., & Williams, M. L. (2013). Policing cybercrime: Networked and social media technologies and the challenges for policing. *Policing and Society*, *23*(4), 409–412. doi:10.1080/10439463.2013.780222

Walther, J. B., & Jang, J.-W. (2012). Communication Processes in Participatory Websites. *Journal of Computer-Mediated Communication 18*(1), 2–15. doi: 10.1111/j.1083-6101.2012.01592.x

Wang, H., & Shi, F. (2018). Weibo use and political participation: The mechanism explaining the positive effect of Weibo use on online political participation among college students in contemporary China. *Information, Communication & Society*, *21*(4), 516–530. doi:10.1080/1369118X.2017.1289234

Wellman, B. (1979). The community question: The intimate networks of East Yorkers. *American Journal of Sociology*, *84*(5), 1201–1231.

Wellman, B. (2001). Physical place and cyber place: The rise of personalized networking. *International Journal of Urban and Regional Research*, *25*(2), 227–252. doi:10.1111/1468-2427.00309

Wellman, B., & Gulia, M. (1999). Virtual communities as communities: Net surfers don't ride alone. In P. Kollock & M. Smith (Eds.), *Communities and Cyberspace* (pp. 167–194). New York, NY: Routledge.

Wellman, B., Haase, A. Q., Wittem, J., & Hampton, K. (2001). Does the Internet increase, decrease, or supplement social capital? Social networks, participation, and community commitment. *American Behavioral Scientist*, *45*(3), 436–455. doi:10.1177/00027640121957286

Williams, D. (2006). On and off the 'Net: Scales for social capital in an online era. *Journal of Computer-Mediated Communication*, *11*(2), 271–307. doi:10.1111/j.1083-6101.2006.00029.x

Williams, D. (2007). The impact of time online: Social capital and cyberbalkanization. *CyberPsychology & Behavior*, *10*(3), 398–406. doi:10.1089/cpb.2006.9939

Williams, M. L., & Burnap, P. (2016). Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. *British Journal of Criminology*, *56*(2), 211–238. doi:10.1093/bjc/azv059

Williams, M. L., Edwards, A., Housley, W., Burnap, P., Rana, O., Avis, N., . . . Sloan, L. (2013). Policing cyber-neighbourhoods: Tension monitoring and social media networks. *Policing and Society*, *23*(4), 461–481. doi:10.1080/10439463.2013.780225

Wong, M. A., Frank, R., & Allsup, R. (2015). The supremacy of online White supremacists—An analysis of online discussions by White supremacists. *Information & Communications Technology*, *24*(1), 41–73. doi:10.1080/13600834.2015.1011845

Wong, R. Y. M., Cheung, C. M. K., & Xiao, B. (2018). Does gender matter in cyberbullying perpetration? An empirical investigation. *Computers in Human Behavior*, *79*, 247–257. doi:10.1016/j.chb.2017.10.022

Yao, M. Z., & Zhong, Z. J. (2014). Loneliness, social contacts and Internet addiction: A cross-lagged panel study. *Computers in Human Behavior*, *30*, 164–170. doi:10.1016/j.chb.2013.08.007

Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on Twitter. *Bulletin of Science, Technology & Society*, *30*(5), 316–327. doi:10.1177/0270467610380011

Ybarra, M. L., Mitchell, K. J., Wolak, J., & Finkelhor, D. (2006). Examining characteristics and associated distress related to Internet harassment: Findings from the Second Youth Internet Safety Survey. *Pediatrics*, *118*(4), 1169–1177. doi:10.1542/peds.2006-0815

Yen, C. (2016). How to unite the power of the masses? Exploring collective stickiness intention in social network sites from the perspective of knowledge sharing. *Behaviour & Information Technology*, *35*(2): 118–133. doi:10.1080/0144929X.2015.1105297

Yip, W., Subramanian, S. V., Mitchell, A. D., Lee, D. T., Wang, J., & Kawachi, I. (2007). Does social capital enhance health and well-being? Evidence from rural China. *Social Science & Medicine*, *64*(1), 35–49. doi:10.1016/j.socscimed.2006.08.027

Zielinski, K., Nielek, R., Wierzbicki, A., & Jatowt, A. (2018). Computing controversy: Formal model and algorithms for detecting controversy on Wikipedia and in search queries. *Information Processing and Management*, *54*, 14–36.

Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., . . . Quattrociocch, W. (2017). Debunking in a world of tribes. *PLoS ONE*, *12*(7), e0181821. doi:10.1371/journal.pone.018182

APPENDIX A: ENGLISH-TRANSLATED VIGNETTES AND MANIPULATIONS USED IN THE SURVEY EXPERIMENT

Positive stance on gambling [**experience-driven**] [**fact-driven**]

[Me and many of my friends] [According to a recent report, 80% of the Finnish people] gamble. Gambling brings [me enjoyment] [enjoyment], and it [has brought significant benefits to me and my family's well-being] [brings significant benefits to the society and people's well-being]. Behind the following link, you can read more [about Finnish people's experiences] [research findings] on gambling.

Negative stance on gambling [experience-driven] [fact-driven]

[Me and many of my friends] [According to a recent report, over 120,000 Finnish people] suffer from gambling problems. Gambling causes [me problems] [problems], and it [has caused significant damage for me and my family's well-being] [causes significant damage for the society and people's well-being]. Behind the following link, you can read more [about Finnish people's experiences] [research findings] on gambling.