Journal of Experimental Criminology
https://doi.org/10.1007/s11292-021-09471-0

Examining the interactive effects of the filter bubble and the echo chamber on radicalization

Michael Wolfowicz[1] · David Weisburd[1,2] · Badi Hasisi[1]

Accepted: 10 May 2021/Published online: 03 August 2021

Abstract

Objectives Despite popular notions of "filter bubbles" and "echo chambers" contributing to radicalization, little evidence exists to support these hypotheses. However, social structure social learning theory would suggest a hereto untested interaction effect.
Methodology An RCT of new Twitter users in which participants were randomly assigned to a treatment of "filter bubble" (personalization algorithm) suppression. Ego-centric network and survey data were combined to test the effects on justification for suicide bombings.
Findings Statistically significant interaction effects were found for two proxies of the echo chamber, the E-I index and modularity. For the treatment group, higher scores on both factors decreased the likelihood for radicalization, with opposing trends in the control group.
Conclusions The echo chamber effect may be dependent on the filter bubble. More research is needed on online network structures in radicalization. While personalization algorithms can potentially be harmful, they may also be leveraged to facilitate interventions.

Keywords Social structure-social learning · Echo chambers · Filter bubbles · Radicalization · Internet

✉ Michael Wolfowicz
michael.wolfowicz@mail.huji.ac.il

1 Institute of Criminology, Faculty of Law, and the Cyber-Security Research Centre, Hebrew University of Jerusalem, Mount Scopus, Jerusalem, Israel

2 Department of Criminology, Law and Society, George Mason University, FairfaxVirginia, USA

Introduction

Despite receiving significant attention in the literature, the internet-radicalization nexus remains poorly understood and under-researched (Gill et al., 2017; Holt et al. 2015). The effects for passive exposure to and active posting of radical content have been found to be small relative to other risk factors (Wolfowicz et al., 2021b, 2021a). In turn, experimental research has found exceptionally small effects for exposure to the types of radical content encountered online (e.g. Rieger et al. 2013; Shortland et al. 2017, 2020). It is therefore important to consider other possible factors that facilitate radicalization through the internet (Richards and Wood 2020). Two such factors, namely "filter bubbles" and "echo chambers", have received increasing attention by both researchers and policy makers in recent years but remain poorly understood (Macdonald and Whittaker 2019).

"Filter bubbles" refer to the cumulative effects of personalization algorithms selectively choosing users' content, or what they will and will not see, based on their user profile and preferences (Bucher 2012; Pariser 2011). In the context of radicalization, the theory is that a user who dabbles in radical content will automatically be fed similar content and recommendations for 'friendships' with others who the algorithms identify as sharing similar interests. As a user's consumption and exposure become increasingly dominated by radical content, and their associations increasingly made up of similar ties, they may begin on a digital or algorithmic "drift" towards radicalization (Goldsmith and Brewer 2015; Richards and Wood 2020).

There is now good evidence that searching for radical content on certain platforms can lead to further recommendations for similar content (e.g. Regnér 2014; Reed et al. 2019; O'Callaghan et al. 2015). However, there is little evidence to support assumptions that such algorithmic dynamics actually increase radicalization (Macdonald and Whittaker 2019). In fact, evidence suggests that algorithmic selection's potential effect on radicalization would be small relative to the effects of self-selection (O'Hara and Stevens 2015; Reed et al. 2019).

The potentially negative consequences of self-selection lie at the heart of the "Echo chamber" hypothesis, according to which users seek out ideologically similar content and associations (self-selection), and coalesce into isolated, insular and homogenous networks. Networks characterized by such properties operate as "Echo chambers", fostering opinion extremism through mutual reinforcement and isolating the network from ideologically non-congruent individuals and ideas (Sunstein 2009).

Research on both the filter bubble and echo chamber frameworks continues to suffer from a number of issues. The first is methodological. Studies in this field have thus far been relegated primarily to the macro-level, with little if any individual-level study (Macdonald and Whittaker 2019). Second, measures of 'echo chamberness' often fail to capture the type of network structure characteristics described by the respective propositions (Bruns 2019a). This ties in to another issues, namely the conceptual ambiguity of both the echo chamber and filter bubble, which has challenged researchers with developing their own way of testing the frameworks' hypotheses (Bruns 2017; Bruns 2019a, b). However, given that current evidence is unsupportive of a direct link between echo chambers or filter bubbles and radicalization, researchers should also seek to develop and test alternative hypotheses (Bruns 2019a; Macdonald and Whittaker 2019).

In this regard, at least one group of criminologists have recently described filter bubbles and echo chambers in the context of social learning perspectives, specifically social structure social learning theory (e.g. Hawdon et al. 2019; Hawdon and Costello 2020). As we describe below, in abstracting the filter bubble and echo chamber to a social structure social learning framework, an alternative hypothesis emerges, one which predicts that it is an interactive effect between these two components of the online social structure that contributes to radicalization.

We test this hypothesis in the context of a randomized experimental study. While randomized designs are quite rare in radicalization research Jugl et al. (2020), it is generally recognized that they have particular strength in linking causes to effects (Dezember et al. 2020). Our sample consisted of 96 new Twitter users from East Jerusalem who were randomly assigned to a treatment of personalization algorithm (filter bubble) suppression. Unlike other studies' focus on whole networks Bright (2018), we examine radicalization—justification of suicide bombings—at the individual level. By linking personal network data with survey data, we are able to provide a more accurate measure of the relationship between online network structures and this attitudinal outcome (Brooks et al. 2014; Eady et al. 2019). Our study confirms earlier findings regarding the lack of direct links between algorithm suppression and radicalization, but points to the potential importance of interaction effects between algorithm suppression and the network characteristics common to the filter bubble.

"Echo chambers", "Filter bubbles" and radicalization: a criminological perspective

Understanding the ways in which the internet may be involved in radicalization is a necessary first step towards developing better approaches for leveraging the internet for intervention and prevention purposes (Wolfowicz et al. 2021b, 2020). The way in which the internet has been theorized to effect radicalization is not through the simple consumption of a few violence promoting videos. Rather, it is through its function as a facilitator of access to a wider network of individuals who may share ideas for which it is more difficult to find support offline (Neumann 2013; Keene 2011; Stevens and Neumann 2009; Suler 2004; Ducol et al. 2016; Sageman 2008).

Given the existing natural tendency to self-select ideologically similar content and associations (McPherson et al. 2001), the online connecting of these individuals can lead to the development of "echo chambers" (Sunstein 2009). These insular and homogenous networks promote a one-sided balance of messages in support of a particular belief or position (Sageman 2008; Sunstein 2009). Echo chambers have been found to be characterized by properties such as insularity, or inwardness, and are most likely to form around the extreme ends of the ideological spectrum, especially as it pertains to political issues (Bright 2018; Kaiser and Rauchfleisch 2020). These structural features are conducive to ideological homophily, attitudinal reinforcement and bias confirmation, intolerance, group-think and ultimately the possibility of radicalization (Bright 2018; Wojcieszak 2010; Ganesh and Bright 2020).

The dynamics described by the echo chamber hypothesis should be familiar to criminologists. In Sutherland's (1947) differential association theory, the adoption of deviant attitudes or behaviours is the result of an excess of definitions favourable to the attitude or behaviour over those opposed to it. Definitions are provided by differential associations—friends, family and media etc.—through normative interactions occurring in the individual's intimate social environment(s), which is part of larger social structures (Akers 1998). While Sutherland's (1947) theory acknowledged the role of social structures in social learning, it was only later re-workings of the theory that specified their roles explicitly. In its most recent iteration, Akers' (1998) social structure social learning (SSSL) defines the role of these social structural factors, namely, (1) differential social (dis)organization, (2) differential location in the social structure, (3) differential social location and (4) other theoretically derived structural factors. Differential social (dis)organization refers to factors such as population size, density and connectivity, which "distinguishes one community, region, society, or social system from another". Differential location in the social structure refers to an individual's characteristics which determine which niches of the wider social structure they are more likely to fall in. Differential social location refers to an individual network's "size, organization, and structure", as well as an individual's position in their network(s) (Akers and Jennings 2016: 237–238).

Some researchers believe that network structure characteristics may be even more important to the development of deviant attitudes than the identity of a network's members (e.g. Haynie 2001; Papachristos 2011). While criminologists have only just begun to explore the effects of online network structures and characteristics on deviant outcomes (e.g. McCuddy and Vogel 2015a, b), the role of offline network structure characteristics has been explored for over two decades. However, results have been mixed with respect to factors such as tie reciprocity (Jose et al. 2016; Turanovic and Young 2016), network density (Swartout 2013; Jose et al. 2016) and network centrality (e.g. Haynie and South 2005; Baron and Tindall 1993; Reynolds and Crea 2017).

The mixed findings can perhaps be explained by the tendency to overlook SSSL's hypothesized moderation effect between social structure and learning variables and interaction effects between social structure variables (Verrill 2008). For example, Kaczkowski et al. (2020) found that while having radical peers increases the likelihood of radicalization, for those with greater social network diversity, having radical peers has a negative effect. In another study, Bélanger et al. (2020a, b) found that obsession increased the likelihood for involvement in radical networks, and in turn radicalization, only when network density was high. In the online domain, Zhu et al. (2020) found that the effects of online political communication on engagement in radical political participation were moderated by social media network heterogeneity.

Accordingly, network structure characteristics should interact with other structural conditions that explain deviancy's (radicalization) relationship with individual and network characteristics (Akers 1998). As a space, the internet is under the structural conditions of personalization algorithms (Castro 2018; Hilbert et al. 2018). Personalization algorithms use other social structure factors to make their decisions, such as language, location, political ideology, preferences for media and news and personality. Based on these factors, personalization algorithms tailor the advertisements, items, content, pages and "friendships" a user will be recommended, and which "friends" will appear on their feeds most often (Bucher 2012; Skeggs and Yuill 2016; Van Dijck 2013). This creates a feedback loop, and for those who have dabbled in and are now being recommended deviant content and associations, a potential source of "algorithmic deviancy amplification" (Wood 2017).

There is evidence that on certain platforms, like YouTube, personalization algorithms will promote radical content to those who have searched for, or previously viewed similar content (Reed et al. 2019). For example, Musa and Bendett (2010) found that even neutral searches on "Islam" on YouTube resulted in recommendations for some radical Islamist content. Similar findings have been made for extreme right-wing content, which may be reached through searches and Twitter links to otherwise innocuous content (O'Callaghan et al. 2015). This algorithmic-generated path to radical content was demonstrated by Regnér (2014) through a simulation of left-wing and right-wing users. In a short time, the simulated profiles' recommendations were virtually free of ideologically oppositional content, and violent content quickly made it to the top of their recommendation lists. This is especially concerning given that approximately 70% of all views that are on YouTube are via playlist recommendations.

In line with the self-selection perspective, the seeking out of radical content can have a significant impact on radicalization. However, active seeking it is not the only way that an individual may come into contact with radical content or associations. It is often the case that such contact or exposure is more accidental or passive (Frissen 2021). But with personalization algorithms designed to entertain curiosities, it is questionable whether such encounters are the result of pure chance, even if unintentional on the side of the user (Brewer et al. 2018). Accounts from violent radicals have often included recollections of their first exposure to radical content on YouTube having occurred through the platform's recommendations. There are even accounts of Facebook's friend suggestions having connected ISIS supporters, and Twitter's recommendations having directed followers of one Al-Qaeda supporting account to 'follow' each other (Reed et al. 2019).

Beyond this, recent research has found evidence that YouTube's personalization creates echo chambers, characterized by the type of insularity measures described above (Kaiser and Rauchfleisch 2020). While current evidence remains primarily at the whole network (macro) level, it suggests that the filter bubble encourages homogeneity of content and ties at the individual level. In line with Granovetter's (1973) perspective on the strength of weak ties, although individuals are more likely to interact with intimate peers, weak ties can serve as an important source of expected rewards, especially social and reinforcement (Akers 1998). This potential is greater for less dense networks, which are theoretically more heterogeneous (Granovetter 1983). In highly dense networks, however, the potential influence of contrasting definitions supplied by weak ties is lessened (McGloin et al. 2014). Similarly, if strong ties (e.g. close friends) provide notably different definitions and reinforcement than the wider network of weak ties, the potential influence of the strong ties can be weakened (Rees and Pogarsky 2011). When attitudes of the strong ties are close to those of the weak ties, the potential influence of the weak ties is increased (Granovetter 1983).

The filter bubble can be seen as a factor that can potentially shape or determine the effects of weak ties in two ways. First, it can facilitate the formation of weak ties between similar individuals who lack any personal connection. Second, it serves as a conduit for direct information flow between weak ties, by giving preference to similarity over distance (Eady et al. 2019). Given that the social learning of deviance occurs “in the context of all other concurrently available schedules and sources of reinforcement” (Akers 1998:70), the filter bubble is a form of technological selection and a structural condition of the internet (Castro 2018; Hilbert et al. 2018), that shapes differential availability, and thereby homophily (Pariser 2011).

The current study

Early research on the topic of echo chambers, filter bubbles and radicalization suggested that to whatever degree the internet may be complicit in the formation of echo chambers, there was no evidence that they necessarily would be associated with an increased likelihood of radicalization O'Hara and Stevens (2015). However, with a lack of individual-level research, it is difficult to make such claims (Whittaker 2017). One way to address this gap is through emerging approaches that combine social media and survey data, which is still lacking in this area (Eady et al. 2019).

The SSSL framework has previously been suggested for the study of radicalization (Akers and Silverman 2004; Akins and Winfree Jr 2016), online radicalization in particular (Pauwels and Schils 2016) and for understanding and testing the filter bubble and echo chamber hypotheses (Hawdon et al. 2019). An alternative hypothesis is borne out of this abstraction of the SSSL framework, namely that an interaction effect exists in which the filter bubble reinforces or alters the effects of echo chamber variables in promoting or reducing the likelihood of radicalization. We test this hypothesis by employing a treatment of personalization algorithm suppression and examining interactions with network structure properties of echo chambers (Bruns 2017; Bruns 2019a, b), combining social network and survey data (Al Baghal et al. 2020). Our study therefore provides both substantive and methodological contributions in a literature that suffers from a dearth of RCTs (Wolfowicz et al. 2021b, 2021a). To date, most RCTs in this field have examined the effects of providing information or content to reduce radicalization (e.g. Amjad and Wood 2009; Bélanger et al. 2020a, b). Thus, to the best of our knowledge, this is the first study to consider the online environment in and of itself.

Participants and procedure

Sample

The sample for this study was made up of young adults from East Jerusalem, a population considered at risk for radicalization (Hasisi et al. 2020). Participants were recruited by responding to printed and digital advertisements in East Jerusalem, including the Hebrew University of Jerusalem’s Mount Scopus campus. Participation was limited to those who used Facebook to access politically driven content but whom had never used Twitter. The study commenced with 115 participants opening new Twitter accounts on a single day in January 2019, with 57 participants randomly assigned to the control group and 58 to the treatment group. All participants continued to use Twitter throughout the observation period. However, 19 participants did not respond to the survey (an attrition rate of 15%), leaving a final sample of an equal 48 in the control group and 48 in the treatment group.

As per Table 1, the groups did not differ on commonly examined risk factors for radicalization (Wolfowicz et al. 2021b, 2021a). There were also no significant differences

Table 1 Descriptive statistics of final sample demographics

Factor	Treatment (N = 48)	Control (N = 48)	Sig. for test statistic
Mean age (SD)	19.02 (2.48)	18.31 (1.43)	.98
% Female	85.1%	83.7%	.78
Education			.59
High school	14.6%	19.6%	
Preparatory	62.5%	52.2%	
University	22.9%	28.3%	
Religiosity			.52
Secular	0%	2%	
Traditional	22.2%	20.45%	
Religious	55.6%	45.45%	
Very religious	22.2%	31.8%	
Worship attendance			.75
Never	4.4%	4.4%	
On holidays	6.7%	2.2%	
Often	75.6%	82.2%	
Regularly	13.3%	11.1%	
Time online			.22
< 1 hour/day	2.1%	6.3%	
1–2 hours/day	20.8%	27.1%	
3–4 hours/day	35.4%	43.8%	
> 5 hours/day	41.7%	22.9%	
Facebook use			.50
Monthly	28.3%	21.3%	
Weekly	17.4%	10.64%	
Daily	30.4%	31.9%	
A few times/day	23.9%	36.17%	
Twitter use			.13
< 1 hour/day	43.75%	64.58%	
> 1 hour/day	45.83%	33.33%	
> 2 hours/day	8.33%	2.08%	
> 3 hours/day	2.08%	0%	

Note: All p values are for test statistics. For age, a t test was used, and χ^2 tests were used for all other factors between the groups across media-related factors, including Facebook usage. Lastly, the groups did not differ in terms of their self-reported level of Twitter usage at the end of the observation period. A joint test of significance further suggested that the randomization was successful in achieving groups with similar characteristics ($F\,(20, 75) = 1.15$, $p = .32$).

Treatment and setting

The treatment of 'algorithm suppression' sought to relax the effects of the filter bubble on Twitter. Unlike a traditional treatment in which something is administered or given to the treatment group, our treatment involved the suppression of an existing condition. This means that our control group is more similar to a 'treatment as usual' condition. To facilitate the creation of these different conditions, the control and treatment groups each received different Arabic language instructional videos created by the research team, accompanied by written instructions, on how to open their accounts. While participants were not given any specific tasks, at the time of recruitment and in the instructional videos, they were encouraged to try utilizing Twitter as their primary social media platform for the next 4 months.

In the video sent to the control group, participants were instructed to use their existing, primary email addresses to sign up to Twitter. Participants were instructed to accept all recommendations provided by Twitter upon opening their accounts, encouraging them to use these recommendations to build their initial network[1]. The intended effect was that the control group would mimic, as closely as possible, the average new Twitter user, including cross-migration of personalization from Facebook and other linked sources.

Participants in the treatment group were instructed to create new email addresses for signing up to Twitter and to skip all automatic recommendations upon signup. They were subsequently guided to deactivate all personalization settings, which would enable them to start their networks 'from scratch'. The intended effect was for treatment group members to represent new Twitter users who were unaffected by the platform's personalization, and with little or no cross-over from Facebook or other platforms.

Within the first hour after having opened their accounts, we manually inspected participants' follow lists to see whether they had 'followed' the types of celebrities, politicians and news outlets commonly recommended upon the opening of a new account. Through this inspection, we confirmed that the participants had followed the instructions they received, with all members of the control group following at least some accounts of this type, and no members of the treatment group following these types of accounts.

Data

Data was collected using a combination of web-based survey and Twitter's API, providing for the possibility to investigate the intersection between online network characteristics and cognitive outcomes (Sasaki et al. 2015; Spiro 2016). This approach helped overcome many of the limitations of self-reported media use, such as underestimation of usage and misspecification of network relationships (Stier et al. 2020), and the biases of the Twitter API that affect macro-level analyses (Bucher 2013). During

[1] The control group video was designed to encourage participants to act like new Twitter users, who will use their real email addresses and accept these recommendations. While some new Twitter users may not do this, we sought to avoid contamination with the treatment.

the course of the 4-month experimental period, social network data was collected from Twitter using the NodeXL software package. At the end of the period, the directed, ego-centric Twitter networks of each participant (1.5-degree network) were created. Subsequently, participants received randomly generated keys to complete an online survey that was linked to their network data. The survey asked participants questions pertaining to their online experience and activities as well as a range of beliefs pertaining to conflict and violence.

Variables

Radicalization

Whilst there are now a number of validated instruments to measure radicalization, we rely on a single-item measure that has been used extensively in the literature and which is considered an acceptable proxy of radicalization (Schmid 2017). Our single-item measure was derived from the well-known PEW survey, used extensively in radicalization research (e.g. McCauley 2012; Victoroff et al. 2012): “Some people think that suicide bombing and other forms of violence against civilian targets are justified in order to defend Islam from its enemies. Other people believe that, no matter what the reason, this kind of violence is never justified. Do you personally feel that this kind of violence is often justified to defend Islam, sometimes justified, rarely justified, or never justified?” As we used an Arabic version of this question, the language was changed slightly. As an ordinal variable, the mean was $M = 2.02$ ($SD = .65$). While we recognize the limitations of this measure, it was chosen in light of the complexities and sensitive nature of our research context (East Jerusalem) and design of combining network and survey data Al Baghal et al. (2020), with a view to increasing participation and response rates.

Treatment

As noted earlier, our experimental variable is the “filter bubble”, which is operationalized by personalization algorithm suppression in the treatment groups. The control group as noted earlier used Twitter with its personalization features operating as usual.

Echo chamber

Across various fields, a diverse set of approaches have been taken for measuring of ‘echo chamberness’. However, many of these approaches are ill-suited for examining individual or ego-centric networks. Based on the measures discussed by Bruns (2017, 2019b), we used two complementary measures for the ‘echo chamber’.

The first measure of ‘echo chamberness’ is *modularity*, which measures the extent to which a network represents a well-defined community structure that exists in isolation to the wider network or structure (Bruns 2017; Markgraf and Schoch 2019; Newman 2003). According to the echo chamber framework, greater modularity should be associated with an increased likelihood of radicalization. In our sample, network modularity was quite low overall ($M = .21$, $SD = .06$).

However, the mere identification of a *modular* network, while perhaps indicative of the existence of an echo chamber-like structure, does not fully capture the nature of the echo chamber. This is because, while modularity may indicate the existence of a community-like cluster, it does not measure the extent to which members of the cluster may also interact with those external to it (Bruns 2017). As such, we also employ Krackhardt and Stern's (1988) "E-I" index. The E-I index is a measure of network inwardness, which can represent insularity of cohesion. Indices range from − 1 and + 1, with indices closer to + 1 indicating more externally focussed networks, and those closer to − 1 more internally focussed networks, and E-I indices of 0 indicate an equal balance between external and internal ties. As such, E-I indices can increase or decrease due to the addition of external or internal ties, respectively (Krackhardt and Stern 1988).

According to the echo chamber framework, larger E-I indices, representing more externally focussed network, should be associated with a lower likelihood of radicalization, and more internal networks with a greater likelihood. In ego-centric networks, negative values are quite rare, as was the case in our sample ($M = .492$, $SD = .262$)[2]. At the end of the observation period, there were no statistically significant differences in E-I indices of the treatment and control groups. Additionally, there were no statistically significant differences in terms of the number of overall ties or external ties. While the number of internal ties appears to be larger for the control group, which would be expected, the difference was not statistically significant. This is because as a proportion of all ties, the groups were almost equal with respect to both internal and external ties (Table 2).

Other network structure characteristics

In line with the echo chamber hypothesis and the SSSL framework, we also sought to examine additional network structure characteristics. First, we examined *reciprocity*, a measure of the extent to which weak and strong ties cluster in an individual's network (Pedersen et al. 2019). Tie reciprocity measures the extent to which two nodes in the network reciprocate friendship (following and followed) and interact with one another. Larger scores indicate greater tie strength for a network's ego ($M = .159$, $SD = .162$). We also included *Network density*, which measures the degree to which the different alters in an individual's network are interconnected with each other. Given the focus on individual networks, we used the clustering coefficient (CC) as a measure of network density ($M = .001$, $SD = .0036$). Lastly, we examined two measures of *network centrality*, a known predictor of deviant outcomes (Jose et al. 2016). In this study, we used the Eigenvector centrality (EC) measure of *relative* position of an individual in an egocentric network ($M = .140$, $SD = .053$) and closeness centrality (CC), which measures how close the ego is to all other nodes in the network ($M = .030$, SD = .029).

[2] The E-I index is calculated as the number of external ties (*Ge*) minus the number of internal ties (*Gi*), divided by the total number of ties.

Table 2 Factors contributing to the E-I index

Factor	Treatment	Control	*t* test
E-I index	.507 (.277)	.477 (.248)	− .550 (.708)
# of ties	79.23 (55.56)	139.06 (434.43)	.947 (.346)
# of external ties	62.54 (48.10)	72.21 (143.69)	.442 (.330)
% external ties	.753 (.020)	.739 (.124)	− .550 (.708)
# of internal ties	16.69 (12.11)	66.85 (310.67)	1.12 (.135)
% internal ties	.247 (.138)	.261 (.124)	.550 (.292)

Note: Above are means and standard deviations with associated t statistics and p values

Analytic strategy

To test the hypotheses, we ran a series of ordinal logistic regression models in which we first tested the direct effects of treatment, and subsequently a series of interactions between the treatment and each of the independent variables, controlling for treatment and the stand-alone variable. For each model, we carried out a Brant test to determine whether the model met the parallel line assumptions. When the assumption was violated, either for the model or any particular variable, we used a generalized ordinal logistic regression model which relaxed the parallel lines constraints for the problematic variables (Weisburd et al. 2020). In order to reduce the effects of inter-correlation, which is common with social network variables, and also because of our focus on interactions, we standardized all variables before the regressions (Aiken and West 1991; Dearing and Hamilton 2006). Robust standard errors were used to account for heteroscedasticity and further reduce the potential effects of multi-collinearity (Marquardt 1980). Following previous research Repke and Benet-Martínez (2017), we had to consider the relatively small size of our sample, and the great variation in social network variables (Brandes et al. 2008), and as such, we used a $p < .10$ level of significance.

Given that the participants hailed from the same geographic region (East Jerusalem) and were all either local high school students, preparatory program students, and current university students (see Table 1), we could not discount the possibility of shared network ties between the participants, either directly or indirectly. To check the robustness of results against non-independence, we used the Wild-t bootstrap method (Cameron, Gelbach & Miller, 2008), clustering on education and also individuals (treated and untreated) within the clusters (MacKinnon and Webb 2018; Roodman et al. 2019).

Results

Our first model (not included in Table 3) tested the direct effects of the treatment and found it to not be statistically significant (OR = .878, SE = .354, p = .748). Accordingly, the filter bubble in itself does not significantly affect radicalization.

Table 3 Ordinal regressions for individual factors and interactions predicting justification of suicide bombings

Factor	I	II	III	IV	V	VI
Treatment	.845 (.347)	.870 (.352)	.867 (.350)	.772 (.334)	.922 (.389)	.761 (.330)
EI-Index	1.849 (.554)**					
Modularity		1.464 (.333)*				
Reciprocity			1.211 (.389)			
Density				2.180 (1.117)		
Eigenvector Centrality (EC)					1.676 (.483)*	
Closeness Centrality (CC)						2.140 (.978)*
EI index × treatment	.452 (.173)**					
Modularity × treatment		.646 (.163)*				
Reciprocity × treatment			.992 (.402)			
Density × treatment				.529 (.351)		
EC × treatment					.909 (.461)	
CC × treatment						.483 (.308)
R^2	.024	.0101	.0052	.0486	.0264	.0552
				gologit2	gologit2	gologit2

Note. **<.05, *<.10

We subsequently ran a series of models which tested each of the interactions between the treatment and the network structure characteristics individually following our alternative research hypothesis (i.e. that the impacts of algorithm suppression will operate in interaction with factors associated with the echo chamber). Across all models, the treatment variable was not statistically significant on its own (see Table 3). Since network structure characteristics were not manipulated experimentally, we do not interpret the coefficients for their direct impacts. In accordance with our hypothesis, we are primarily interested in the interaction between these elements and treatment.

The interaction between the EI index and treatment is statistically significant ($p <$ .05) as is the interaction between modularity and treatment ($p < .10$). In Figs. 1 and 2, we plot the interactions at two standard deviations above and below the mean. As our dependent variable was ordinal, there are separate lines for each of the three outcomes. As described above, the dependent variable measures the degree to which respondents agree that suicide bombings can be justified, choosing between never justifiable (outcome 1), rarely justifiable (outcome 2) and often justifiable (outcome 3).

For the control group, higher levels of the EI index are associated with a greater likelihood of outcome 3, with 36% and 51% likelihoods at 1 and 2 standard deviations above the mean, respectively. For the treatment group, the opposite is true, with larger

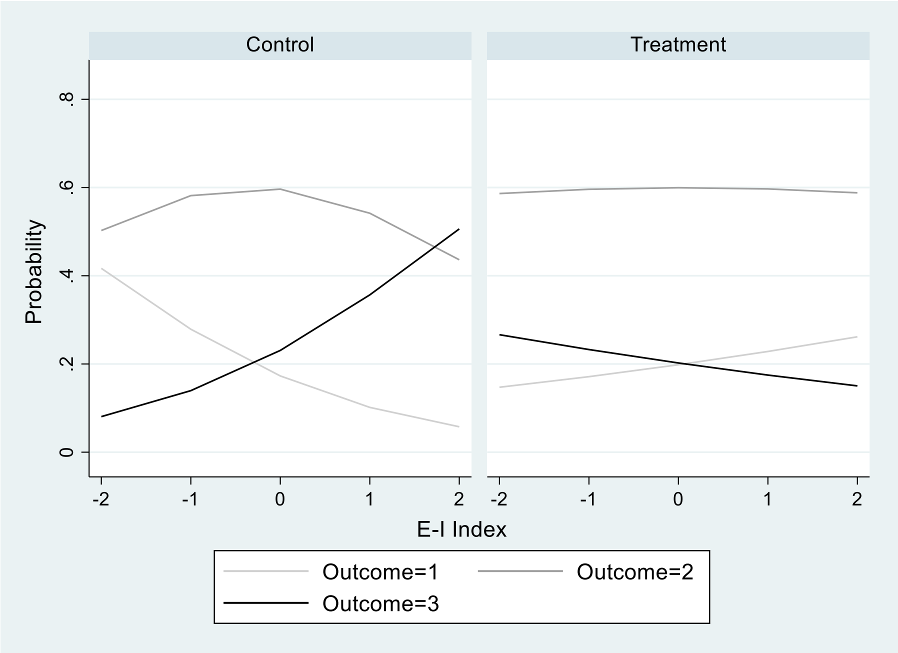

Fig. 1 Predictive margins for E-I index at 2 SD above and below mean

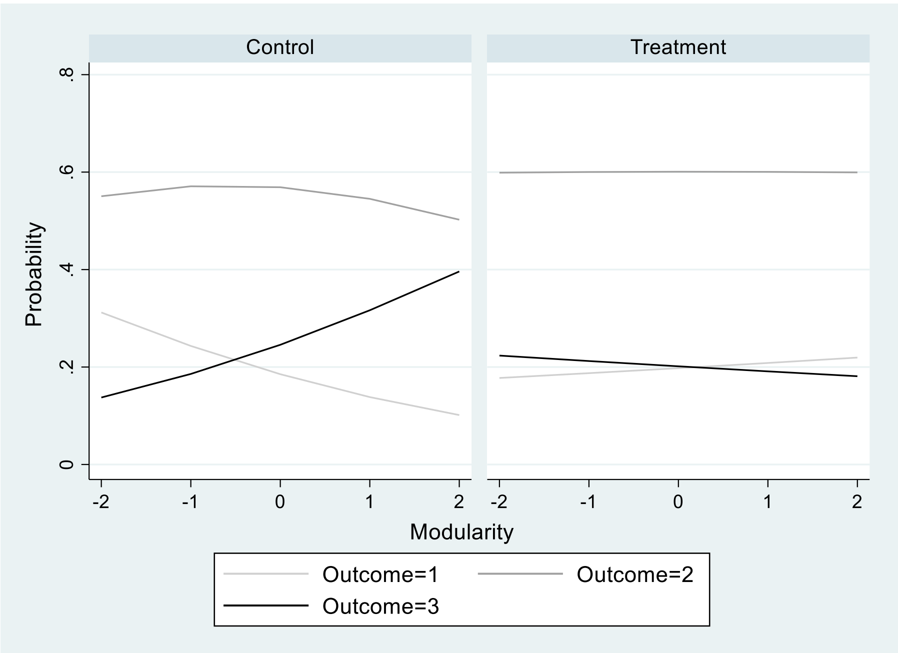

Fig. 2 Predictive margins for network modularity at 2 SD above and below mean

Table 4 Marginal predictions for E-I index and modularity across outcomes 1–3

	Outcome 1		Outcome 2		Outcome 3	
Factor	Treatment	Control	Treatment	Control	Treatment	Control
E-I index						
− 2 SD	.15†	.42**	.59***	.50***	.27*	.08†
− 1 SD	.17**	.28***	.60***	.58***	.23**	.14**
Mean (0)	.20***	.17***	.60***	.60***	.20***	.23***
+ 1 SD	.23***	.10*	.60***	.54***	.17***	.36***
+ 2 SD	.26**	.06[ns]	.59***	.44**	.15*	.51**
Modularity						
− 2 SD	.19***	.33**	.59***	.55***	.22***	.12*
− 1 SD	.20***	.25***	.59***	.58***	.21***	.17***
Mean (0)	.21***	.18***	.59***	.59***	.21***	.23***
+ 1 SD	.21***	.13*	.59***	.56***	.20***	.30***
+ 2 SD	.22**	.09†	.59***	.52***	.19***	.39***

Note: ***< .001, **< .01, *< .05, †< .10

EI indices predicting a greater likelihood of outcome 1, with 23% and 26% likelihoods at 1 and 2 standard deviations above the mean, respectively. Conversely, for the control group, there is a declining probability of outcome 1 (non-radicalization) as the E-I index increases, with 10% and 6% likelihoods at 1 and 2 standard deviations above the mean, respectively (see Table 4).

The finding for the control group, which represents a normal Twitter user exposed to the effects of personalization, demonstrates that the theorized effects of more open networks are lost in the presence of personalization. Conversely, when personalization is suppressed, these theorized protective effects are observed. Given that there are no significant differences in the E-I indices between the groups or the makeup of internal and external ties, this finding may point to the effects of personalization-driven homophily, as will be discussed below.

For outcome 2, an attitude that suicide bombings can rarely be justified, the results are somewhat different. For the control group, there is a curve-linear

Table 5 Wild-T bootstrapped p values

Factor	Education	Education/individuals
E-I index	.00	.04
E-I index × Treatment	.00	.04
E-I index+E-I index × treatment	.00	.09
Modularity	.02	.09
Modularity × treatment	.09	.08
Modularity + modularity × treatment	.02	.23

Note: Based on 9999 permutations

pattern, in which there is an increased probability for EI-indices at 1 and 2 standard deviations below the mean, and at the mean, followed by a sharp decline in the probability for increasingly positive scores. For the treatment group, however, there is an essentially flat pattern, with equal probabilities for this outcome across the range of EI indices. However, even in the presence of the filter bubble (control group), there is a similar probability for both the smallest and largest EI indices. Together with the trends for outcomes 1 and 3, the pattern for outcome 2 highlights what are essentially opposite relationships between the EI index and radicalization in the absence or presence of personalization.

In the case of the interaction between treatment and modulatory, the results are somewhat different. Here, it is the control group that demonstrates a pattern consistent with what would be expected. For the control group, modularity scores at 1 and 2 standard deviations above the mean are associated with a 30% and 39% probability of greater radicalization, respectively (outcome 3). At 1 and 2 standard deviations below the mean, there is a 25% and 33% likelihood, respectively, for lower radicalization (outcome 1). For outcome 2, the pattern is similar to the one observed for the EI-index, although the curve is less pronounced, with nearly indistinguishable probabilities across modularity scores. For the treatment group, the probabilities for outcomes 1 and 2 are nearly indistinguishable across modularity scores, although the overall trend is in the opposing direction to that of the control group. For outcome 2, like the EI-index, the pattern is entirely flat. In this case, modularity on its own may not necessarily have a positive or negative impact on radicalization. Rather, it is only when personalization is in operation that extreme modularity is associated with increased or decreased probabilities of radicalization (see Table 4).

Robustness check

We examined whether the statistically significant results were robust to non-independence for within the three clusters of education. We computed two estimates of the Wild-T bootstrap p values. First clustering on the education grouping and second clustering on individuals (treated and untreated) within the education clusters (MacKinnon and Webb 2018; Roodman et al. 2019). In both cases, the E-I index and its interaction term remained statistically significant (< .05). Taken together, the factors were jointly significant in the first analysis and marginally significant (< .10) in the second. Modularity was also found to be statistically significant in the first analysis (< .05), and its interaction term was marginally significant (< .10). Taken together, the factors were jointly significant (p = .02). However, in the second analysis, modularity and its interaction term were each only marginally significant (< .10), and taken together, the factors were jointly non-significant (p = .23). Taken together, the results demonstrate a relative degree of robustness (see Table 5).

Discussion

Despite the significant attention that both the filter bubble and echo chamber hypotheses have received in the literature, there has been little evidence linking them to radicalization. What evidence does exist has supported a failure to reject the null hypothesis concerning any direct effects. Given the ambiguity of these concepts, we developed a theoretical framework that situates the key components of the filter bubble and echo chamber within a social structure social learning theory framework. Based on this framework, as well as findings from recent research, we developed an alternative hypothesis that there is an interaction effect between the filter bubble and echo chamber related factors. Using an experimental design, we combined ego-centric network analysis with survey-based methods, enabling us to provide accurate measures of online network characteristics and cognitive radicalization. Our results provide support for our alternative hypothesis, with statistically significant interaction terms between the treatment and our proxies for the echo chamber, the E-I index and network modularity.

Our results provide for some interesting findings that highlight the need to revisit and revaluate the poorly understood concepts of the filter bubble, echo chamber and their potential role in radicalization (Macdonald and Whittaker 2019). According to its functions, in the context of network formation, the filter bubble should theoretically increase the likelihood for homophily by bringing users into contact with similar ties. For new Twitter networks, initial recommendations and suggestions are more likely to be from celebrities, organizations, "influencers" and political figures and groups. These are the types of "weak ties" that may increase the likelihood of radicalization, as Twitter's algorithms have chosen profiles that are consistent with a user's profile and preferences. However, even though ideologically similar, these types of profiles are unlikely to reciprocate with a "follow", and at the network level would be calculated as external ties. We believe that this accounts for our control group's high E-I indices being associated with greater radicalization scores.

Alternatively, we may also consider what the E-I index is actually measuring. Whilst the echo chamber literature has used it as a proxy of network inwardness, insularity and even homophily, it has elsewhere been considered a measure of tie strength and cohesion, both being a reflection of social bonds (Krackhardt 1992). Previous research has found that social bonds of different types have both protective and risk effects for radicalization, especially for highly similar ties which have been found to be one of the most salient of known risk factors (Wolfowicz et al. 2021b, 2021a). In line with this approach, we may consider that in the presence of the filter bubble, larger E-I indices reflect networks with weak but similar ties and in its absence weak, dissimilar ties. It has previously been suggested that weak, similar ties can provide reinforcement for deviant attitudes, including radicalization (Everton 2016; Kennedy and Weimann 2011). Accordingly, the potential protective effects of more open networks may be lost, and even reversed if the weak ties are ideologically similar, in line with concerns about the homophily produced by the filter bubble.

In line with these possible explanations, Barberá (2015) found that new Twitter users are more likely to follow more ideologically extreme profiles in the early stages of their network development. Over time, users increasingly diversify their networks and create ties with more moderate ties. In this context, our findings concerning the E-I index may be demonstrative of the strength of the more extreme, weak ties that are more likely to be prevalent for the control group (Granovetter 1983).

Our findings of a significant interaction effect between the treatment and network modularity also highlight the potential role of the filter bubble in altering the hypothesized effects of network structure characteristics. According to the echo chamber hypothesis, greater network modularity should increase the likelihood of radicalization. However, we identified that this does not occur to any observable degree in the treatment group. Meanwhile, the control group followed the echo chamber's theorized direction for network isolation. Previous research has found that modularity may be insufficient for distinguishing between radicalized and non-radicalized individuals (Benigni et al. 2017). As such, our findings suggest that other factors may come to explain the observed differences between the groups. For example, one of the theorized effects of modular networks is that its members may develop a perception that their attitudes are more widely shared than they really are, an effect likely reinforced through by the filter bubble (Wojcieszak 2010).

Like any study, ours is not without its limitations. We recognize that our sample represents a significant limitation, both in terms of statistical power and generalizability. This pertains to the specific context, being from East Jerusalem, as well as the demographics. In this regard, personalization on platforms like Twitter may have differential effects based on the individual's characteristics (e.g. gender, language), as well as local and national setting (Kaiser and Rauchfleisch 2020). In this regard, our sample was heavily made up of female participants. While many studies in the field of radicalization have samples consisting of large proportions of females (see the description of studies in Wolfowicz et al. 2021b). This is an often overlooked issue in the literature, especially given that males are often considered those most at risk of radicalization, or at least those who pose the greatest risk of going on to become engaged in radical violence Wolfowicz et al. (2021b). We therefore acknowledge the importance of additional research being carried out in other contexts and with samples of other socio-demographic compositions. In terms of sample size, our study highlights the difficulties of recruiting participants who are prepared to link their survey and network data. As larger samples will be needed to confirm the results of this study, future research should consider these challenges when designing an effective recruitment strategy (Al Baghal et al. 2020).

Additionally, most of the evidence concerning these network structure characteristics are derived from networks constructed from surveys, giving us little to compare to directly. Furthermore, our data did not enable us to test the full SSSL model, specifically with regard to the hypothesized moderation between social structure and learning variables. While few applications of SSSL have tested the full model (Verrill 2008), we believe that the online environment provides an opportunity to do so and we encourage such an inquiry to be taken up in future research. Lastly, whilst our single item measure of radicalization has been widely used in the literature, it may be quite limited (Schmid 2017). Evidence of this limitation may be seen in our results pertaining to outcome 2—an attitude that suicide bombings can be justifiable in rare, or certain circumstances. We would encourage future research to makes use of more general scales, such as the Sympathies for Radicalization scales (SyFor; Bhui et al. 2014) and the Activism-Radicalism-Intentions-Scales (ARIS; Moskalenko and McCauley 2009).

Despite the limitations, we believe that our findings provide both substantive and methodological contributions. Our study highlights not only the importance of online network structure as a predictor of radical attitudes but a new direction for understanding the relationship between the filter bubble, echo chamber, and radicalization. In this regard, despite the centrality of personal networks in prominent criminological and radicalization models, this is the first study that we are aware of to incorporate egocentric social network analysis—especially for online social networks—in the study of radicalization (Möller 2020). Moreover, our study demonstrates the utility of combining such data with survey data (Al Baghal et al. 2020; Eady et al. 2019). Our results also add to a growing body of literature that demonstrates the importance of criminological frameworks—both theoretical and methodological—to the study of radicalization LaFree et al. (2020).

Finally, the results of this study provide important information for policy makers. Until now, policy makers' fears that "filter bubbles" and "echo chambers" may promote polarization and fragmentation in society, and radicalization at the individual level, have not been supported by the evidence O'Hara and Stevens (2015). While we do not suggest that personalization algorithms are inherently bad, they are deserving of greater attention because they may encourage radicalization in interaction with network characteristics. In this regard, evidence suggests that personalization algorithms can also enable for the exposure to ideologically cross-cutting content from outside self-selected echo chambers (Flaxman et al. 2016). It has previously been suggested that they can also be leveraged in improving the automated detection of radicalization, and for facilitating targeted counter-messaging interventions to combat radicalization (Schmitt et al. 2018).

Conclusions

The current study is one of only a small number of randomized controlled experiments relating to online radicalization. While previous studies have focussed primarily on the role of exposure to different types of content encountered online, we demonstrated the importance of other online mechanisms in the internet-radicalization nexus. Network characteristics represent a key aspect of an individual's online experience and play a role in shaping and representing a range of cognitions. In this regard, the literature has often made reference to the ambiguous echo chamber and filter bubble metaphors, and to date, there was little evidence to support their hypotheses. Drawing on a well-established criminological theory, our study demonstrated that the filter bubble and echo chamber, as elements of the online social structure, have interactive effects that are relevant for the understanding of online radicalization.

Future studies examining the effects of personalization algorithms and the potential for algorithmic deviancy amplification should also seek to assess a range of radical ideological strains, since differential effects are likely to be found. Additionally, stronger treatments to suppress algorithms, perhaps through the use of specially designed software, may provide a better understanding of their direct and indirect effects on both network structure characteristics and radicalization. While laboratory experiments may be able to better control algorithms or implement experimental platforms, we think the approach of using a randomized field study allows for a more natural usage of social media platforms. We think that such studies more widely applied would provide strong policy guidance for reducing radicalization.

Funding This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 699824. Additional funding and assistance were received from The Federmann Cyber Security Center—Cyber Law Program at the Hebrew University of Jerusalem.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.

Akers, R. L. (1998). *Social learning and social structure: A general theory of crime and deviance*. Transaction Publishers.

Akers, R. L., & Jennings, W. G. (2016). Social learning theory. *Wiley Handbooks in Criminology and Criminal Justice*, 230–240.

Akers, R. L., & Silverman, A. (2004). *Toward a social learning model of violence and terrorism. Violence: From theory to research* (pp. 19–35).

Akins, J. K., & Winfree Jr., L. T. (2016). Social learning theory and becoming a terrorist: New challenges for a general theory. *The Handbook of the Criminology of Terrorism, 133*.

Al Baghal, T., Sloan, L., Jessop, C., Williams, M. L., & Burnap, P. (2020). Linking Twitter and survey data: The impact of survey mode and demographics on consent rates across three UK studies. *Social Science Computer Review, 38*(5), 517–532.

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: Official Journal of the International Society for Research on Aggression, 35*(6), 514–519.

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political analysis, 23*(1), 76–91.

Baron, S. W., & Tindall, D. B. (1993). Network structure and delinquent attitudes within a juvenile gang. *Social Networks, 15*(3), 255–273.

Bélanger, J. J., Nisa, C. F., Schumpe, B. M., Gurmu, T., Williams, M. J., & Putra, I. E. (2020a). Do counter-narratives reduce support for ISIS? Yes, but not for their target audience. *Frontiers in psychology, 11*.

Bélanger, J. J., Robbins, B. G., Muhammad, H., Moyano, M., Nisa, C. F., Schumpe, B. M., & Blaya-Burgo, M. (2020b). Supporting political violence: The role of ideological passion and social network. *Group Processes & Intergroup Relations, 23*(8), 1187–1203.

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.

Bhui, K., Warfa, N., & Jones, E. (2014). Is violent radicalisation associated with poverty, migration, poor self-reported health and common mental disorders? *PloS one, 9*(3), e90718.

Brandes, U., Lerner, J., Lubbers, M. J., McCarty, C., & Molina, J. L. (2008, March). Visual statistics for collections of clustered graphs. *IEEE Pacific visualization symposium*, 47–54.

Brewer, R., Cale, J., Goldsmith, A., & Holt, T. (2018). Young people, the internet, and emerging pathways into criminality: A study of Australian adolescents. *International Journal of Cyber Criminology, 12*(1), 115–132.

Bright, J. (2018). Explaining the emergence of political fragmentation on social media: The role of ideology and extremism. *Journal of Computer-Mediated Communication, 23*(1), 17–33.

Brooks, B., Hogan, B., Ellison, N., Lampe, C., & Vitak, J. (2014). Assessing structural correlates to social capital in Facebook ego networks. *Social Networks, 38*, 1–15.

Bruns, A. (2017). *Echo Chamber? What echo chamber? Reviewing the evidence*. School of Communication. Digital Media Research Centre.

Bruns, A. (2019a). It's not the technology, stupid: How the 'Echo Chamber' and 'Filter Bubble' metaphors have failed us. IAMCR Conference, Mediated Communication, Public Opinion and Society Section, Submission No.19771. Madrid, Spain.

Bruns, A. (2019b). Filter bubble. *Internet Policy Review, 8*(4).

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*.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics, 90*(3), 414–427.

Castro, J. C. L. D. (2018). Social networks as a model of algorithmic governance. *Matrizes, 12*(2), 165–191.

Dearing, E., & Hamilton, L. C. (2006). Best practices in quantitative methods for developmentalists: V. Contemporary advances and classic advice for analyzing mediating and moderating variables. *Monographs of the Society for Research in Child Development*.

Dezember, A., Stoltz, M., Marmolejo, L., Kanewske, L. C., Feingold, K. D., Wire, S., et al. (2020). The lack of experimental research in criminology—evidence from Criminology and Justice Quarterly. *Journal of Experimental Criminology*, 1–36.

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 The Canadian Network for Research on Terrorism, Security, and Society 16-05.

Eady, G., Nagler, J., Guess, A., Zilinsky, J., & Tucker, J. A. (2019). How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. *Sage Open, 9*(1), 2158244019832705.

Everton, S. F. (2016). Social networks and religious violence. *Review of Religious Research, 58*(2), 191–217.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly, 80*(S1), 298–320.

Frissen, T. (2021). Internet, the great radicalizer? Exploring relationships between seeking for online extremist materials and cognitive radicalization in young adults. *Computers in Human Behavior, 114*, 106549.

Ganesh, B., & Bright, J. (2020). Countering extremists on social media: Challenges for Strategic communication and content moderation. *Policy & Internet, 12*, 6–19.

Gill, P., Corner, E., Conway, M., Thornton, A., Bloom, M., & Horgan, J. (2017). Terrorist use of the Internet by the numbers: Quantifying behaviors, patterns, and processes. *Criminology & Public Policy, 16*(1), 99–117.

Goldsmith, A., & Brewer, R. (2015). Digital drift and the criminal interaction order. *Theoretical Criminology, 19*(*1*), 112–130.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology, 78*(6), 1360–1380.

Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory, 1*, 201–233.

Hasisi, B., Carmel, T., Weisburd, D., & Wolfowicz, M. (2020). Crime and terror: Examining criminal risk factors for terrorist recidivism. *Journal of Quantitative Criminology, 36*(3), 449–472.

Hawdon, J., & Costello, M. (2020). *Learning to hate: Explaining participation in online extremism. In Radicalization and Counter-Radicalization*. Emerald Publishing Limited.

Hawdon, J., Bernatzky, C., & Costello, M. (2019). Cyber-routines, political attitudes, and exposure to violence-advocating online extremism. *Social Forces, 98*(*1*), 329–354.

Haynie, D. L. (2001). Delinquent peers revisited: Does network structure matter? *American journal of sociology, 106*(4), 1013–1057.

Haynie, D. L., & South, S. J. (2005). Residential mobility and adolescent violence. *Social forces, 84*(*1*), 361–374.

Hilbert, M., Ahmed, S., Cho, J., Liu, B., & Luu, J. (2018). Communicating with algorithms: A transfer entropy analysis of emotions-based escapes from online echo chambers. *Communication Methods and Measures, 12*(*4*), 260–275.

Holt, T., Freilich, J. D., Chermak, S., & McCauley, C. (2015). Political radicalization on the Internet: Extremist content, government control, and the power of victim and jihad videos. *Dynamics of Asymmetric Conflict, 8*(*2*), 107–120.

Ill, P., Corner, E., Conway, M., Thornton, A., Bloom, M., & Horgan, J. (2017). Terrorist use of the Internet by the numbers: Quantifying behaviors, patterns, and processes. *Criminology & Public Policy, 16*(*1*), 99–117.

Jose, R., Hipp, J. R., Butts, C. T., Wang, C., & Lakon, C. M. (2016). Network structure, influence, selection, and adolescent delinquent behavior: Unpacking a dynamic process. *Criminal Justice and Behavior, 43*(*2*), 264–284.

Jugl, I., Lösel, F., Bender, D., & King, S. (2020). Psychosocial prevention programs against radicalization and extremism: a meta-analysis of outcome evaluations. *European journal of psychology applied to legal context, 13*(1), 37–46.

Kaczkowski, W., Swartout, K. M., Branum-Martin, L., Horgan, J. G., & Lemieux, A. F. (2020). Impact of perceived peer attitudes and social network diversity on violent extremist intentions. *Terrorism and Political Violence.*

Kaiser, J., & Rauchfleisch, A. (2020). Birds of a feather get recommended together: Algorithmic homophily in YouTube's Channel Recommendations in the United States and Germany. *Social Media+ Society, 6*(4), 2056305120969914.

Keene, S. D. (2011). Terrorism and the internet: a double-edged sword. *Journal of Money Laundering Control, 14*(4), 359.

Kennedy, J., & Weimann, G. (2011). The strength of weak terrorist ties. *Terrorism and Political Violence, 23*(2), 201–212.

Krackhardt, D. (1992). The Strength of Strong Ties: The Importance of Philos in Organizations. In N. Nohria & R. G. Eccles (Eds.), *Networks and Organizations: Structure, Form, and Action* (pp. 216–239). Harvard.
Krackhardt, D., & Stern, R. (1988). Informal networks and organizational crises: an experimental simulation. *Social psychology quarterly, 51*(2), 123–140.
LaFree, G., Weerman, F., & Bijleved, C. (2020). Editor's Introduction: Terrorism and Violent Extremism. *Journal of Quantitative Criminology, 36*, 339–405.
Macdonald, S., & Whittaker, J. (2019). Online radicalization: Contested terms and conceptual clarity. In John R. Vacca (Ed.) *Online terrorist propaganda, recruitment, and radicalization*. Boca Raton:CRC Press 33-45.
MacKinnon, J. G., & Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *The Econometrics Journal, 21*(2), 114–135.
Markgraf, M., & Schoch, M. (2019). "Quantification of Echo Chambers: A Methodological Framework Considering Multi-Party Systems". In Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden, June 8-14
Marquardt, D. W. (1980). Comment: You should standardize the predictor variables in your regression models. *Journal of the American Statistical Association, 75*(369), 87–91.
McCauley, C. (2012). Testing theories of radicalization in polls of US Muslims. *Analyses of Social Issues and Public Policy, 12*(1), 296–311.
McCuddy, T., & Vogel, M. (2015a). Beyond traditional interaction: Exploring the functional form of the exposure-offending association across online network size. *Journal of Criminal Justice, 43*(2), 89–98.
McCuddy, T., & Vogel, M. (2015b). More than just friends: Online social networks and offending. *Criminal Justice Review, 40*(2), 169–189.
McGloin, J. M., Sullivan, C. J., & Thomas, K. J. (2014). Peer influence and context: the interdependence of friendship groups, schoolmates and network density in predicting substance use. *Journal of youth and adolescence, 43*(9), 1436–1452.
McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology, 27*(1), 415–444.
Möller, V. (2020). Bedeutung und Nutzen der egozentrierten Netzwerkanalyse in der Radikalisierungsforschung: Ein Werkstattbericht zum Einsatz der Methode. *Monatsschrift für Kriminologie und Strafrechtsreform, 103*(2), 158–168.
Moskalenko, S., & McCauley, C. (2009). Measuring political mobilization: The distinction between activism and radicalism. *Terrorism and Political Violence, 21*(*2*), 239–260.
Musa, S., & Bendett, S. (2010). *Islamic Radicalization in the United States. New Trends and a Proposed Methodology for Disruption.* Washington D.C.: National Defense University, Washington DC Center for Technology and National Security Policy. Retrieved from https://apps.dtic.mil/sti/pdfs/ADA532696.pdf
Neumann, P. R. (2013). Options and strategies for countering online radicalization in the United States. *Studies in Conflict & Terrorism, 36*(6), 431–459.
Newman, M. E. (2003). Mixing patterns in networks. *Physical review E, 67*(2), 026126.
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.
Olfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2020). A field-wide systematic review and meta-analysis of putative risk and protective factors for radicalization outcomes. *Journal of Quantitative Criminology, 36*(*3*), 407–447.
Papachristos, A. V. (2011). The coming of a networked criminology. *Advances in criminological theory, 17*, 101–140.
Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. 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.
Pedersen, M. Y., Smets, S., & Ågotnes, T. (2019). Analyzing echo chambers: a logic of strong and weak ties. In *International Workshop on Logic, Rationality and Interaction*. Springer, Berlin, Heidelberg 183-198.
Reed, A., Whittaker, J., Votta, F., & Looney, S. (2019). *Radical filter bubbles: Social media personalization algorithms and extremist content*. Global Research Network on Terrorism and Technology.
Rees, C., & Pogarsky, G. (2011). One bad apple may not spoil the whole bunch: Best friends and adolescent delinquency. *Journal of Quantitative Criminology, 27*(2), 197–223.

Regnér, L. (2014). The YouTube-Born Terrorist. *Journal Exit-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur, 2*, 139–189.

Repke, L., & Benet-Martínez, V. (2017). Conceptualizing the dynamics between bicultural identification and personal social networks. *Frontiers in Psychology, 8*, 469.

Reynolds, A. D., & Crea, T. M. (2017). The integration of immigrant youth in schools and friendship networks. *Population Research and Policy Review, 36*(4), 501–529.

Richards, I., & Wood, M. (2020). Responding to online violent extremism. In C. A. Ireland, M. Lewis, A. Lopez, & J. L. Ireland (Eds.), *The Handbook of Collective Violence: Current Developments and Understanding* (pp. 1–13). Routledge.

Rieger, D., Frischlich, L., & Bente, G. (2013). Propaganda 2.0: Psychological effects of right-wing and Islamic extremist internet videos. Luxemburger:Wolters Kluwer Deutschland.

Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., & Webb, M. D. (2019). Fast and wild: Bootstrap inference in Stata using boottest. *The Stata Journal, 19*(1), 4–60.

Sageman, M. (*2008*). *Leaderless Jihad: Terror networks in the twenty-first century*. University of Pennsylvania Press.

Sasaki, Y., Kawai, D., & Kitamura, S. (2015). The anatomy of tweet overload: How number of tweets received, number of friends, and egocentric network density affect perceived information overload. *Telematics and Informatics, 32*(4), 853–861.

Schmid, A. (2017). Public opinion survey. Data to measure sympathy and support for Islamist terrorism. *International Centre for Counter-Terrorism-The Hague*.

Schmitt, J. B., Rieger, D., Rutkowski, O., & Ernst, J. (2018). Counter-messages as prevention or promotion of extremism?! the potential role of youtube: Recommendation algorithms. *Journal of communication, 68*(4), 780–808.

Shortland, N., Nader, E., Imperillo, N., Ross, K., & Dmello, J. (2017). The interaction of extremist propaganda and anger as predictors of violent responses. *Journal of interpersonal violence, 0886260517747599*.

Shortland, N., Nader, E., Thompson, L., & Palasinski, M. (2020). Is extreme in the eye of the beholder? *An experimental assessment of extremist cognitions. Journal of interpersonal violence, 0886260520958645*.

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.

Spiro, E. S. (2016). Research opportunities at the intersection of social media and survey data. *Current Opinion in Psychology, 9*, 67–71.

Stevens, T. & Neumann, P. R. (2009). Countering online radicalization: A strategy for action. The International Centre for the Study of Radicalization and Political Violence. Retrieved from http://icsr.info/wp-content/uploads/2012/10/1236768491ICSROnlineRadicalisationReport.pdf.

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. *Social Science Computer Review, 38*(5), 503–516.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & behavior, 7*(3), 321–326.

Sunstein, C. R. (2009). Going to extremes: How like minds unite and divide. Oxford University Press.

Sutherland, E. (1947). *Principles of criminology*. Lippincott.

Swartout, K. M. (2013). The company they keep: How peer networks influence male sexual aggression. *Psychology of Violence, 3*(2), 157.

Turanovic, J. J., & Young, J. T. (2016). Violent offending and victimization in adolescence: social network mechanisms and homophily. *Criminology, 54*(3), 487–519.

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

Verrill, S. W. (2008). *Social structure-social learning and delinquency: Mediation or moderation?* LFB Scholarly Publications.

Victoroff, J., Adelman, J. R., & Matthews, M. (2012). Psychological factors associated with support for suicide bombing in the Muslim diaspora. *Political Psychology, 33*(6), 791–809.

Weisburd, D., Wilson, D., Wooditch, A., & Britt, C. (2020). 'Multivariate regression with multiple category nominal or ordinal measures'. In Basic Statistics in Criminology and Criminal Justice, Springer

Whittaker, J. (2017). International Centre for Counter-Terrorism, The Hague (ICCT). Retrieved from The Sound of an Echo: https://icct.nl/publication/the-sound-of-an-echo/

Wojcieszak, M. (2010). 'Don't talk to me': Effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism. *New Media & Society, 12*(4), 637–655.

Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2021a). Cognitive and behavioral radicalization: A systematic review of the putative risk and protective factors. *Campbell Systematic Reviews*. https://doi.org/10.1002/cl2.1174.

Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2020). A field-wide systematic review and meta-analysis of putative risk and protective factors for radicalization outcomes. *Journal of Quantitative Criminology, 36*(3), 407–447.

Wolfowicz, M., Perry, S., Hasisi, B., & Weisburd, D. (2021b). Faces of radicalism: Differentiating between violent and non-violent radicals by their social media profiles. *Computers in Human Behavior, 116*, 106646.

Wood, M. A. (2017). Antisocial media and algorithmic deviancy amplification: Analysing the id of Facebook's technological unconscious. *Theoretical Criminology, 21*(2), 168–185.

Zhu, A. Y. F., Chan, A. L. S., & Chou, K. L. (2020). The pathway toward radical political participation among young people in Hong Kong: A communication mediation approach. *East Asia*, 1–18.

Michael Wolfowicz, Ph.D. is a researcher at the Institute of Criminology, Hebrew University of Jerusalem. His research focuses on the criminological study of radicalization and terrorism, with a specific focus on the internet-radicalization relationship.

David Weisburd is the Distinguished Professor of Criminology, Law and Society at George Mason University and Walter E. Meyer Professor of Law and Criminal Justice at the Hebrew University of Jerusalem. He is also the Executive Director of the Center for Evidence Based Crime Policy. He was the Founding Editor of the Journal of Experimental Criminology. He has received many international prizes and awards for his research, including the Stockholm Prize in Criminology (2010), the Sutherland (2014) and Vollmer Awards (2017) from the American Society of Criminology, and the Israel Prize (2015).

Badi Hasisi serves as a full professor and Chair of the Institute of Criminology, Faculty of Law, The Hebrew University. His work focuses on the interaction between community and the criminal justice agencies, with specific emphasis on the particular problems faced by minority groups with the criminal justice system. He also specializes in homeland security and crime-terrorism nexus. He received the 2018 best article prize of the Israeli Organization of Law and History and the Fattal Prize for Excellence in Legal Research & Criminology. Prof. Hasisi also served as the Executive Editor of the Journal of Quantitative Criminology and acts as the current chair of the Israeli Society of Criminology.