

A social learning theory examination of the complicity of personalization algorithms in the creation of echo chambers of online radicalization to violent extremism

Michael Wolfowicz

Abstract

Today's internet is a space that is governed by personalization algorithms which dictate what information a user will be provided with and shielded from based on their previous online activities. The search results, friend recommendations and video suggestions that all users receive are outputs of these algorithms. Some have referred to this as a "filter bubble", in which a user's future online exposures are decided for them and where certain content will be made more available, or even recommended to them based on the algorithms' perceptions of their preferences. One of the negative side-effects of the "filter bubble" is that it can lead users to be disproportionately exposed to material and associations that confirm and reinforce their previously held biases. The result is that a user becomes part of an online "echo chamber", characterized by homogeneity and where polarization of ideas can lead to the adoption of more extreme stances. These processes are especially important considerations for the study of online radicalization to extremist violence. Since the likelihood that an individual will be radicalized on the internet is a function of a network of factors, this raises the question, *are internet personalization algorithms responsible for providing a user with a greater volume, frequency and intensity of radicalizing material and associations in comparison to a situation in which internet personalization algorithms would not be influencing the user's online experience, and does this have a significant impact on the likelihood that an individual will be radicalized?*

The proposed research examines the issue from the perspective of Social Learning Theory, testing both the strength of the theory and the degree to which internet personalization, especially with respect to new social media (NSM) platforms, may be complicit in online radicalization. The proposed research will employ a unique methodology that examines for differences between user experiences under the influence of personalization compared to experiences in a space where algorithms and personalization settings are suppressed, and the impact on radical beliefs. The proposed research's findings may have usefulness in directing countering violent extremism (CVE) policies, regulations and strategies, and may assist in focussing efforts on effective prevention and intervention approaches in an area that has received little attention to date.

The proposed research will be conducted within the framework of the joint grant program "H-CSRC – HUJI Cyber Security Research Center – Cyber Law Program", bringing together the fields of criminology, radicalization research and cyber regulation.

A Social Learning Theory (SLT) approach to the filter bubble, echo chambers and online radicalization to violent extremism

The current body of research has focused on how the internet acts as a facilitator in the creation of social circles, virtual communities, and environments through the collation of likeminded individuals under one roof (Radlauer, 2007). The internet bridges geographic boundaries, provides a degree of anonymity, and creates an environment conducive to the free expression of ideas and opinions that may be less acceptable in mainstream discourse. In turn, the internet facilitates the types of associations that provide reinforcement to ideas and beliefs that may be difficult to find offline (Stevens & Neumann, 2009; Suler, 2004). While much of this is based on self-selection and the individual's active pursuit to be part of a larger social network, some believe that 'personalization' features of the internet may play a role in creating these associations automatically. This perspective holds that the algorithms that make up online personalization decide *for* the user what they will see and what will be shielded from them, creating a "filter bubble" (Pariser, 2011) that leads individuals into "echo chambers" (Von Behr et-al, 2013; Sunstein, 2007, 2009) which are characterized by their homogeneity (Wojcieszak, 2010). Both the "filter bubble" and the "echo chamber" represent important functions in the learning and adoption of deviant behaviors. With regards to the "filter bubble", it acts as an environmental factor which dictates which opportunities and immediate situations are made available to an individual and which are conducive to either conformity or deviance (Akers, 1998). Following from this, the "echo chamber" then becomes a place in which messages in support of a given behavior (or belief), versus those against it are presented and made available in a biased manner. (Ramakrishna, 2011; Saddiq, 2010; Stevens & Neumann, 2009; Sunstein, 2008; Von Behr et al., 2013; Warner, 2010; Wojcieszak, 2009, 2010). When an echo chamber is one of deviant or radical beliefs, its function becomes one in which a user's online world consists of a greater frequency of exposure to deviant messages versus normative messages, increasing the likelihood that deviant beliefs and behaviors will be adopted (Sutherland, 1947; Akers, 1998). As such, the proposed research seeks to examine the extent to which personalization algorithms as environmental factors, are complicit in exposing a user to radicalizing content and associations that they may not have otherwise been exposed to through self-selection, and to what extent this contributes to the creation of radical echo chambers where there is an increased likelihood of radicalization.

The proposed research approaches the issue within the framework of Social Learning Theory (SLT), considered to be the most empirically proven of the criminological theories, in part because its primary elements are easily converted into measurable variables of: 1) Differential associations (*priority, frequency, duration and intensity*), 2) Differential reinforcement, 3) Definitions, and 4) Imitation. SLT has been well suited for studying cyber related crime (Higgins & Makin, 2004; Higgins & Wilson, 2006) cyber enabled crime (D'ovidio et-al, 2009) and online radicalization (Pauwels & Schils, 2016; Pauwels et-al, 2014; Holt et-al 2010; Hawdon, 2012). SLT approaches radicalization to terrorism as being a learnt form of aggressive

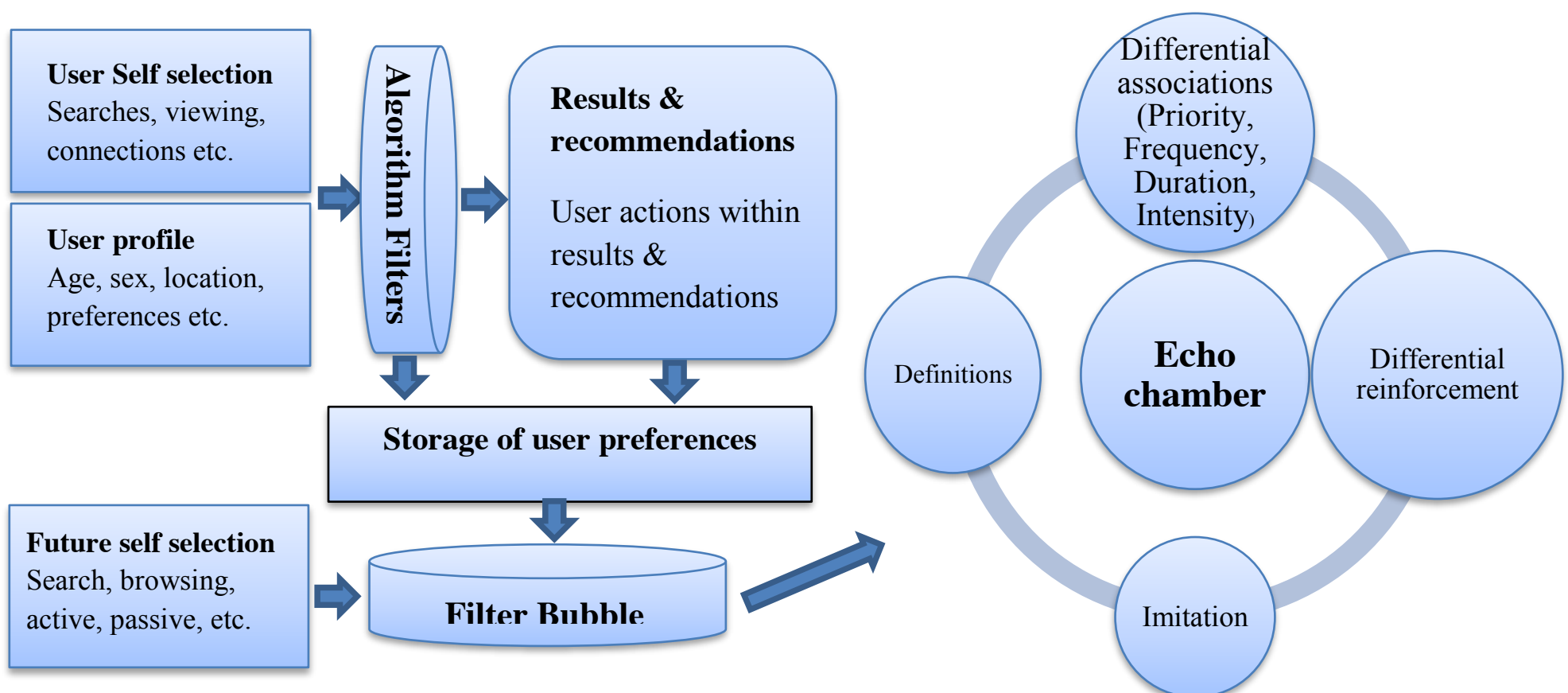
behavior that is no different than the learning of other aggressive behaviors (Oots & Wiegele, 1985; Akers & Silverman, 2004), providing also a theoretical understanding of how filter bubbles are problematic due to their role in determining availability and opportunities. The learning of deviant behaviors is dependent on the reinforcers being not only effective but also *available*. In other words, "A given behavior must be seen in the context of all other concurrently available schedules and sources of reinforcement" (Akers, 1998, p.70). Here the filter bubble plays a role in determining availability and accessibility, pre-requisites for the selection of differential associations and reinforcements.

Internet users are likely to select and view online content, or accept 'friends' when specifically recommended or first made available to them (Zhou, Khemmarat & Gao, 2010; Figueiredo, Benevenuto & Almeida, 2011). It is widely understood that humans have a natural tendency to gravitate towards similar individuals with similar beliefs and views, however, the filter bubble functions to streamline such processes (Pauwels et-al, 2014; Pauwels & Schils, 2016). When a user's online activities display some interest in a hateful ideology, personalization introduces them to content and associations of the same genre. Each time the user engages, passively or actively with such content and associations, the filter bubble further refines the information they will later see. Eventually, a user is bound to their filter bubble which "isolates us because each person's bubble is unique, and it presents us with a bias—our own—but does so without us knowing it", and "since we are largely unaware that the information we are consuming is biased, we likely believe it is not" (Hawdon, 2012:43). As such, individuals thereby enter not only virtual communities but virtual worlds in which they are more likely to be exposed to information that provides for reinforcement, and where opposing views remain hidden (Hawdon, 2012). In this way, the filter bubble may actually be pushing individuals towards radical and deviant associations and lead to the creation of echo chambers.

Echo chambers are networks of likeminded people and are prone to polarization, where their worldview reflects only their radical ideology and simultaneously hinders exposure to potential counter-messaging. Echo chambers are places in which the frequency, duration, and intensity of definitions, as well as associations themselves are essentially homogenous, with opposing definitions hardly existing (Hawdon, 2012; Wojcieszak, 2010; Geeraerts, 2012). According to Haynie (2001:1049) "when delinquent peer networks are very cohesive, network members are at heightened exposure to definitions and behavioral patterns favorable to delinquency involvement". Echo chamber are characteristically cohesive and "the principal part of the learning of criminal behavior occurs within intimate personal groups" (Sutherland, 1947, p.6). Additionally, online echo chambers' accessibility and availability provides for constant differential reinforcement and provision of definitions in favor of a deviant behavior (Miller & Morris, 2016). This aspect of online echo chambers may be referred to as "algorithmic deviancy amplification", where a user engaging with deviant material, content and associations is fed only additional material, information, and associations but at an increased frequency which validates definitions and provides for strong differential reinforcement (Wood, 2016:10-13).

Regarding radicalization to violent extremism, echo chambers may provide one of the best explanations for the role of the internet since: “no single item of extremist propaganda is guaranteed to transform people into terrorists. Rather, in most cases, online radicalization results from individuals being *immersed* in extremist content for *extended periods of time*, the amplified effects of graphic images and video, and the resulting *emotional desensitization*”(Neumann, 2013:435)¹. Echo chambers provide the type of immersion that increases the rate at which an individual can adopt the shared views (Klausen et-al, 2016). Differential reinforcement is also inherent in the echo chamber's abhorring of dissent. When opposing content or views are not simply removed (Geeraerts, 2012), they are met by a 'drowning out' that is reminiscent of mob mentality. This type of environment can lead to a desensitization to violence, creating a sort of "online disinhibition" that may increase the propensity for radicalization to result in actual violence (Davies et-al, 2016; Ducol et-al, 2016; Suler, 2004), making echo chambers prime "criminogenic environments" (Sutherland, 1947; Von Behr et-al, 2013).

While some research has examined the effects of the filter bubble in the context of left-wing and right-wing extremism in NSM (Hawdon, 2012; Regner, 2014; O'Callaghan, 2013, 2015; Bright 2016), to date there are no studies that have examined the issue with regards to radical jihadism extremism. Additionally, whilst "Algorithms for content promotion are supposed to be the main determinants of the polarization effect arising out of online social media...not much is known about the role of cognitive factors" (Bessi et-al, 2016). As such, the proposed research explores the issues within the case study of online radicalization to radical Jihadi extremism and Jewish extremism, whilst building on recent work in the areas of online radicalization, personalization and echo chambers. The SLT approach provides the framework for understanding the processes through which the filter bubble may act as an environmental factor in the creation of radical echo chambers, and how echo chambers provide the environment for radicalization.



¹ Italics added

Current knowledge

Online hate groups and bloggings are made up of dense, tightly-knit connections of likeminded people (see Chau & Xu, 2007; Burris et al., 2000), and the same can be said for NSM platforms such as Facebook, whose algorithms dictate “how your network is shaped over time” and “how you interact – with whom, when, how” (Skeggs & Yuill, 2015, 2016:391). Facebook's algorithms also “control the ‘visibility’ of friends, news, items, or ideas” (van Dijck, 2013:49). User consumption behaviors and echo chamber characteristics appear to be similar across both the Facebook and Youtube platforms, with their respective algorithms resulting in few differences. Users and echo chambers on these platforms also function similarly regardless of the topic/ideology (Bessi et-al, 2016). Additionally, while Facebook users tend to view their online associations and networks as generally sharing their views, Goel et-al (2010) found that while associations do share many views, there is significant disagreement over important issues such as politics and religion which users were unaware of. Whilst the hidden potential for a change in balance of differential associations therefore exists, groupthink is clearly present.

In a large scale Facebook commissioned research project examining emotional contagion on the platform, Kramer et-al (2013) conducted an experiment of a manipulated news feed in which 689,903 users were randomly exposed to either overly positive or overly negative messages and expressions. In analyzed the posting characteristics over a one-week period they found that users exposed to fewer negative posts, or fewer positive posts, displayed lowered levels of those respective expressions. Additionally, a “withdrawal effect” was found in which some users were less likely to post emotionally expressive messages or content in general. In another study commissioned by Facebook, Bakshy et-al (2015) found that algorithmic personalization resulted in an 8% reduction in exposure to ideologically opposing content for Liberal users and a 5% reduction for Conservative users. This was done after ranking and factoring in the makeup of a user's network. However, according to a study conducted by Nikolov et-al (2015), Facebook algorithms result in a decreased exposure to ideologically opposing news of 25% for politically conservative users and 50% for Liberal, with these independent findings contradicting those of the Facebook commissioned studies.

Research conducted regarding political fragmentation and online echo chambers has thus far found only partial evidence to support the theory regarding the complicity of internet technologies in their creation (Sunstein, 2007; Pariser, 2011). Such research has primarily been limited to Twitter, blogs, and conservative-liberal or right-wing-left wing extremism (Bright, 2016; Hargittai et-al, 2008).

The current knowledge relating to internet personalization, NSM patterns and patterns of user behaviors in the context of radicalization to violent extremism has been quite limited (except for O'Callaghan et-al, 2013, 2015; Regner, 2014; Bucher, 2012, 2013; Tateo, 2005; Bright, 2016).

Proposed research design

The proposed research will compare user experiences in surfing the internet and accessing radical material when normal personalization algorithms and settings are active in providing the results, with what a user would experience and be exposed to when personalization algorithms and settings are being suppressed. The proposed research will conduct a Randomized Controlled Trial (RCT) by recruiting 100 participants which will ideally be split into two groups: Arab N=50, Jewish N=50. The groups will be randomly allocated with normal computers N=25/group (Control group) and computers outfitted with algorithm suppressing software and settings N=25/group (Treatment group). Participants will engage in normal web browsing as well as carry out specific research tasks that will require them to conduct searches on topics of religion, politics and violence and access related content. Participants' exposure and attitudes will be checked at different intervals (such as 1 week, 1 month, 3 months etc.).

In the first analysis the groups will be compared and analyzed for differences in the ratios of exposure to extremist material and associations, against exposure to material and associations that represent opposing ideologies and counter-narratives in order to test the hypothesis:

H1: Personalization algorithms are responsible for streaming a significantly greater volume, frequency and intensity of radicalizing material to users who have engaged in some self-selection of radical material compared to what a user would have been exposed to without personalization.

The second part of the analysis will examine differences and changes in user beliefs and attitudes towards extremist ideologies between the groups in order to test the hypothesis:

H2: The increased exposure to radical material and associations created by personalization algorithms leads to a significant increase in the likelihood that a user will be radicalized.

For the first analysis a mixed methodology combining Social Network Analysis (SNA) & Qualitative Content analysis (QCA) will be employed. QCA is well suited for small-N samples and cross-comparison as it provides for "more robust causal explanations and descriptions of the multiple paths to violent extremism" and the "'multiple" roles of the Internet in the radicalization trajectories" (Ducol et-al, 2016:113). QCA was recently used to examine different pathways of domestic radicalization (Jensen & LaFree, 2016). QCA also enables differentiation in coding between radical violent material and non-violent material of the same radical genre as well as normative material (Smith 2004; Smith et-al, 2008). SNA is well suited for visualizing, analyzing and simulating social learning in online environments, especially with regards to NSM and its constant variability (Sie et-al, 2012). In this research, as different sites, pages, profiles and other associations are presented, they will be treated as 'nodes' in the model, representing a user's potential different options (O'Callaghan et-al, 2013). Radical content, links, or contacts will be coded as "exposure", since even if a user does not engage in active consumption, passive consumption has already occurred (Regner, 2014). Additionally, SNA is well suited to be used in conjunction with Krackhardt and Stern's "E-I" ratio (Hargittai et-al, 2008).

Dependent variable constructs

The dependent variable for the first analysis of the study is 'echo chamber' and is based on the degree of insularity and fragmentation of a user's online environment. This follows Bright (2016) in using Krackhardt and Stern's "E-I" ratio (Krackhardt & Stern, 1988) for determining network/individual insularity. The measurement can be constructed for both users and networks as follows:

$$E - I \text{ Ratio} = \frac{Ge - Gi}{Ge + Gi}$$

The dependent variable for the second analysis will be "Normative-Extremist beliefs and attitudes" and will be constructed as a composite measure of attitudes and beliefs towards radical violence and ideology. The measure will use a scale of 1-4 and follows the approaches of Cherney & Povey (2013), Amjad & Wood (2009) and Sharma (2016). More specifically, the variable will include measures of 1) support for radical ideology, 2) support for violence, 3) support for radical groups, 4) willingness to join a radical group and 5) willingness to engage in violence.

Independent variables

Independent variables will include all forms of self-selected and algorithm generated exposure to either radical or non-radical content or associations and will be used for both analyses:

Exposure to radical content NSM	Exposure to algorithm generated content NSM
Exposure to opposing content	Exposure to algorithm generated opposing content
Exposure to radical content TM	Exposure to algorithm generated content TM
Exposure to radical contact	Exposure to algorithm generated radical contact
Exposure to radical communications	Exposure to algorithm generated radical communications
Active/passive radical contact	Active/Passive algorithm generated radical contact
Online radical communications	Online opposing communications

Control variables

Age, sex, location, language and education will be used as control variables as these are all believed to be important inputs that internet personalization algorithms take into consideration.

Innovation and implications for cyber regulations

To date there has been little research examining the nexus between internet personalization and radicalization, especially not from the user perspective. Additionally, while there is significant evidence for the existence of radical online echo chambers (Del Vicario et-al, 2016; Quattrociocchi et-al, 2016), including left-wing and right-wing (Hawdon, 2012; Regner, 2014; O'Callaghan et-al, 2013, 2015), no research has examined the issue as it relates to radical jihadist extremism. In fact, there is little empirical evidence to support any of the competing theories of online radicalization. The proposed research thereby provides innovation in a number of important areas and seeks to fill important gaps in the body of knowledge. Primarily, the proposed research examines the extent to which internet personalization provides users with radical content, recommends radical associations, and shields users from opposing content and associations. The proposed research therefore also seeks to examine the extent to which internet personalization is complicit in the formation of radical echo chambers. Additionally, the proposed research will examine from a user perspective, what is experienced in radical echo chambers and how radicalization may occur in such places. Lastly, the proposed research contributes to the literature by its study design, with only one known RCT having been conducted in the area of radicalization (Amjad & Wood, 2009).

The proposed research is also highly relevant to cyber regulation and policies. While government interventions are overwhelmingly focused on negative measures such as content removal, personalization issues may represent an important and unattended to risk. If the proposed research finds evidence of personalization contributing to radicalization processes, then it could be argued that IT companies be targeted to re-write algorithms in order to combat these processes, such as accessibility etc. Currently, governments are focusing on efforts to bring IT companies under increased regulation and supervision. Primarily this relates to efforts to force IT companies to take negative measures such as content removal and the closing of extremist profiles and pages. However, such measures have been referred to as 'the least desirable' and least effective of countering violent extremism (CVE) approaches (Neumann, 2012, 2013). Meanwhile, it appears that positive measures such as exposure to counter-messaging and counter-narratives are more effective approaches to prevention and interventions of online radicalization (Vidino, 2013). As noted in this proposal one of the issues being examined is the way in which the internet limits and prevents exposure to such efforts. As such, altering NSM algorithms could increase users' serendipitous exposure to such content whilst also making the availability of extremist content more difficult. It is therefore important to first examine the factors that contribute to the formation of radical echo chambers and how they impact radicalization processes.

References

- Akers, R. (2008). Social Learning Theory. In A. Walsh, & C. Hemmens, *Introduction to Criminology: A Text/Reader* (pp. 163-171). Thousand Oaks: Sage.
- Akers, R. L. (1998). *Social Learning and Social Structure: A General Theory of Crime and Deviance*. Boston: Northeastern University Press.
- Akers, R. L., & Silverman, A. (2004). Toward a social learning Model of Violence and Terrorism. In M. A. Zahn, H. H. Brownstein, & S. L. Jackson, *Violence: From Theory to Research* (pp. 19-35). Cincinnati: Lexis Nexis-Anderson Publishing.
- Amjad, N., & Wood, A. M. (2009). Identifying and changing the normative beliefs about aggression which lead young Muslim adults to join extremist anti-Semitic groups in Pakistan. *Aggressive behavior*, 35(6), 514-519.
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130-1132.
- Ben-David, A., & Fernández, A. M. (2016). Hate speech and covert discrimination on social media: Monitoring the Facebook pages of extreme-right political parties in Spain. *International Journal of Communication*, 10, 1167-1193.
- Bessi, A., Zollo, F., Del Vicario, M., Puliga, M., Scala, A., Caldarelli, G., . . . Quattrociocchi, W. (2016). Users Polarization on Facebook and Youtube. *arXiv preprint arXiv:1604.02705*.
- Bright, J. (2016). Explaining the emergence of echo chambers on social media: the role of ideology and extremism. *arXiv:1609.05003v1*.
- Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New media & society*, 14(7), 1164-1180.
- Bucher, T. (2013). Objects of intense feeling: The case of the Twitter API. *Computational Culture*, 3.
- Burris, V., Smith, E., & Strahm, A. (2000). White Supremacist Networks on the Internet. *Sociological Focus*, 33(2), 215-235.
- Castro, J. C. (2014). Conclusion: The code we learn with. In V. Venkatesh, *Educational, Psychological, and Behavioral Considerations in Niche Online Communities* (pp. 402-410). Hershey, PA: IGI Global.
- Chau, M., & Xu, J. (2007). Mining communities and their relationships in blogs: A study of online hate groups. *International Journal of Human-Computer Studies*, 65(1), 57-70.
- Cherney, A., & Povey, J. (2013). Exploring Support for Terrorism Among Muslims. *Perspectives on terrorism*, 7(3).

- Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 1, 92-112.
- Davies, G., Neudecker, C., Ouellet, M., Bouchard, M., & Ducol, B. (Spring 2016). Toward a Framework Understanding of Online Programs for Countering Violent Extremism. *Journal for Deradicalization*, 6, 51-86.
- Del Vicario, M. B., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3), 554-559.
- Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). Echo Chambers: Emotional Contagion and Group Polarization on Facebook. *Scientific Reports*, 6(37825). doi:http://doi.org/10.1038/srep37825
- D'Ovidio, R., Mitman, T., El-Burki, I. J., & Shumar, W. (2009). Adult-Child Sex Advocacy Websites as Social Learning Environments: A Content Analysis. *International Journal of Cyber Criminology*, 3(1), 421-440.
- Ducol, B., Bouchard, M., Davies, G., Ouellet, M., & Neudecker, C. (2016). *Assessment of the state of knowledge: Connections between research on the social psychology of the Internet and violent extremism*. Waterloo: TSAS (Canadian Network for Research on Terrorism, Security & Society).
- Egebark, J., & Ekström, M. (2011). *Like What You Like or Like What Others Like? Conformity and Peer Effects on Facebook*, No. 886. . Stockholm: Research Institute of Industrial Economics.
- Ferretti, S., Mirri, S., Prandi, C., & Salomoni, P. (2016). On personalizing Web content through reinforcement learning. *Universal Access in the Information Society*, 1-16.
- Figueiredo, F., Benevenuto, F., & Almeida, J. M. (2011). The tube over time: characterizing popularity growth of youtube videos. *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 745-754). ACM.
- Fleder, D., & Hosanagar, K. (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science*, 55(5), 697-712.
- Geeraerts, S. B. (2012). Digital radicalization of youth. *Social cosmos*, 3(1), 25-32.
- Gilbert, E., Bergstrom, T., & Karahalios, K. (January 2009). Blogs are echo chambers: Blogs are echo chambers. *HICSS'09. 42nd Hawaii International Conference on IEEE* (pp. 1-10). System Sciences.
- Goel, S., Mason, W., & Watts, D. J. (2010). Real and perceived attitude agreement in social networks. *Journal of personality and social psychology*, 99(4), 611.
- Gross, E. F. (2004). Adolescent Internet use: What we expect, what teens report. *Journal of Applied Developmental Psychology*, 25(6), 633-649.

- Hargittai, E., Gallo, J., & Kane, M. (2008). Cross-ideological discussions among conservative and liberal bloggers. *Public Choice*, 134(1-2), 67-86.
- Hawdon, J. (2012). Applying differential association theory to online hate groups: a theoretical statement. *Research on Finnish Society*, 5, 39–47.
- Haynie, D. L. (2001). Delinquent peers revisited: Does network structure matter? *American journal of sociology*, 106(4), 1013-1057.
- Higgins, G. E., & Makin, D. A. (2004). Does social learning theory condition the effects of low self-control on college students' software piracy? *Journal of Economic Crime Management*, 2(2), 1-22.
- Higgins, G. E., & Wilson, A. L. (2006). Low self-control, moral beliefs, and social learning theory in university students' intentions to pirate software. *Security Journal*, 19(2), 75-92.
- Holt, T. J., Burruss, G. W., & Bossler, A. M. (2010). Social Learning and Cyber-Deviance: Examining the Importance of a Full Social Learning Model in the Virtual World. *Journal of Crime and Justice*, 33(2), 31-61.
- Jensen, M., & LaFree, G. (2016). *Empirical Assessment of Domestic Radicalization (EADR), "Final Report to the National Institute of Justice, Office of Justice Programs*. College Park: START.
- Klausen, J., Campion, S., Needle, N., Nguyen, G., & Libretti, R. (2016). Toward a Behavioral Model of "Homegrown" Radicalization Trajectories. *Studies in Conflict & Terrorism*, 39(1), 67-83.
- Krackhardt, D., & Stern, R. N. (1988). Informal networks and organizational crises: An experimental simulation. *Social psychology quarterly*, 51(2), 123-140.
- Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788-8790.
- Livingstone, S., Haddon, L., Görzig, A., & Ólafsson, K. (2010). *Risks and safety for children on the internet: The UK report*. London: LSE: EU Kids Online.
- Livingstone, S., Haddon, L., Görzig, A., & Ólafsson, K. (2011). *Risks and safety on the internet: The perspective of European children. Full Findings*. London: LSE: EU Kids Online.
- Miller, B., & Morris, R. G. (2016). Virtual peer effects in social learning theory. *Crime & Delinquency*, 62(12), 1543–1569.
- Nagulendra, S., & Vassileva, J. (2014). Understanding and controlling the filter bubble through interactive visualization: a user study. *Proceedings of the 25th ACM conference on Hypertext and social media* (pp. 107-115). ACM.
- Neumann, P. (2012). *Countering Online Radicalization in America*. Washington D.C.: Bipartisan Policy Center.

- Neumann, P. R. (2013). Options and strategies for countering online radicalization in the United States. *Studies in Conflict and Terrorism*, 36(6), 431-459.
- Nikolov, D., Oliveira, D., Flammini, A., & Menczer, F. (2015). Measuring online social bubbles. *PeerJ Computer Science*, 1(e38).
- O'Callaghan, D. G. (2013). *The Extreme Right Filter Bubble*. arXiv:1308.6149v1.
- O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2015). Down the (White) Rabbit Hole The Extreme Right and Online Recommender Systems. *Social Science Computer Review*, 33(4), 459-478.
- O'Hara, K., & Stevens, D. (2015). Echo chambers and online radicalism: Assessing the Internet's complicity in violent extremism. *Policy & Internet*, 7(4), 401-422.
- Oots, K. L., & Wiegele, T. C. (1985). Terrorist and victim: Psychiatric and physiological approaches from a social science perspective. *Terrorism*, 8(1), 1-32.
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. London: Penguin UK.
- Pauwels, L., & Schils, N. (2016). Differential Online Exposure to Extremist Content and Political Violence: Testing the Relative Strength of Social Learning and Competing Perspectives,. *Terrorism and Political Violence*, 28(1), 1-29.
- Pauwels, L., Brion, F., Schils, N., Laffineur, J., Verhage, A., De Ruyver, B., & Easton, M. (2014). *Explaining and understanding the role of exposure to new social media on violent extremism: an integrative quantitative and qualitative approach*. Gent: Academia Press.
- Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). Echo chambers on facebook. *Harvard Law School Discussion Paper No. 877*.
- Radlauer, D. (2007). Virtual Communities as Pathways to Extremism. In B. Ganor, K. von Knop, & C. A. Duarte, *Hypermedia Seduction for Terrorist Recruiting* (pp. 67-75). Amsterdam: IOS Press.
- Ramakrishna, K. (2010). Self-Radicalisation and the Awlaki Connection. *RSIS Commentary*(75).
- Regnér, L. (2014). The YouTube-Born Terrorist. *Journal Exit-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur*, 2, 139-189.
- Resnick, P., Garrett, R. K., Kriplean, T., Munson, S. A., & Stroud, N. J. (2013). Bursting your (filter) bubble: strategies for promoting diverse exposure. *Proceedings of the 2013 conference on Computer supported cooperative work companion* (pp. 95-100). ACM.
- Saddiq, M. A. (2010). *Whither e-jihad: evaluating the threat of internet radicalisation (RSIS Commentaries No. 083)*. Singapore: Nanyang Technological University.

- Segatto, B. (2012). The Internet and Children. A look at online risks among adolescents. *Italian Journal of Sociology of Education*, 4(3), 123-137.
- Sharma, K. (2016). *What Causes Extremist Attitudes Among Sunni and Shia Youth? Evidence from Northern India*. Washington D.C. : George Washington University Program on Extremism.
- Sie, R. L., Ullmann, T. D., Rajagopal, K., Cela, K., Bitter-Rijkema, M., & Sloep, P. B. (2012). Social network analysis for technology-enhanced learning: review and future directions. *International Journal of Technology Enhanced Learning*, 4(3-4), 172-190.
- Skeggs, B., & Yuill, S. (2015). The methodology of a multi-model project examining how Facebook infrastructures social relations. *Information, Communication & Society*, 19(10), 1356-1372.
- Skeggs, B., & Yuill, S. (2016). Capital experimentation with person/a formation: how Facebook's monetization refigures the relationship between property, personhood and protest. *Information, Communication & Society*, 19(3), 380-396.
- Smith, A. G. (2004). From words to action: Exploring the relationship between a group's value references and its tendency to engage in terrorism. *Studies in Conflict and Terrorism*, 27, 409-473.
- Smith, A. G., Suedfeld, P., Conway III, L. G., & Winter, D. G. (2008). The language of violence: distinguishing terrorist from nonterrorist groups by thematic content analysis. *Dynamic of Asymmetric Conflict*, 1(2), 142-163.
- Stevens, T., & Neumann, P. (2009). *Countering Online Radicalisation: A Strategy for Action*. London: International Centre for the Study of Radicalisation and Political Violence.
- Suler, J. (2004). The Online Disinhibition Effect. *Cyber-Psychology & Behavior*, 7(3), 321-326.
- Sun, E., Rosenn, I., Marlow, C. A., & Lento, T. M. (2009). Gesundheit! Modeling Contagion through Facebook News Feed. *Proceedings of the Third International ICWSM Conference*.
- Sunstein, C. (2007). *Republic.com 2.0*. Princeton: Princeton University Press.
- Sunstein, C. R. (2002). The Law of Group Polarization. *Journal of Political Philosophy*, 10(2), 175-195.
- Sunstein, C. R. (2009). *Going to extremes: How like minds unite and divide*. New York: Oxford University Press.
- Sutherland, E. H. (1947). *Principles of Criminology. Fourth Edition* . Philadelphia: J. B. Lippincott.
- Tandukar, U., & Vassileva, J. (2012). Ensuring relevant and serendipitous information flow in decentralized online social network. *International Conference on Artificial Intelligence: Methodology, Systems, and Applications* (pp. 79-88). Springer Berlin Heidelberg.

- Tateo, L. (2005). The Italian Extreme Right On-line Network: An Exploratory Study Using an Integrated Social Network Analysis and Content Analysis Approach. *Journal of Computer-Mediated Communication*, 10(2).
- TNS. (2015). *Arab Social Media Report*. Dubai: Arab Social Media Influencers Summit.
- Tsang, A., & Larson, K. (2016). The Echo Chamber: Strategic Voting and Homophily in Social Networks. In J. Thangarajah, k. Tuyls, C. Jonker, & S. Marsella, *Proceedings of the 15th International Conference on Autonomous agents and Multiagent Systems (AAMAS 2016)* (pp. 368-375). Singapore: International Foundation for Autonomous Agents and Multiagent Systems.
- Van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. New York: Oxford University Press.
- Vidino, L. (2012). *European Strategies Against Jihadist Radicalization*. ETH Zurich: Center for Security Studies (CSS).
- Von Behr, I., Reding, A., Edwards, C., & Gribbon, L. (2013). *Radicalisation in the digital era*. Santa Monica: RAND.
- Warner, B. (2010). Segmenting the Electorate: The Effects of Exposure to Political Extremism Online. *Communication Studies*, 61(4), 430–44.
- Weisburd, D., & Britt, C. (2007). *Statistics in criminal justice*. New York: Springer.
- Wojcieszak, M. (2009). Carrying Online Participation Offline—Mobilization by Radical Online Groups and Politically Dissimilar Offline Ties. *Journal of Communication*, 59(3), 564–86.
- Wojcieszak, M. (2010). Don't Talk to Me: Effects of Ideologically Homogeneous Online Groups and Politically Dissimilar Offline Ties on Extremism. *New Media and Society*, 12(4), 637–55.
- Wood, M. A. (2016). Antisocial media and algorithmic deviancy amplification: Analysing the id of Facebook's technological unconscious. *Theoretical Criminology*, 1-18.
- Zhou, R., Khemmarat, S., & Gao, L. (2010). The impact of YouTube recommendation system on video views. *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement* (pp. 404-410). ACM.