Social Media: A Source of Radicalization and a Window of Opportunity- Lessons from Israel

Michael Wolfowicz

The Institute of Criminology and The Cyber-Security Research Center

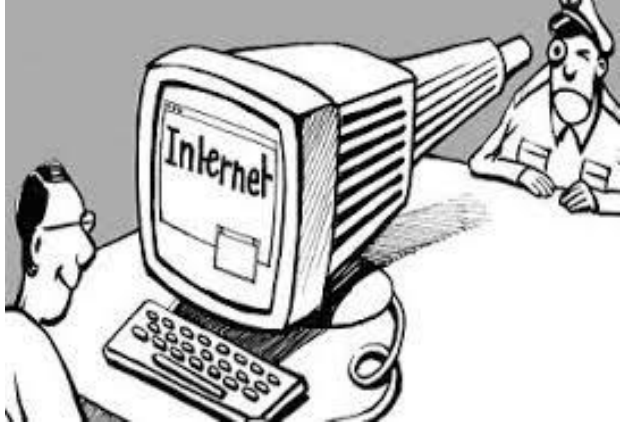
Hebrew University of Jerusalem

המכון לקרימינולוגיה
הפקולטה למשפטים
האוניברסיטה העברית בירושלים

מרכז המחקר להגנת הסייבר
CYBER SECURITY RESEARCH CENTER

PROTON
Modelling the processes leading
to organised crime and terrorist networks

Two sides to the social media coin



Radicals

- Leveraged by radical groups to incite and encourage supporters to engage in acts of radical violence, including violent protests, riots, and terrorism.
- Leveraged to create social movements that can lead to violence and unrest.
- A tool for propaganda, communications, and organization.

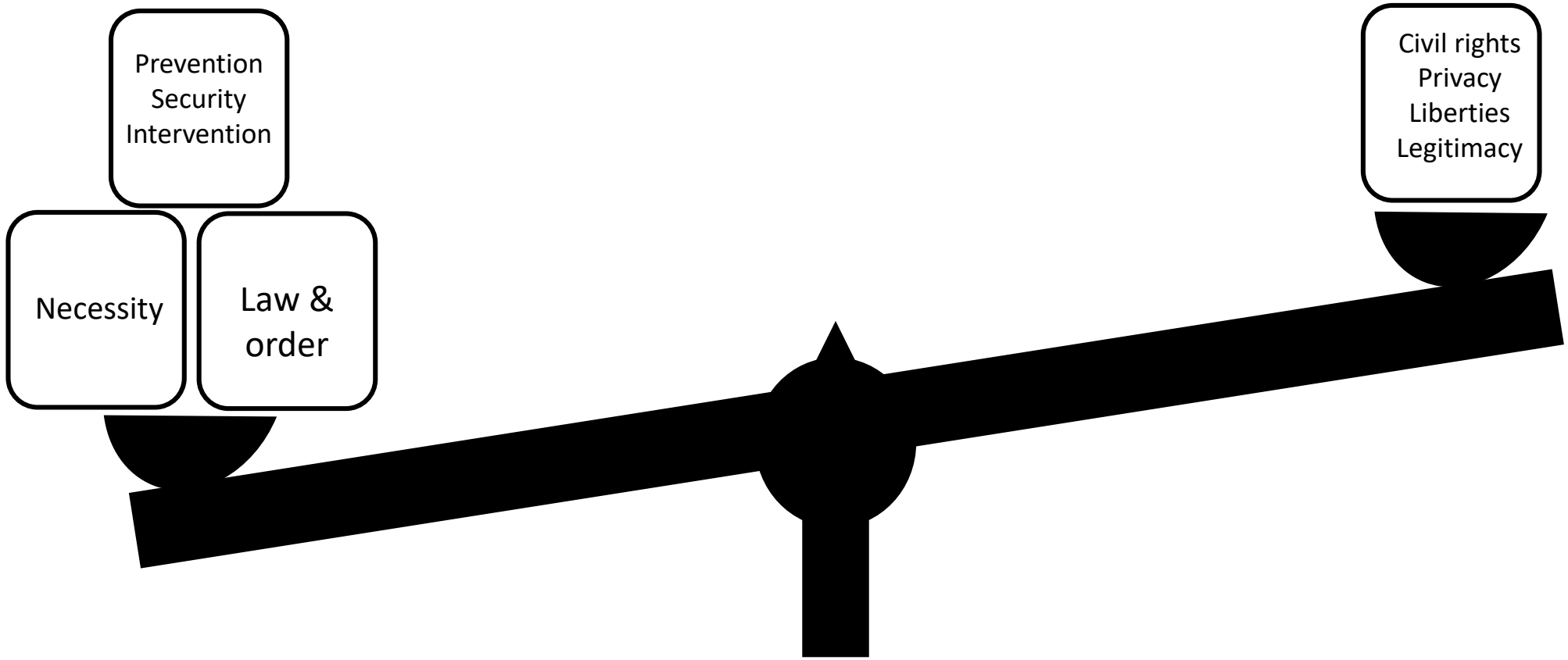
Government agencies

- Superior surveillance tool which is mostly non-invasive.
- Allows for the dissemination of counter-messaging.
- Provides access to the small window of opportunity for intervention and prevention

Balancing security needs and rights

- We have to find a balance between maintaining democratic principles and maintaining effective prevention strategies

- What is proportional?
- What is effective?



To delete or not to delete? that is the question

- Sometimes necessary, even mandated under international humanitarian law (Fidler, 2015; Shefet, 2016).
- The “least desirable” approach (Neumann, 2013).
  - Evidence to support claims and arguments, thereby generating mc support (Weirman &Alexander, 2018).
  - May cause radicals to move to more secure platforms (e.g. Telegram).
- May limit legitimate free speech
- Automated tools may flag legitimate and innocuous content, impinge on privacy (EU, 2011) and may lack proportionality (Granger & Irion, 2014).

Other considerations

- Content removal requires mass surveillance and the use of automated detection tools.
- Large number of opinion radicals but only a small proportion will act (Schmid, 2013; Hafez & Mullins, 2015).
- Keywords more likely to be used by non-violent radicals than violent radicals, simply because they outnumber them (Shortland, 2016).
- Automated detection tools built on data from radicals or synthetic data (Pelzer, 2018)
- Low accuracy rate, many false arrests (Munk, 2017; Brumnik, Podbregar, and Ivanuša, 2011).

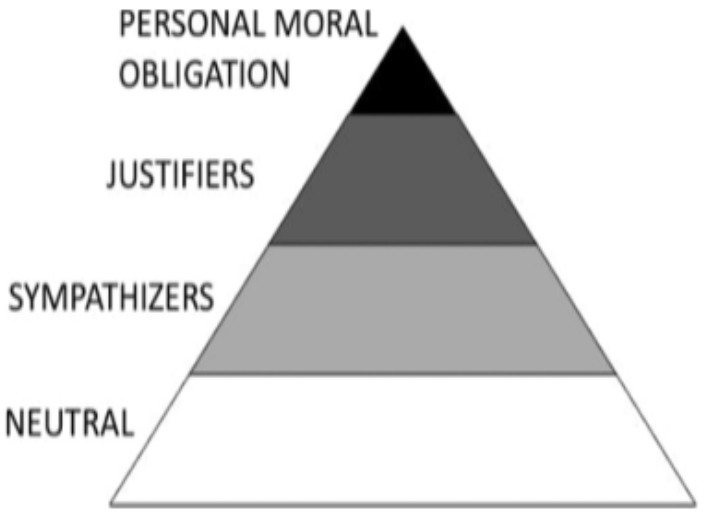


Figure 1. Opinion radicalization pyramid.

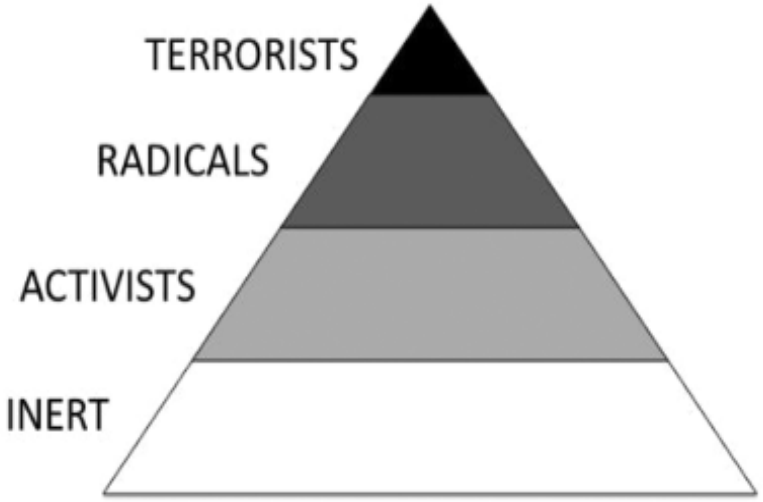
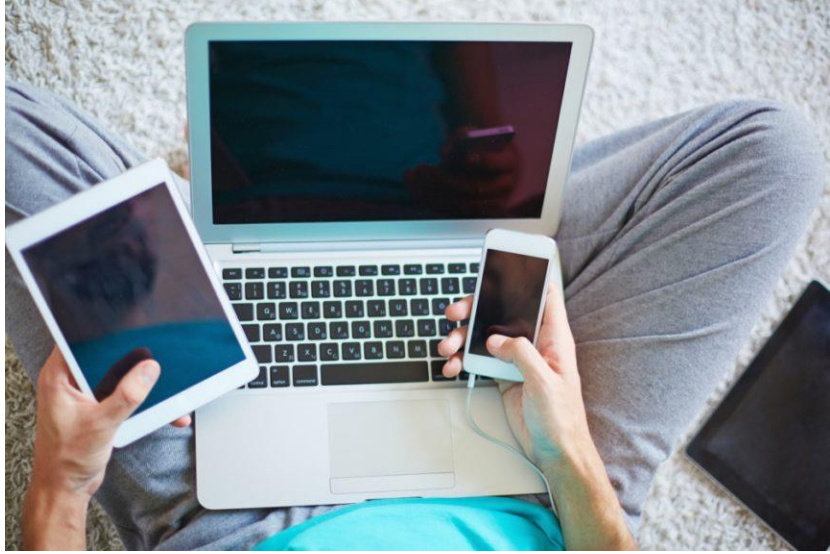


Figure 2. Action radicalization pyramid.

Can online radical content be a protective factor?

- By providing an essentially non-violent outlet to voice grievances, increased social media posting can potentially act as a protective factor against extremism (Barbera, 2014; Helmus, York and Chalk, 2013; Özdemir & Kardas, 2014, 2018).
  - Keeps them busy
  - Makes them feel like they are contributing to 'the cause'
- In Chile, using Facebook for self-expression was unrelated to engaging in offline, violent activism (Valenzuela, Arriagada and Scherman, 2012).

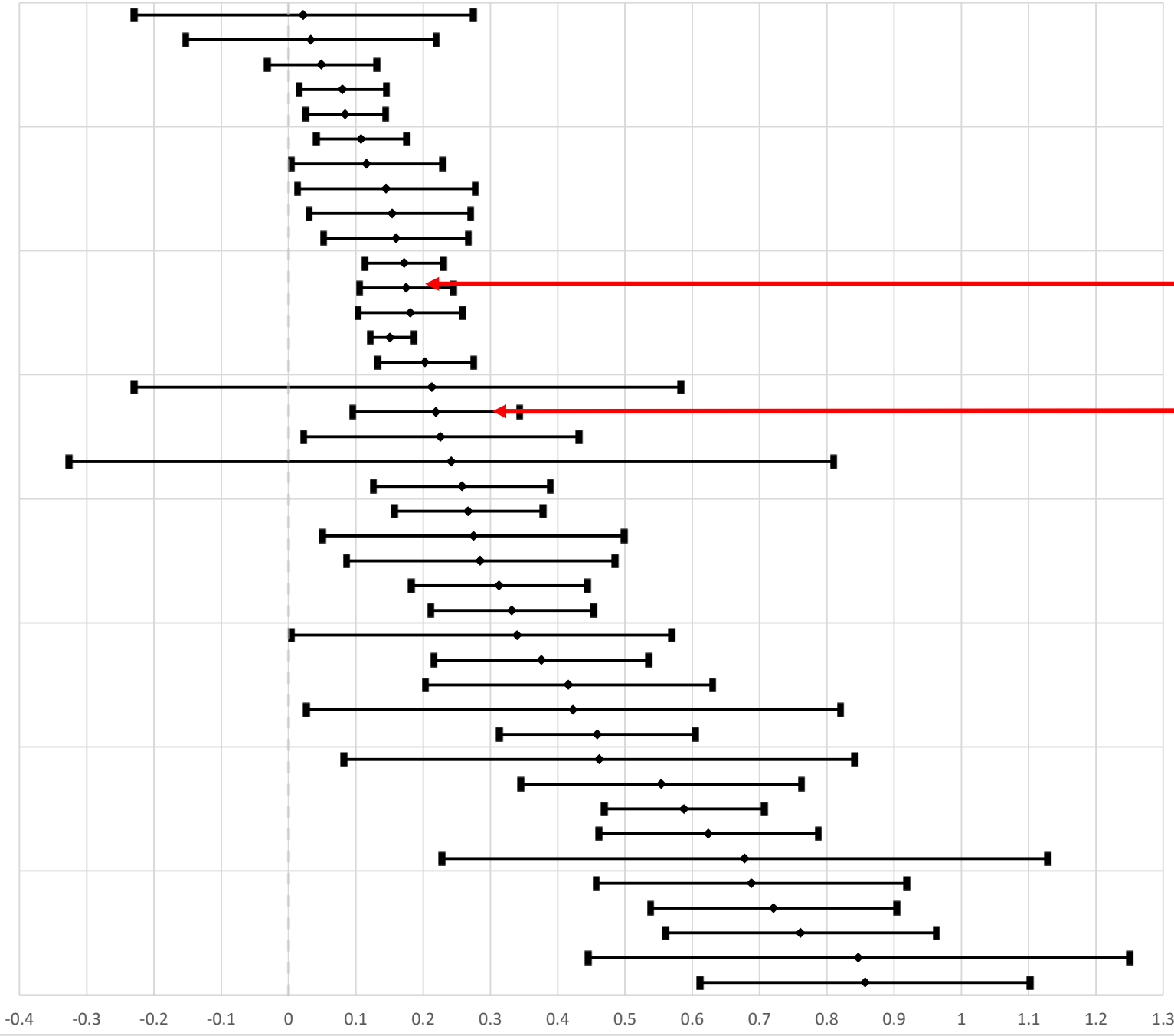
Is it as big of a problem as we think?



The internet's role in radicalization (Gill et al., 2017):

- Passive
  - Reinforcing prior beliefs
  - Seeking legitimization for action
  - Consuming propaganda (Videos, images, recordings, text based media etc.)
- Active
  - Disseminating propaganda (Videos, images, recordings, text based media etc.)
  - Communications
  - Planning
- Passive/active
  - Support groups

Risk factors for radicalization



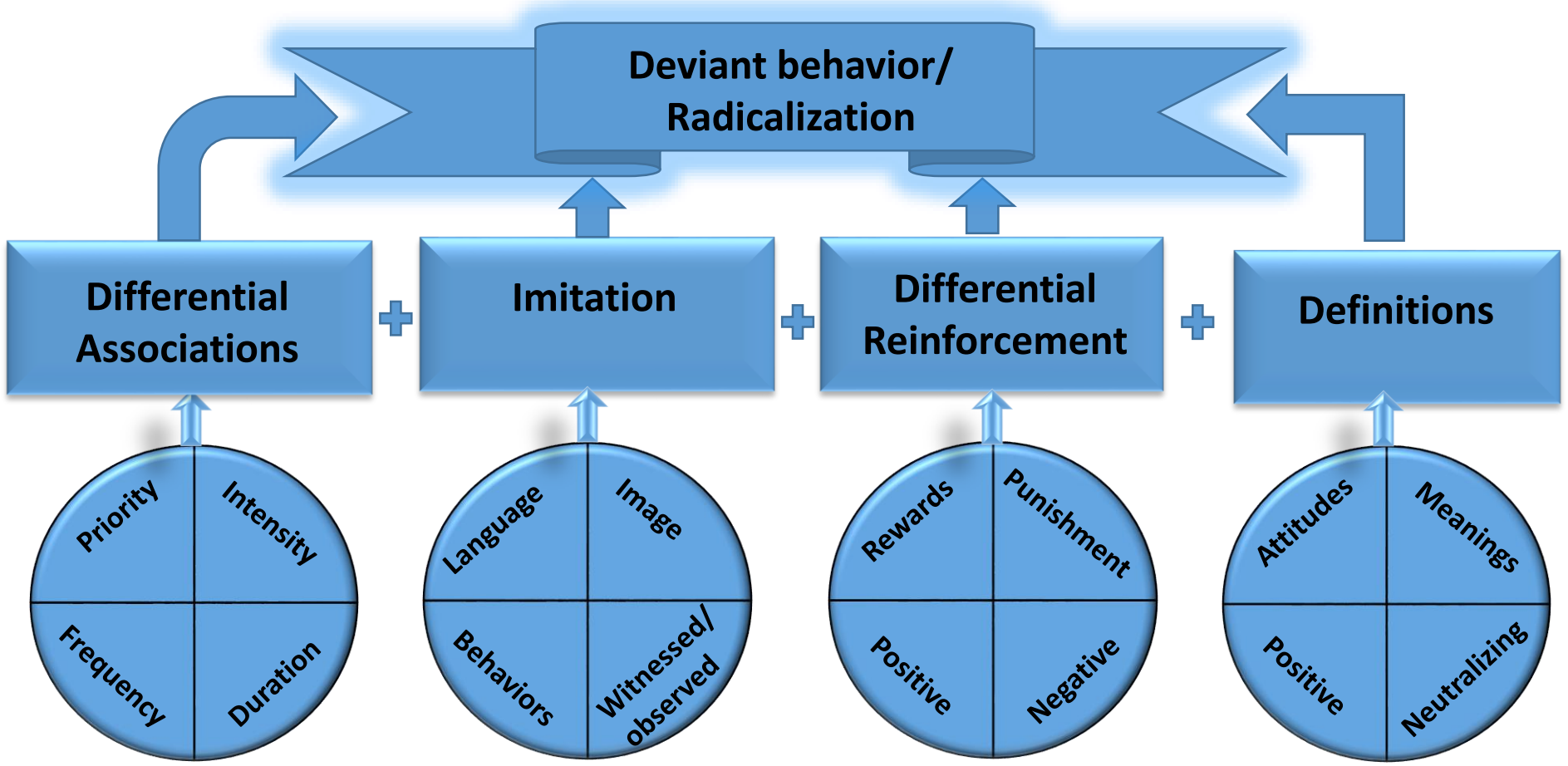
- Political efficacy (.022 NS)
- Uncertainty (.033 NS)
- Worship attendance (.049 NS)
- West Vs Islam (.08*)
- Immigrant (.084**)
- Welfare recepient (.108**)
- Unemployment (.116*)
- Religiosity (.145*)
- Discrimination (.154**)
- Political Grievance (.16**)
- Prayer frequency (.172***)
- **Violent media Exp. (.175***)** — ***Passive***
- Perceived injustice (.172***)
- Violence exposure (.186***)
- Male (.203***)
- APD/Narcissism (.213 NS)
- **NSM posting (.219**)** — ***Active***
- Aggression (.226**)
- SES (High) (.242 NS)
- Relig/Nat identity (.258***)
- Personal strains (.267***)
- Anti Democratic (.275*)
- Ind. Rel. Dep. (.285**)
- Educ. Low (.313***)
- Coll. Rel. Dep. (.332***)
- Anger/Hate (.34 NS)
- Low integration (.376***) — ***Offline peers***
- Deviant peers (.416***)
- Legal cynicism (.423*)
- Segregation (.459***)
- Moral neutralization (.462*)
- Law legitimacy (.554***)
- Low Self Control (.588**)
- Thrill/risk seeking (.624***)
- Criminal History (.678**)
- Symbolic threat (.688***)
- Police Contact (.721***)
- Realistic threat (.761***)
- Group superiority (.847***)
- Authoritarian/fundamentalism (.857***)

What is our goal?

- Identifying potentially violent radicals from the non-violent radical pool; not radicals from the general population.
- Moving beyond text-based analysis.
- Minimizing impingements on rights without compromising on security.

Social learning theory

- Deviant beliefs and behaviors are learnt as normative ones (Sutherland, 1947)
- The peer/network effect is stronger online than offline (Sunstein, 2017)



The study

- 48 violent radicals (terrorists)
  - All male
  - Aged 15-57 (M=21)
  - Carried out a combination of stabbings (49%), vehicular attacks (17%), shootings (8.5%), and other types of attacks (25.5%) (including 1 bombing)
- 96 matched non-violent radicals (two matches for each violent radical).
  - Matched by age, gender, location
  - Had to be friends with the terrorist
- Compared 100 days of Facebook activity across social learning metrics
- Only a small number displayed clear intentions of action

Theoretically driven social media level metrics

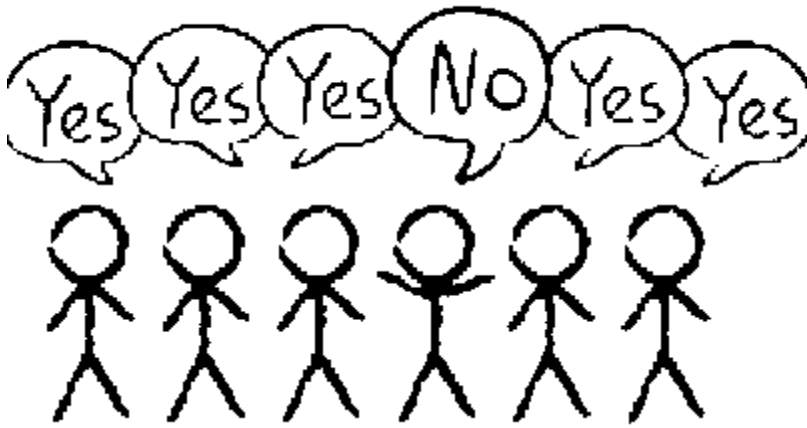
Social learning variable	Facebook metric
Differential associations (Deviant peers)	Measured as a dichotomous variable of whether the subject has posted content relating to a terror attack committed by an online network member.
Frequency	Measured as posts/day Measured as fluctuations in posting activity: non-activity
Duration	Measured as the time on Facebook prior to attack
Network size	Measured as the number of friends
Imitation	Measured as the proportion of posting types: Text post, image post, video post, shared post
Definitions	Measured as the ratio between radical and non-radical posts
Differential reinforcement	Measure of likes/post received Measure of comments/post received Measure of shares/post received

Results

Variable	Actions (N=48)	Beliefs (N=96)	T	U (Standardized)
Differential associations with terrorists	0.542 (SD=0.504)	0.219 (SD=0.416)	3.837***	3.880***
Network size (Computed)	478.104 (SD=214.673)	528.083 (SD=270.561)	-1.116	.199
Posts/day (Frequency)	0.555 (SD=0.795)	0.469 (SD=0.442)	0.696	-1.344
Duration	38.688 (SD=20.886)	34.365 (SD=17.685)	1.300	1.134
Definitions (radical post ratio)	0.696 (SD=0.397)	0.578 (SD=0.377)	1.738 †	1.804†
Differential reinforcement				
Likes/post	45.001 (SD=47.136)	44.037 (SD=36.296)	0.136	-.687
Comments/post	7.538 (SD=6.813)	9.110 (SD=9.167)	-1.051	-.161
Shares/post	0.469 (SD=0.729)	0.156 (SD=0.326)	2.834**	3.383***
Imitation (post type)				
Text posts (%)	17.938 (SD=23.089)	31.271 (SD=22.089)	-3.363**	-3.907***
Shared posts (%)	32.792 (SD=32.854)	15.271 (SD=20.637)	3.377***	2.556*
Picture posts (%)	45.083 (SD=33.285)	45.577 (SD=26.517)	-0.090	-.352
Video posts (%)	4.20 (SD=.121)	8.00 (SD=.121)	-1.798†	-2.835**

***< 0.001, ** <.01, *<.05, †<.10

What does it mean?



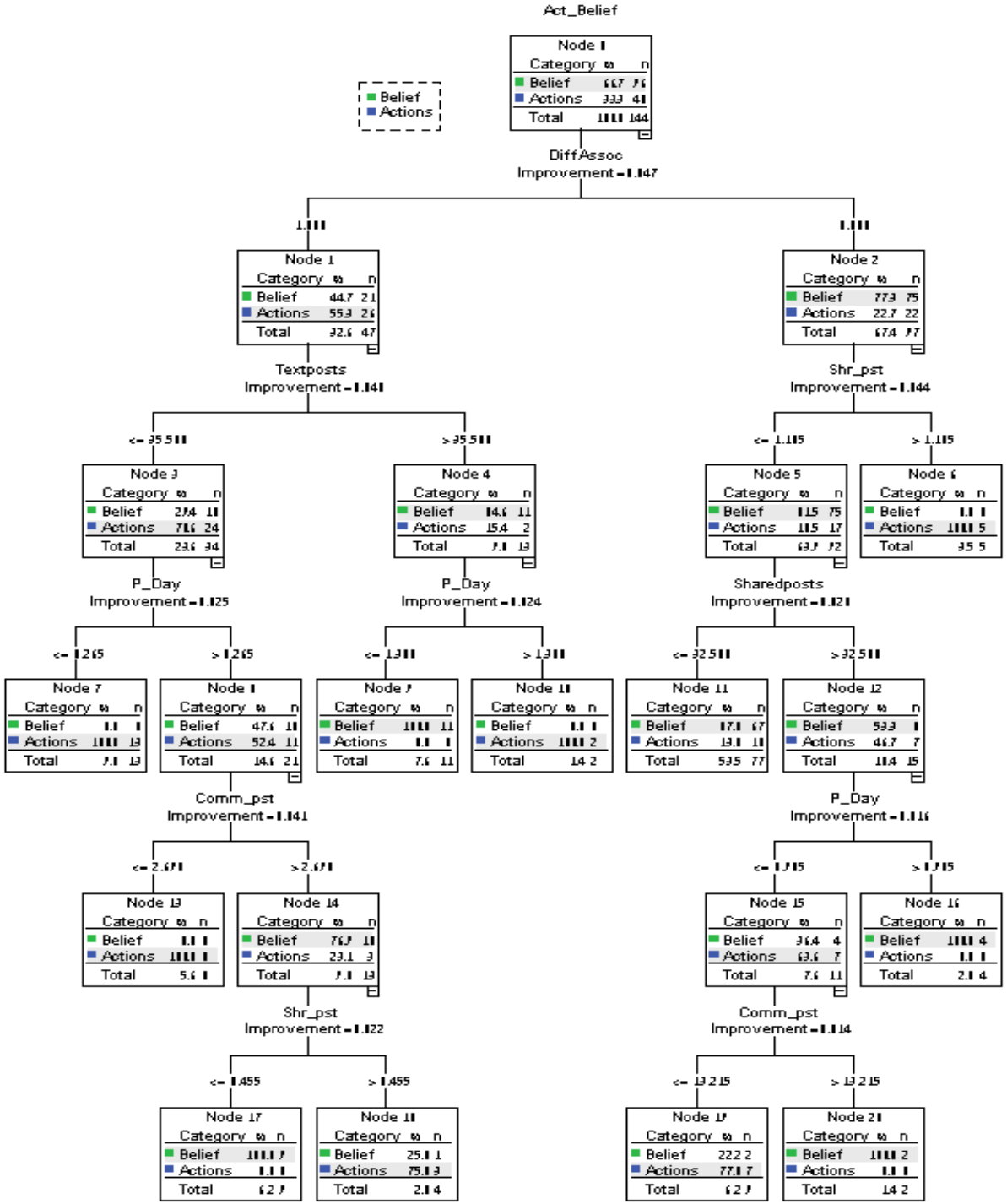
1) Differential associations (Pauwells & Schills, 2016).

2) Opinion leaders (Oeldorf-Hirsch & Sundar, 2015)

3) Lower cognitive sophistication (Baele, 2017)

- Fixation (Meloy et-al, 2012)
- Identification/imitation (Meloy et-al, 2012).
- More self expression is a protective factor(Barbera, 2014; Helmus, York and Chalk, 2012; Özdemir & Kardas, 2014, 2018).
- Supported by the findings from the study in Chile (Valenzuela, Arriagada and Scherman, 2012).

4) Using text-based analysis ignores most of the content, especially for violent radicals



Examples of rules:

If Type 1 in [22.5, 92.31[and Radical3 in [0, 2.735[then 0/1 = 0 in 100% of cases

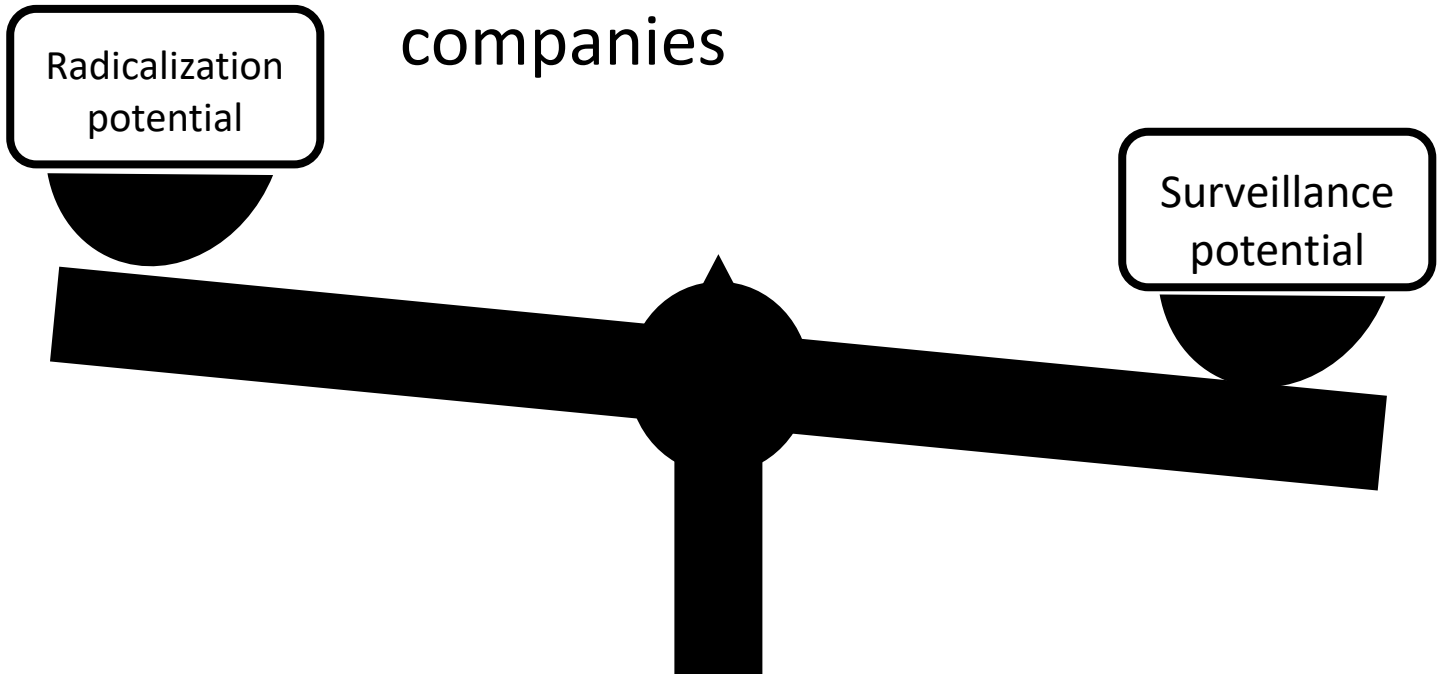
If Posts/day in [1.335, 1.66[and Radical3 in [8.13, 16.415[then 0/1 = 1 in 100% of cases

Model	AUC	Overall	Actions	Beliefs
Logistic Regression	.827	78.47%	77.08%	79.17%
CART	.918	91.0%	79.2%	96.9%
CHAID	.837	81.9%	60.4%	92.7%

Important decisions

- The most active writers are less likely to be violent.
- The internet may provide a better window of opportunity for identification, prevention and intervention than it does for radicals to radicalize (Benson, 2014; Sageman, 2010; Hughes, 2016).

- Leaving content up leaves the windows open.
  - Allows for counter-messaging
  - Improves maintenance of rights and freedoms
  - Improves relationships with IT companies



Success in Israel

- Combine online detection with offline warnings (The Economist, 2017; Barnea, 2018).
- This combines situational prevention with intelligence-led efforts and focussed deterrence.
- A well rounded approach such as this has been shown to be effective against crime.
- Warnings are taken more seriously and legitimacy is maintained (Braga & Weisburd, 2015).
- In Israel, claims of 800 arrests (Santos, 2018), but 400 of them terrorists (Barnea, 2018).
- This is well above the rates of automated detection tools alone.

Conclusions

- Content removal only when necessary (like high-policing in general)
- The internet can act as a protective factor, and may for the most active
- Leaving content untouched has benefits that outweigh removal:
  - Protects free speech
  - Enables more targeted surveillance (better privacy protection)
  - Decreases chances of radicals moving underground
  - Provides legitimacy
  - Keeps the window of opportunity for counter-messaging open
- Automated tools need to move beyond text based analysis
- Automated tools should not replace the analyst but are a ‘tool’ to be used in conjunction with offline tools

References

- Baele, S. J. (2017). Lone-Actor Terrorists' Emotions and Cognition: An Evaluation Beyond Stereotypes. *Political Psychology*, *38*(3), 449-468.
- Barberá, P. (2014). How social media reduces mass political polarization. Evidence from Germany, Spain, and the US. *Job Market Paper, New York University*, *46*.
- Barnea, A. (2018). Challenging the "Lone Wolf" Phenomenon in an Era of Information Overload. *International Journal of Intelligence and CounterIntelligence*, *31*(2), 217-234.
- Benson, D. C. (2014). Why the internet is not increasing terrorism. *Security Studies*, *23*(2), 293-328.
- Braga, A. A., & Weisburd, D. L. (2015). Focused deterrence and the prevention of violent gun injuries: Practice, theoretical principles, and scientific evidence. *Annual review of public health*, *36*, 55-68.
- Fidler, D. P. (2015). Countering Islamic State exploitation of the internet. *Digital and Cyberspace Policy Program, June 2015*
- 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.
- Granger, M. P., & Irion, K. (2014). The Court of Justice and the Data Retention Directive in Digital Rights Ireland: telling off the EU legislator and teaching a lesson in privacy and data protection. *European Law Review*, *39*(4), 835-850.
- Helmus, T. C., York, E., & Chalk, P. (2013). *Promoting online voices for countering violent extremism*. Rand Corporation.
- Hughes, S. (2016). Countering the Virtual Caliphate. *Written testimony submitted before the US House of Representatives, Foreign Affairs Committee*.
- Kardaş, T., & Özdemir, Ö. B. (2018). The making of European foreign fighters: Identity, social media and virtual radicalization. In *Non-State Armed Actors in the Middle East*(pp. 213-235). Palgrave Macmillan, Cham.
- Meloy, J., Hoffmann, J., Guldimann, A., & James, D. (2012). The role of warning behaviors in threat assessment: An exploration and suggested typology. *Behavioral sciences & the law*, *30*(3), 256-279.
- Oeldorf-Hirsch, A., & Sundar, S. S. (2015). Posting, commenting, and tagging: Effects of sharing news stories on Facebook. *Computers in Human Behavior*, *44*, 240-249.
- Özdemir, Ö. B., & Kardaş, T. (2014, October). The Making of European Foreign Fighters. SETA.
- Neumann, P. R. (2013). Options and strategies for countering online radicalization in the United States. *Studies in Conflict & Terrorism*, *36*(6), 431-459.
- 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.
- Sageman, M. (2010). Small Group Dynamics. *Protecting the Homeland from International and Domestic Terrorism Threats: Current Multi-Disciplinary Perspectives on Root Causes, the Role of Ideology, and Programs for Counter-radicalization and Disengagement*, 133.
- Shefet, D. (2016). Policy options and regulatory mechanisms for managing radicalization on the Internet. Retrieved from www.en.unesco.org/sites/default/files/rapport_dan_shefet.pdf
- Shortland, N. D. (2016). "On the Internet, Nobody Knows You're a Dog": The Online Risk Assessment of Violent Extremists. In *Combating Violent Extremism and Radicalization in the Digital Era* (pp. 349-373). IGI Global.
- Sunstein, C. (2017). Hashtag Republic., Princeton: Princeton University Press
- Sutherland, E. H. (1947). Principles of Criminology. Philadelphia: J. B. *P Lippincott*.
- The Economist (2017, June 8), The stabbing intifada: How Israel spots lone-wolf attackers, The Economist. Retrieved from https://www.economist.com/international/2017/06/08/how-israel-spots-lone-wolf-attackers
- Valenzuela, S., Arriagada, A., & Scherman, A. (2012). The social media basis of youth protest behavior: The case of Chile. *Journal of Communication*, *62*(2), 299-314.