



UvA-DARE (Digital Academic Repository)

Developing a social media response to radicalization

The role of counter-narratives in prevention of radicalization and de-radicalization

van Eerten, J.-J.; Doosje, B.; Konijn, E.; de Graaf, B.; de Goede, M.

Publication date

2017

Document Version

Final published version

[Link to publication](#)

Citation for published version (APA):

van Eerten, J.-J., Doosje, B., Konijn, E., de Graaf, B., & de Goede, M. (2017). *Developing a social media response to radicalization: The role of counter-narratives in prevention of radicalization and de-radicalization*. University of Amsterdam, Department of Psychology. https://www.wodc.nl/onderzoeksdatabase/2607-effectiviteit-van-counter-narratives.aspx?refTitle=rhOh-n_RA1rsAzOSCo9H7DBzPhYTfdGS1FErzBhIMluidYMB-elwNOmdZcncrUzeuzaXgKmsnEHR2wqwxqRqhrW-5WZ4hUNMTeLYDVzh0nSBS8-YXKleIFne5Uwfeow0&refId=SmyZrcVMoat3wpl4PoCqW0V8-zmc0W8pdsjKUxUiNkzk1A-6TZP0dY4rskNd11LvoFYNr-sFkQINvJ7hhdUSrA2

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Developing a social media response to radicalization

The role of counter-narratives in prevention of radicalization and de-radicalization.

September, 2017

Jan-Jaap van Eerten, Bertjan Doosje, Elly Konijn,
Beatrice de Graaf & Mariëlle de Goede



COLOPHON

Correspondence address:

B. Doosje

University of Amsterdam

Department of Psychology and Department of Political Science

P.O. Box 15900, 1001 NK Amsterdam, the Netherlands

Email: e.j.doosje@uva.nl

Tel: +31205256885

Cover design: Martijn van de Zuidwind

Copyright © 2017, WODC

Table of contents

ACKNOWLEDGMENTS	6
ABSTRACT	7
1. INTRODUCTION	11
2. RADICALIZATION, NARRATIVES AND SOCIAL MEDIA	13
2.1. Introduction	13
2.2. What is radicalization?	13
2.3. The Jihadi-Salafi Movement	14
2.4. Salafi-Jihadi strategic communication and narratives	15
2.5. Salafi-Jihadi narratives and social media	19
2.5.1. YouTube: Propaganda & Psychological	21
2.5.2. Twitter: Online Out-reach	22
2.5.3. Facebook: Social Networking and Recruitment	23
2.6. Online violent radicalization	24
2.7. Countering the narrative	27
2.7.1. Counter-narratives	28
2.7.2. Alternative narratives	29
2.7.3. Government strategic communication	30
2.8. Summary	31
3. DE-RADICALIZATION THROUGH ONLINE COUNTER-NARRATIVES CAMPAIGNS?	32
3.1. Introduction	32
3.2. The narrow focus of a radical person	32
3.3. Theory 1: Cognitive dissonance theory	33
3.4. Theory 2: Psychological reactance theory	36
3.5. Theory 3: Social Judgement theory	37
3.6. Theory 4: Social identity approach	39
3.7. Theory 5: Side model of deindividuation effects	41
3.8. Summary	44
4. WHAT MAKES A COUNTER-NARRATIVE CAMPAIGN EFFECTIVE?	46
4.1. Introduction	46
4.2. Principles of an effective counter-narrative campaign?	46
4.3. Start a campaign with research and planning	47
4.3.1. Use of theory	48
4.3.2. Define and know the audience	49
4.3.3. Learn from (former) extremist narratives	51
4.3.4. Analyze available resources	53
4.3.5. Determine goals and objectives	54

4.4. Designing your program	54
4.4.1. Developing the message	54
4.4.2. Select credible messengers	66
4.4.2.1. Former extremists	69
4.4.2.2. Victims	70
4.4.2.3. Family members	71
4.4.2.4. Youth	72
4.4.2.5. Other community actors	73
4.4.3. Determine appropriate dissemination channels	74
4.5. Implementing a campaign	77
4.5.1. Disseminating the message	77
4.5.2. Risks and challenges	79
4.5. Summary	81
5. DETERMINING THE POTENTIAL EFFECTIVENESS OF A PROGRAM	84
5.1. Introduction	84
5.2. Types of evaluation	84
5.2.1. Formative evaluation	84
5.2.2. Process evaluation	85
5.2.3. Summative evaluation	85
5.2.4. Realistic evaluation method	85
5.3. Social media monitoring and analytics	87
5.4. Other evaluation tools and techniques	88
5.4.1. (Online) focus groups & interviews	88
5.4.2. One-on-one interviews	90
5.4.3. (Online) surveys	90
5.4.4. Experiments	91
5.4.5. Netnography	92
5.4.6. Sentiment & Content Analysis	93
5.4.7. Social Network Analysis	96
5.4.8. Other evaluation methods	97
5.5. Summary	98
6. THE ROLE OF THE GOVERNMENT	100
6.1. Introduction	100
6.2. The Government and counter-narrative campaigns	100
6.3. Streamlining government strategic communications	101
6.4. Fostering public-private partnerships	102
6.5. Empower effective messengers	104
6.6. Capacity building	105
6.7. Evaluation and research	107
6.8. Strengthening digital literacy and consumption skills	108
6.9. Summary	109

7. SUMMARY, CONCLUSIONS, LIMITATIONS AND FUTURE DIRECTIONS	110
7.1. Introduction	110
7.2. Summary of answers to research questions	110
7.3. Discussion	112
7.4. Limitations of the current study	113
7.5. Future directions	114
REFERENCES	117
APPENDIX A: METHOD	146
APPENDIX B: NEDERLANDSE SAMENVATTING [DUTCH SUMMARY]	147

ACKNOWLEDGMENTS

Narratives are powerful tools in the hands of extremist groups such as ISIS. To what extent is it possible to produce convincing and effective counter-narratives? In this report, we aim to answer this question. This study was commissioned by the Research and Documentation Centre (WODC-2607), Ministry of Security and Justice, The Netherlands.

We like to thank the committee that has supervised us along the way: Daniël Wigboldus (Chair of the committee; Radboud University, Nijmegen), Michael Kowalski, Frederike Zwenk and Casper van Nassau (all from Ministry of Security and Justice, The Hague), Stijn Sieckelinck (Utrecht University, Utrecht) and Reint-Jan Renes (University of Applied Sciences, Utrecht). We have appreciated their valuable and constructive comments. Their general as well as specific feedback clearly has improved this report substantially. We are very grateful to them for their effort and input.

Amsterdam/Utrecht, September, 29, 2017

Jan-Jaap van Eerten, Bertjan Doosje, Elly Konijn, Beatrice de Graaf & Mariëlle de Goede

ABSTRACT

Radical groups thrive by spreading their message. They have increasingly used social media to spread their propaganda and promote their extremist narratives. Including on websites such as YouTube, Facebook, and Twitter. In this study, the main question that we investigate is to what extent it is possible to use counter-narrative programs via social media to de-radicalize individuals or prevent violent extremism. We focus on violent radicalization connected to Salafi-Jihadism. Due to the political and societal context in which this project was initiated, NCTV and WODC were interested in this specific form of radicalization. The method we used consisted of a literature study, interviews (n=8) and three focus groups (n=6, n=7 and n=8) with people with different expertise and backgrounds (including academics, field workers, social media students and a former radical).

In this study, narratives are conceptualized as strategically constructed storylines that are projected and nurtured through (online) strategic communication activities by state and non-state actors in attempts to shape how target audiences feel about or understand events or issues, and ultimately, guide their behavior in a manner that is conducive to their aims and goals. We conceptualize counter-narratives as strategically constructed storylines that are projected and nurtured through strategic communication (or messaging) activities with the intention to undermine the appeal of extremist narratives of violent extremist groups. Narratives have been projected by radical groups through a wide range of social media, including YouTube, Twitter and Facebook. For counter-narratives on social media, we identify three domains: (1) counter-messaging (e.g., activities that challenge extremist narratives head on); (2) alternative messaging (e.g., activities that aim to provide a positive alternative to extremist narratives); and (3) strategic communication by the government (e.g., activities that provide insight in what the government is doing).

A central outcome of our analysis is that to the extent that people have become more radical and strongly identify with an extremist ideology or group, they are *less* likely to be persuaded by any counter-narrative campaign. In general, it's challenging to persuade people with strong convictions through (online) campaigning. They are typically not very receptive to messages that aim to change their views. However, it will be very difficult to challenge people with extremist attitudes. They may not pay attention to these messages or even may adopt a stronger attitude in the other direction than intended as a reaction to this persuasive attempt. Thus, we argue that counter-narrative campaigns may be most useful for prevention purposes. They could perhaps focus on those that appear to be sensitive to extremist views and information – to the extent that they can be identified meaningfully.

In addition, we have raised the question whether or not such counter narrative efforts may have some effect for the individuals from radical groups who show a glimpse of doubt about their group. For such individuals, a counter-narrative message may fall on fertile grounds and a seed may be planted, although this is highly speculative at the moment and more research is needed to support this notion.

A sound starting point for designing a counter-narrative campaign is research. One should develop a sound understanding of (the context of) 'the problem', underlying determinants, where

communication can potentially contribute to a solution. Here it is advisable to consider relevant theories and insights about radicalization and attitude and behavioral change. Beyond borrowing from existing theory, program planners do well to develop a sound theory of change that explicates how campaign efforts are expected to lead to the desired result.

In addition, the *audience* should be carefully defined and segmented on the bases of meaningful variables. This requires thorough audience research. Campaigns targeting an overly broad and general target audience run the risk of being ineffective and may be counterproductive. It is therefore recommendable to focus on narrow, specific audience segments.

In addition to studying the audience, it is important to develop an in-depth understanding of the *extremist narratives* one aims to counter. When deciding upon which elements of, for example, the Salafi-Jihadi narrative one aims to counter, we argue it is also important to consider *why* members of the target audience in question may be attracted to these narratives. Radicalizing individuals that are looking for extremist content may have different motives. Different motives of radical people may request different alternative or counter-narrative content. Some may be drawn to an extremist ideology. However, there are a variety of other reasons that may motivate people, such as the need for group-membership and camaraderie, a search for meaning, the need for excitement and adventure.

As suggested, it is important to clarify the desired outcome, and the steps towards that outcome. Relatedly, at the outset of a campaign, *goals and objectives* should be delineated that clearly specify which change one aims to achieve. Well-defined objectives are specific, measurable, achievable, relevant, and time-bound. Running an effective social media effort also requires adequate *resources* (time, finances, people) which should be assessed at the outset of a program. It is also important to give consideration to potential partners. Involving partners from within local communities in campaign development and delivery may benefit a campaign effort.

Control over the *message* is not strictly assured when it is released on social media. Of course, it is still important to give consideration to the content of the message. As put forward earlier, to the extent that this is possible, it may be fruitful to consider the different motives of radical people. Besides content, it is important to think about different aspects of the message. Messages that use the narrative format have the advantage of using subtle ways to influence people and may overcome various forms of resistance. In addition, double-sided messages (in which one side is being discredited) and messages with a strong emotional appeal may prove to be particularly persuasive.

In terms of *sources*, it is clear that they need to be perceived as credible by the audience in order to function as trustworthy messengers. Commonly mentioned candidates are former extremists, victims of violence, peers and family, as well as key members of communities and civil society actors. The potential of the government to serve as a credible source might be limited. Among other things, the government often suffers from a say-do gap in the eyes of relevant target audiences. Yet, which messenger will in fact effective will depend on the context, the target audience, and the message one wants to relay. Also, it should be noted that source is a murky entity in social media environments. For instance, processes of information transfer may obscure the original source of a message or make it difficult to interpret who the source is. Some

researchers also maintain that social media environments complicate traditional insights about credibility indicators.

In terms of selected *channels*, obviously they should be frequently utilized by the audience one intends to reach. Using multiple channels is argued to be beneficial. Importantly, social media may not be appropriate in all circumstances, and linking online with offline activities is suggested to be particularly effective. In terms of *message dissemination*, one may join already on-going ‘conversations’. Also, it may be beneficial to partner up with people and organizations that have already established a strong social media presence and are popular among the target audience. In some cases, it may also be useful to ask established online (news) outlets to share the message. Furthermore, it may be important to create opportunities for audience engagement. One should also give consideration to aspects such as timing and volume.

Finally, we argue it is important to assess the potential risks, challenges and limitations when developing a communication strategy. These might include losing control of the message, counter-campaigns, threats to safety and well-being of messengers, and multi-tier approval processes that hamper campaign efforts, as well as the notorious difficulty of demonstrating the effect of counter-narrative efforts.

Generally speaking, it is not easy to establish an effect of a policy or measure in counter violent extremism (CVE). For the potential effectiveness of a counter-narrative campaign, it is important to distinguish between a formative evaluation, a process evaluation and a summative evaluation. Specifically tailored at online interventions, we argue that a combination of awareness (e.g., reach and views) and engagement metrics (e.g., likes, shares, comments and emoji responses) can provide some insight into the extent to which a counter-narrative campaign achieved its desired effect. More traditional research techniques such as interviews and focus groups can give insight in the reception of online materials (thus useful in the formative phase), but can also provide input in the interpretation of quantitative data in the summative evaluation phase. Experiments (either offline or online) offer insight in the cause-and-effect-relationship, while “netnography” makes use of data usually obtained from observations of people in their regular social media environments. Sentiment and content analyses are methods to distract the overall evaluation towards a counter-narrative campaign as well as the specific narratives that were put forward in a particular online community. Finally, social network analyses may give insight into the structure of a group and potentially how this might change over time, although the use of this technique in this field has not yet been fully established.

Government actors are not well-suited to act as a counter-narrative producer or messenger. They tend to lack credibility as effective messengers with relevant target audiences. In part, because they are perceived to suffer from a say-do gap, which means their actions and words do not always match in the view of the audience. However, they still can still their part. First, governments can fruitfully engage in streamlining their own strategic communications in terms of explaining their own actions locally and in an international context. Second, the government can play a valuable role by facilitating grassroots and civil society actors best placed to act as counter-narrative messengers. They can do so by establishing an infrastructure to support these initiatives and by sponsoring such efforts (providing help, expertise or financial support) as well as. Yet, they should be cautious, as government endorsement or support for such initiatives may

act as a ‘kiss of death’ and undermine them. Thirdly, they could stimulate thorough monitoring and evaluation, as there is only limited evidence for the effectiveness of counter-narrative efforts. Finally, government efforts could focus on supporting the development of programs in the area of strengthening digital literacy and critical consumption skills

In terms of limitations of the current study, in our analysis, we have focused only on counter-narratives with respect to the Jihadi ideology. As such, it is not possible to generalize these findings to other groups (e.g., extreme right wings groups). Furthermore, due to the fact that there is hardly any strong empirical research available, we had to base some of our analysis on the general literature on persuasion and communication campaigns, as well as some “grey literature” (e.g., policy papers, working papers, and recommendations). This undermines the options to draw strong conclusions from the current analysis, as one might question the extent to which it is possible to generalize from the general field of persuasion to the specific field of online counter-narratives.

We conclude that, although presenting online counter-narratives appears to be intuitively an appealing strategy to employ, our analysis shows that this may not be an ideal option to de-radicalize people. It might be more fruitful to use counter-narratives in a prevention context, in which people’s minds are still open enough to register and process the information presented. Alternatively, it might be possible to explore the options to expose known individuals (e.g., a convicted prisoner) with specific counter-narratives, but ideally not in an online context, but an offline context, in order to directly monitor the reactions of the individuals.

1. INTRODUCTION

People are social beings. As such, they continuously influence each other. When like-minded people influence each other, they have a tendency to become more extreme in their views and attitudes. When supported by an ideology supporting violence, this attitude polarization can translate into a radical belief system, in which people can start to justify the use of violence to reach political goals and/or societal changes. Importantly, nowadays, these social influence attempts frequently occur via social media (e.g., via Internet, Twitter, Facebook, WhatsApp, etc.). In these social media, people are being influenced by the narratives that radical groups and their supporters project. Important questions are: is it possible to use social media to present counter-narratives, in an attempt to prevent radicalization? And once people have adopted a radical belief system, can social media be used to de-radicalize individuals?

In this report, we explore the role of the use of counter-narratives via social media as a potential means to prevent radicalization of potentially susceptible individuals, or to de-radicalize people with radical beliefs. These are important questions as radicalization is a process that may lead to extreme and violent behavior. Because it has proven to be difficult to distract a clear terrorist profile (Silke, 1998), people (e.g., researchers, policy makers and first-line workers) have directed their attention to radicalization. Radicalization is a process by which people increasingly adopt more extreme attitudes and behavior that might involve approval of the use of violence by others or displaying this violence themselves to stimulate fear in the general population in an attempt to instigate changes in society.

Narratives (in the present context briefly described as strategically constructed storylines – see Chapter 2 for a full discussion and definition of narratives and counter-narratives) can be used by extremist groups to influence susceptible individuals in a radical direction and attract support. By projecting narratives via social media, extremists may encourage support for their ideology which involves (a) a clear us vs. them distinction, accompanied by a perceived superiority of their own group (the in-group) and a clearly inferior and de-humanized perception of the enemy (the out-group); (b) a strong perception of injustice or grievances: the in-group is threatened by the out-group; (c) a lack of trust in current institutions (politics, justice system) to address their grievances; (d) a perceived need to use violence to achieve societal and/or political changes (including the belief that such violence will be effective) (Doosje, Loseman & van den Bos, 2013).

In this report, we examine the extent to which counter-narrative initiatives via social media can be effective in preventing people from radicalization or can de-radicalize people. Specifically, we formulate the following research questions:

- (1) How can we conceptualize narratives and counter-narratives?
- (2) How are narratives and counter-narratives used via social media?
- (3) To what extent is it possible to use counter-narrative programs via social media to de-radicalize individuals or prevent violent extremism?
- (4) What are the pre-requisites for a counter-narrative program for it to be effective?
 - a. Which social media are most suitable and why?

- b. What can we learn from examples of counter-narrative programs that have been operational in other democratic countries?
 - c. What can we learn from examples of social media campaigns in other domains, such as health care and environmental issues?
 - d. What are the potential risks for unwanted side effects?
- (5) How can the potential effectiveness of such a counter-narrative program be determined?
- (6) What can be the role of the government in such a counter-narrative program?

We examine all these research questions using three methods¹: (1) a literature study, (2) interviews with experts (academics, policy makers and first-line workers) and (3) focus groups with experts and young people involved in social media and/or radicalization topic. Our methodological approach is further described in Appendix A. In the following, we present a brief outline.

In Chapter 2 (addressing research questions 1 and 2), we first present theoretical conceptualizations of radicalization in the existing literature and describe the role of narratives and counter-narratives in social media, focusing on Jihadi-narratives². Subsequently, in Chapter 3 (dealing with research question 3), we examine how the potential effectiveness of counter-narrative programs via social media to prevent radicalization or instigate de-radicalization may depend on the phase of the radicalization process of a person. Chapter 4 addresses research question 4, namely whether and how counter-narratives initiatives via social media can play a role in preventing radicalization, by examining current examples as well as investigating the possibilities by focusing on communication campaigns in other domains. We outline potential strong elements of counter-narrative programs as well as potential limitations of such an approach. In Chapter 5, we describe how the potential effectiveness of such a program can be determined, including the question whether it is possible to test such a counter-narrative program in experimental research designs (research question 5). In Chapter 6, we discuss the potential role of the government in producing or stimulating counter-narratives (research question 6). Finally, in Chapter 7, we summarize and discuss the main results, outline the limitations of the current study and present potential avenues for further research. In Appendix A, we specify the method that we have used to answer the research questions. For every research question, we used a combination of the three methods (i.e., literature study, interviews and focus groups).

¹ In Appendix A, we describe our method in more detail and explain our choices for the people we have selected for the interviews and the focus groups.

² We realize that this does not cover other forms of radicalization. Given the political and societal context when this project was initiated (2015), the NCTV/WODC specifically were interested in this form of radicalization. This implies a specific focus and we are not in a position to generalize the current findings to other radical contexts or groups.

2. RADICALIZATION, NARRATIVES AND SOCIAL MEDIA

2.1. INTRODUCTION

In this chapter, we aim to provide answers to research questions 1 and 2, namely (1) how can we conceptualize narratives and counter-narratives?; and (2) how are narratives and counter-narratives used via social media? Because identifying a counter narrative is dependent on the perspective taken by the narrator, it is important to first provide a general introduction in the field of radicalization (Section 2.2) and more specifically the Salafi-Jihadi movement (Section 2.3), and how they make use of narratives (Section 2.4) through strategic communication via social media (Section 2.5) to attract youngsters from The Netherlands, Europe and other parts of the Western world. Only when we have outlined how online violent radicalization might work (Section 2.6), we are in a position to present our conceptualization of counter-narratives (Section 2.7) and provide concrete examples of how they currently are used in social media (Box 1 to Box 3). We end with a summary of this chapter (Section 2.8).

2.2. WHAT IS RADICALIZATION?

What exactly is radicalization? There is no consensus definition of radicalization, and the concept remains highly contested in the literature. Although the literature offers a wide variety of definitions, many share the aspect that it is a *“process that involves the adoption of beliefs and attitudes that are in opposition to the mainstream status quo and dominant sociopolitical discourses”* (Macnair & Frank, 2017, p. 148).

In and of itself, adopting radical beliefs or attitudes is not necessarily problematic or negative (Davies, Neudecker, Ouellet, Bouchard, & Ducol, 2016; Macnair & Frank, 2017). It is not altogether uncommon that people hold radical views, and, in most cases this does not result in any violent or otherwise unlawful acts (Macnair & Frank, 2017). In fact, radicalization may actually be a driver of positive social change. As pointed out by Pauwels et al. (2014) *“In the course of history, “radicals” have fought for and changed a lot of things that are now considered as normal and necessary”* (p. 16). Think for example of Nelson Mandela or Martin Luther King who were at one time considered radical by many as they opposed prevailing views in society (ibid.).

However, radicalization becomes a concern when it leads to beliefs and attitudes that sanction, legitimize and compel violence as a means to achieve social change (Davies et al., 2016). As such, the present report is concerned with *violent radicalization* or radicalization into extremism, a process whereby radical views develop into *“a willingness to directly support or engage in violent acts”* (Dalgaard-Nielsen, 2010, p. 798). An important distinction that follows from this is that between radicalism and extremism. The most important difference is that the latter accept to exercise violence as a legitimate means to achieve particular goals (Feddes, Nickolson, & Doosje, 2015).

Violent radicalization may ultimately, but not necessarily, lead to people to commit acts of terrorism and other violent extremist actions. It is important to note, however, that this transfer is not inevitable (Hafez & Mullins, 2015; Pauwels et al., 2014). In fact, most individuals who support the use of violence never engage in violent acts (Khalil, 2014). In this regard, various

authors highlight the critical distinction between cognitive and behavioral dimensions of radicalization (e.g., Borum, 2011; Hafez & Mullins, 2015; Khalil, 2014). Furthermore, the literature emphasizes radicalization in terms of developing extremist attitudes and beliefs is one of many 'pathways' into violent extremist action (Borum, 2011; Neumann, 2013b; Pauwels et al., 2014). As put by Borum (2011): *“Different pathways and mechanisms operate in different ways for different people at different points in time and perhaps in different contexts”* (p. 8).

Violent radicalization can be related to a number of movements and ideologies (e.g., right-wing, left-wing, nationalist/separatist, single issue and so forth). In this study, we predominantly focus on violent radicalization connected to the Salafi-Jihadism. Given the political and societal context when this project was initiated (2015), the NCTV and WODC of the Ministry of Security and Justice in The Netherlands, who commissioned this research, specifically were interested in this form of radicalization. This implies a specific focus and it is important to note that this means that we are not in a position to generalize the current findings to other radical contexts or groups.

2.3. THE JIHADI-SALAFI MOVEMENT

Jihadi-Salafism refers to a distinct ideological movement in Sunni Islam, which seeks to use violence and terrorism in order to establish a restoration of the al-khilafah (i.e., the Caliphate) and a reinstatement of perfect Divine law or sharia (which is in fact conflated with fiqh, which may be understood as the imperfect effort to interpret Divine law; (Beutel et al., 2016; Rane, 2016). Salafi-Jihadism is an outgrowth of the Salafi movement, a global Sunni Islamic renewal movement which originated in Saudi Arabia (Al Raffie, 2012). The meaning of Salafism is derived from the Arabic word al-salaf al-salih - meaning 'pious forefathers' - which refers to first three generations of Muslims who in the view of Salafists exemplify the way that all Muslims should live today (De Koning, 2013). The Salafi movement aims to “revitalize Islam by promoting an idealized vision of the lives of the first Muslims of the seventh century AD, and by persuading Muslims to live according to that vision; an entreaty that its followers find more just and satisfying than the life and circumstances of the present” (De Koning, 2013, p. 19).

In line with this, Salafists typically adhere to *“literalist, conservative and often puritanical approaches to Islam based on an interpretation of what they perceive to be the religion’s original beliefs and practices...”* (Rane, 2016, see chapter 9, section *“Islamist Extremism”*, 2nd paragraph). As Salafists wish to purify the Islam, they are keen to distinguish themselves from what they perceive to less strict adherents, and those groups that they consider deviant in general (Wagemakers, 2009). They also reject Muslim religious traditions and schools of jurisprudence which have developed over many centuries, and hence, what can be described as Islamic orthodoxy.

Importantly, Salafism should not be reduced to Jihadi-Salafism. In fact, some Salafi groups respect democracy and challenge violent fractions (Roex, 2014). Scholars typically distinguish three branches: a-political Salafism (quietist/purist), political Salafism, and Salafi-Jihadism (or Jihadi-Salafism) (e.g. Schmid, 2014; Wagemakers, 2009; Wiktorowicz, 2006). Our focus here is on the latter. Salafi-Jihadism is described to fall outside of the mainstream, and is only adhered to by a small minority (Rabasa & Benard, 2015; Rane, 2016). The main distinction between Jihadi-

Salafism and the non-violent a-political and political strands, is the formers willingness to support and use violence (including violence against other Muslims and non-combatants)³.

Jihadi-Salafists legitimize violence through a selective literal interpretation of Islamic scripture. This involves a reinterpretation of the concept of jihad⁴ that elevates offensive armed struggle, typically argued to be defensive, over a broader repertoire of methods of struggle that have been established in Islamic theology and jurisprudence (Lentini, 2013). Here, Salafi-Jihadi groups such as Al-Qaeda and affiliates interpret the term defensive as follows, “*the meaning of ‘defensive’ extends to striking at the enemies of God wherever they are. They see themselves as vanguards who are willing to fulfill the religious obligations of jihad, even if other Muslims do not*” (Rabasa & Bernard, 2015, p. 31). By exception, ISIS⁵ self-proclaimed “caliph” Abu ‘Umar al-Baghdadi has also emphasized the offensive form of jihad, premised on the uprooting of shirk (idolatry) wherever it is found (see Bunzel, 2016).

In recent years, the Jihadi-Salafist movement has made increasing use of social media to help spread Jihadi-Salafist ideology and actively reach out to young people who may be susceptible to radicalization and recruitment. ISIS in particular has gained a solid reputation with regard to its on-going and technologically sophisticated strategic communication campaign in which mainstream social media platforms have played a pivotal role. At the core of strategic communication are strategic narratives which are projected to the outside world via mainstream social media platforms such as YouTube, Twitter, Facebook, and more or less public / insular forums. In the next two sections, we outline how we conceptualize strategic communication as well as strategic narratives.

2.4. SALAFI-JIHADI STRATEGIC COMMUNICATION AND NARRATIVES

What, exactly, strategic communication is and what it includes remains matter of debate. For our current endeavor, we follow the definition provided by Halverson, Corman, and Goodall (2013), who defined strategic communication as “*communication intended to persuade an audience in support of one or more specific goals*” (p. 179). When applied to Jihadi-Salafist groups, they argue that strategic communication supports three main goals: The first is ‘resisting’, which refers to fighting foreign invaders who have exploited Muslims in their lands as well as enemies who seek to destroy the *Ummah* from within. The second goal, ‘rebuking’, involves the discrediting apostate leaders of Muslim countries who have persecuted Muslims and are complicit in the acts of foreign

³ Whereas only the Salafi-Jihadism is inherently violent, it should be noted that also the non-jihadist variants are often regarded as extreme by the prevailing norms of West European societies, at odds with European secular freedoms (e.g., the separation of state and religion, popular sovereignty, and respect for minority rights), and opposed to integration into mainstream West European societies (De Koning, 2013; Schmid, 2013). Moreover, some believe that due to a lack of “firebreaks” between more mainstream Salafism groups and movements and extremist fringes, involvement in former can potentially provide a gateway to the latter (Rabasa & Bernard, 2015). In this sense, it has been suggested these groups and movements are a “conveyer belt”. Yet, others argue that the more mainstream Salafi groups may actually function as a “firewall” that prevents those radicalizing from strolling further along the path towards violent extremism (Schmid, 2013).

⁴ It needs to be noted that the concept of jihad refers in its most fundamental form to an internal spiritual struggle. This spiritual struggle is also referred to as the Jihad al-Akbar or the ‘greater jihad’.

⁵ Islamic State of Iraq and Al-Sham (ISIS). Also known as Dawla al-Islamiya fi al-Iraq wa al-Sham (DAISH), Islamic State of Iraq and the Levant (ISIL), Islamic State (IS).

invaders; The last goal is ‘renewing’, the goal of reversing the decline of the Islamic Civilization by reestablishing the Caliphate⁶.

Like strategic communication, the concept of narrative has garnered considerable academic debate. As noted by Corman (2011), *“for all this interest in narrative, there is little consensus about what it is and how it is used [...] It is not far from the truth to say there are as many theories of narrative as there are theorists”* (p. 36). Indeed, there are a variety of conceptualizations, spanning different disciplines. However, in the literature on strategic communication, narratives are argued to be at the core of a strategic communication strategy (Dimitriu & Graaf, 2016). In this view, narratives are inherently strategic.

To explain this more clearly, the point of departure for those utilizing the concept is that narratives are a key means by which people impose order on their experiences and make sense to them (de Graaf, Dimitriu, & Ringsmose, 2015). Or as succinctly put by Sedgwick (2012) *“[i]n order to have meaning, events must be interpreted and placed in a narrative”* (p. 363). People employ more or less sophisticated storylines through which *“a shared sense is achieved representing a past, present and future, an obstacle and a desired end-point”* (Miskimmon, O’Loughlin, & Roselle, 2012, p. 1). Thereby they connect phenomena that may seem otherwise unrelated around some causal transformation (ibid.). In a similar fashion, strategic narratives tie together apparently unconnected events and actions by providing an overarching structure through which they can be understood (de Graaf et al., 2015). Importantly, strategic narratives do not arise spontaneously – rather they are consciously and deliberately designed and nurtured (de Graaf et al., 2015; Freedman, 2006).

‘Strategic’ narratives have been defined as *“compelling storylines that can explain events convincingly and from which inferences can be drawn”* (Freedman, 2006, p. 6). They are a form of deliberately constructed discourse through which state and non-state actors (e.g., violent extremist groups) attempt *“to give determined meaning to past, present, and future”* (Miskimmon, O’Loughlin, & Roselle, 2015, Chapter 4, section “strategic narratives”, paragraph 1). Strategic narratives are constructed and projected by these actors in order to shape how target audiences feel about or understand events or issues, and guide their behavior in a manner that is conducive to their goals and interests (Ibid.). At the most basic level, they frame issues and suggest an appropriate response (Freedman, 2006). However, as Arquilla and Ronfeldt (2001) posit, they also *“express a sense of identity and belonging”* and *“a sense of cause, purpose, and mission”* (p. 328).

⁶ A complementary perspective is offered by Corman and Schiefelbein (2006) who argue that the three main goals are: legitimation, propagation, and intimidation. First, given that these groups engage in acts of violence and as such violate the norms of the overwhelming majority of Muslims as well as how they perceive the tenets of Islam, there is a constant need to legitimate their cause and establish their social and religious viability in relation to their own faith group. Second, Salafi-Jihadi groups are expansionist in nature, hence, they aim to propagate their movement by spreading messages among people that reside in areas where they want to expand and may potentially be sympathetic. Furthermore, they utilize strategic communication as a means to intimidate opponents, both inside and outside their own faith group. They do so in order to bolster their own position as well as inhibit dissent and resistance (Corman & Schiefelbein, 2006; Rane, 2016).

Strategic narratives do not necessarily need to be rational or analytical to be effective (Freedman, 2006). While they may be grounded in evidence and experience, they can often rely on “*appeals to emotion, or on suspect metaphors and dubious historical analogies*” (ibid., p. 23). Various authors have highlighted important ingredients of a strategic narrative (e.g., de Graaf et al., 2015; Freedman, 2015; Schmid, 2014). Among other things, it is suggested that strategic narratives must relate to the culture, experience, beliefs and concerns of the intended audience and resonate with their historical and cultural understandings. They should provide a clear mission or purpose and a promise of success. Moreover, they have to be presented in a consistent and coherent manner, and must be able to survive counter-narratives that challenge them.

Importantly though, crafting and projecting such a narrative is far from a straightforward matter. In part, because it is hardly ever construed or projected by a single “spokesperson”. Rather, they are constantly negotiated social products and articulated through multiple ‘voices’ which all may well see and (consciously) promote a slightly different interpretation of events or circumstances (Maley, 2015). As put by Maley (2015), “[i]n complex governmental systems, often a multiplicity of interests jockey with each other for access to the airwaves, offering subtly different narratives in defense of their own specific organizational interests” (See Chapter 5, section “Rhetoric, strategic narratives...”, paragraph 2). As such, he argues that a strategic narrative is, to some extent at least, metaphorical. Moreover, as hinted at earlier, strategic narratives are continuously contested. Other actors, groups and entities (e.g., the press, NGO’s, religious or community leaders, etc.) can attempt to intervene, influence or undermine the process of narrative construction. Furthermore, as we will explore more thoroughly later on, groups may compete and nourish their own strategic (counter-)narratives to further their interest and or undermine those of their adversaries (Price, 2015). Also, it should be emphasized that while particular interpretations and perspectives may be carefully promoted, whether they get appropriated as intended and are subsequently retold – and thus continue to exist - is something over which ‘the promoter’ has some influence but no control (Archetti, 2013; Bushell, Workman, & Colley, 2016). This holds especially in the porous online environment, where one “*can no longer simply convey a narrative to a single audience – instead it will be commented on, interpreted, appropriated and retold by multiple actors, to multiple audiences. In this way the narratives take on a ‘life of their own’ once they are put out into the public realm*” (Bushell et al., 2016, p. 8).

Now that we have a clearer understanding of what a strategic narrative encompasses, what its key ingredients are, as well as what may hinder their impact, we will examine what strategic narratives Salafi-Jihadi groups (such as Al-Qaeda and ISIS) have projected in attempts to sway audiences to view events and issues on their terms. The narratives that groups such as ISIS and Al-Qaeda project diverge in important aspects, and may vary depending on the purpose, location, audience and time (Beutel et al., 2016; Weimann, 2016). Yet, it is argued that, in their ‘core’, they share three key elements which are captured in what Sedgwick (2012) has termed the ‘ABC-model’. According to this model, Salafi-Jihadi narratives provide: an (A)ccount of the existing order (alternatively described as a ‘basic grievance’ (Schmid, 2014), a vision of a (B)etter world, and a means of (C)hange. That is, (1) it provides an account of a Muslim World in chaos, a ‘basic grievance’ caused by non-Muslim others, commonly an alliance of Western (Christian) ‘crusaders’ and Zionists and/or with local secular dictators; (2) it provides a vision of an ideal society, the restitution of the Caliphate governed in accordance with Sharia, which serves to replace the corrupt Western-

backed governance structures of local apostate regimes; and, (3) it aims to present a promising pathway to move from the grievance to the ideal society, namely, the violent Jihad⁷ (Beutel et al., 2016; Schmid, 2014).

In essence, this overarching ‘core’ narrative encapsulates four separate narratives: political, moral, religious, and social psychological (Leuprecht, Hataley, Moskalenko, & McCauley, 2010; Weimann, 2015). The ‘political narrative’ that focuses on the wrongdoings and evils of the (Christian-Zionist) West and local secular dictators; the ‘moral narrative’ that highlights inherent contradictions in liberal democracies and designates Western values as drivers of moral decay; the ‘religious narrative’ that justifies the use of violence to defend Islam from or fight back against the evil West; and, lastly, the ‘social-psychological narrative’, which employs “*a classic in-group/out-group strategy to brand as infidels those who do not buy into this syllogism, while promoting the brotherhood of arms as a means of countering social exclusion and of fulfilling a yearning for adventure and sacrifice that compels the ‘true believer’*” (Leuprecht et al., 2010, p. 43).

The strategic narratives these groups project holds power, because they “*provide simplistic, pre-packaged explanations of past events and contemporary conditions*” (Mahood & Rane, 2017, p. 3), and appeal to real and perceived grievances of many Muslims (Schmid, 2014). In this regard, they are also backed up by facts and examples such as Muslim casualties as a result of foreign interventions in the Middle-East, the actions of repressive regimes in Muslim-majority countries, acts of marginalization and discrimination against members of Muslim minority communities, and other real and perceived grievances (Heffelfinger, 2010). They also offers hope for a better world, and provides clear directions when it comes to working towards this goal, relaying a sense of purpose (Mair, 2015). By strategically appropriating and transforming key concepts and elements from Islamic sacred texts and Muslim histories in their narratives⁸ – which can relate strongly to identity, they give them a unique appeal and apparent justification. Moreover, it makes them relatively difficult to challenge, as “*any attack on it can be portrayed as an attack on Islam itself, rather than as an effort of debunking an eclectic patchwork of cherry-picked elements from sources considered sacred*” (Schmid, 2014, January, p. 4).

In sum, narratives are at the core of strategic communication. State and non-state actors strategically construct and project them in their (online) communications in order to persuade audiences to perceive particular events and issues on their terms, and ultimately, guide their

⁷ Schmid (2014, pp. 6-7) mentions seven additional key elements in his examination of al-Qaeda’s overarching narrative. These include: (1) ‘shahid’ (suicide/martyrdom) operations are a legitimate tactic of warfare; (2) in the struggle against enemies no distinction should be made between civilian and military targets; (3) killing other Muslims is proportionally just on the grounds of their complicity with either the West or with apostate regimes; (4) the practice of ‘takfir’ (excommunication) should be performed when Muslims fail to conform with sharia law or, more generally, when they do not accept the beliefs and practices that are deemed right by Salafi-Jihadis; (5) it is an individual obligation for righteous Muslims to pursue violent Jihad; (6) there is a clash of civilizations between the Muslim (Dar-al-Islam) and those that remain outside the Islamic frontier (Dar-al-Harp) until entire world-system has submitted to sharia rule; (7) the establishment of a government ruled in accordance with sharia law is as stepping stone towards a Shariah-based system of world governance.

⁸ Halverson, Corman and Goodall (2013) identify 13 “master” narratives (i.e., trans-historical narratives that are deeply culturally embedded) which originate from Islamic sacred sources and Muslim histories and are frequently invoked by Salafi-Jihadi groups. Examples are the narrative of the abolishment of the Caliphate, crusaders, and *jahiliyyah* (conveying society is in a state of godless ignorance). These last two appear to permeate ISIS propaganda (Mahood & Rane, 2017).

behavior in a manner that is conducive to their goals. When applied to Salafi-Jihadi narratives, we noted that they can take a variety of forms, yet, three overarching elements can be identified: (a) they project a basic grievance paired with an assignment of blame - often times the West; (b) they offer an alternative vision of the good society - the caliphate; and, (c) offer a promising pathway to realize this vision - violent Jihad.

Social media have created new opportunities for Salafi-Jihadi groups to propagate their narratives (Weimann, 2015). In the next section, we explore how Salafi-Jihadi groups have utilized internet and social media to project their strategic narratives through online strategic communication (or messaging). Although they are by no means the only instrument Salafi-Jihadi groups use to get their message out, their participatory culture and architecture (which allows users to easily connect with each other, and consume, produce and distribute content) has provided ample opportunities to legitimate and propagate their cause as well as to intimidate opponents (Weiman, 2015). In the next section, we describe what we exactly mean by social media and how they have been used to project Salafi-Jihadi strategic narratives.

2.5. SALAFI-JIHADI NARRATIVES AND SOCIAL MEDIA

Salafi-Jihadi groups have long used the internet for strategic communication purposes and have followed suit with technological developments. Initially, they relied heavily on static webpages, and subsequently, interactive fora (some public, some “hidden”)⁹. More recently mainstream, ‘social media’ platforms have become a primary avenue of communication (Bartlett & Reynolds, 2015). Social media are Internet-enabled means of social interaction among people in which they create, share and/or exchange information and ideas in virtual communities and networks (Ahlqvist, Bäck, Halonen, & Heinonen, 2008). They are dependent on mobile and web-based technologies to create highly interactive platforms or tools that are used by individuals and communities to share, co-create, discuss, and alter (user-generated) content (Kietzmann, Hermkens, McCarthy, & Silvestre, 2011). Such content may take a wide variety of forms such as texts, photos, pictures, videos, audio tracks, but also presence information (geotags), tags and bookmarks (Nissen, 2015). Categories of platforms and tools that allow users to tap into networks and facilitate such interactions include, but are not limited to, (micro)blogs, social networking services, content communities (i.e., file-sharing platforms) and instant-messaging services (Kaplan & Haenlein, 2010).

Some distinguishing features of social media are that they are relatively inexpensive to use and easy to access, allow for a fast flow of information and are generally designed to be user-friendly (Weimann, 2015). Moreover, social media have made it increasingly easy to reach out to and connect with geographically dispersed audiences (Amble, 2012; Nissen, 2015; Weimann, 2015). Unlike traditional media such as newspapers and television, which are controlled by gatekeepers and characterized by top-down one-to-many transmissions, social media allow anyone to become a consumer and a producer of information at the same time (i.e., ‘prosumers’) (Amble, 2012; Weisman, 2015). They provide users with the opportunity to simultaneously interact with both

⁹ Barlett and Reynolds (2015) describe that in 1999 almost all known terrorist groups had established a presence on the internet with text-heavy websites. Half a decade later most groups had completed the transition to interactive forums.

broad and narrow audiences and vice versa, allowing two-way, one-to-many, and many-to-many communications: *'from user to user, user to audience, audience to user, or audience to audience...'* (Carr & Hayes, 2015, p. 52).

The proliferation of social media and mobile technologies has enabled Salafi-Jihadi groups to project their narratives more widely and farther than ever (Fink & Barclay, 2013). Now, rather than waiting for individuals to stumble upon their websites and forums, social media permit them to directly target and market their ideas to potential sympathizers and engage in direct person-to-person interaction. At the same time, social media provide sympathizers with opportunities to reach out to them under conditions of relative anonymity (Weimann, 2014; 2015). Given their popularity among mainstream audiences, in particular youth (which are, generally speaking, more impressionable), they also provide a particularly potent means to 'get their message out' to large audiences. They have made it increasingly easy and cheap for extremist groups to disperse their content that supports their narratives. Not only because they can be utilized free (Weimann, 2015), but also because they can rely on a dispersed network of ideologically aligned groups and supporters who 'repackage' content for consumption by local audiences. This process has been referred to as the 'glocalization' of extremist messaging, and has allowed Salafi-jihadi groups to reach audiences outside their traditional orbit (Amble, 2012; Fink & Barclay, 2013). Yet, such content changes do result in some loss of control of their narratives (Nissen, 2015).

The fact that they can rely on a decentralized social media distribution network has also made it difficult to combat the circulation of extremist content on the web. Unlike static websites and forums that have been prone to attacks by law enforcement agencies and activist, social media allow Jihadi groups to operate as 'swarm casts'. Disseminating content across multiple social media platforms and through dispersed networks of supporters allows them to respond rapidly to account deletions. This has permitted them to mitigate some of the effects of internet policing and censorship (Bartlett & Reynolds, 2015; Fisher, 2015; Weimann, 2015). By utilizing social network sites which are inherently difficult to sensor to their media repertoire - both on the 'dark web' as well distributed open platforms - they have also ensured that content remains widely available and accessible, even when mainstream service providers take steps to undermine their presence and remove their communications (Bartlett & Reynolds, 2015).

To illustrate, below we describe how Jihadi-Salafist violent extremist have harnessed the power of three of the most popular social media applications – Facebook, Twitter, and YouTube – for strategic communication purposes. While all of these service providers have policies and reporting systems in place that prohibit and curtail any violent, criminal or hate content, and have undertaken increasingly proactive efforts to suspend accounts that misuse their platforms, they have nonetheless been effectively leveraged by extremist groups for strategic communication purposes. These are of course by no means the only platforms extremist groups have utilized. Other well-known examples are Ask.fm, Instagram, WhatsApp, PalTalk, Viper, JustPaste.it, and Tumblr (Ferguson, 2016). Importantly, technologies constantly evolve and new platforms emerge continuously. Audiences usage patterns tend to change relatively quickly (Stevens, 2010). Likewise, opportunities for online policing are constantly in flux. This is reflected in the tools extremist groups use and the way they make use of them (Bartlett & Reynolds, 2015). For instance, terrorist networks like Al-Qaeda and ISIS have increasingly come to rely on spreading propaganda and planning terrorist attacks through their own encrypted mobile apps. They do not like to use

software created in the west, liable to government subpoenas or searches. Instead, they make good use of commercial apps like the Russian originated Telegram, which features almost unbreakable encrypted data services. ISIS has relied on Telegram to plan and coordinate a series of attacks between 2013 and 2017, and to use it for media distribution as well (Alkhouri & Kassirer, 2016)

2.5.1. YOUTUBE: PROPAGANDA & PSYCHOLOGICAL

YouTube is probably the most well-known ‘content communities’ of which the main purpose is the sharing of video content between users. According to YouTube (2017), it has over a billion users worldwide. Besides posting and watching videos, one can comment on and link to videos. It also provides basic features of a social networking platform. Users can create a user channel with a basic personal profile, which displays such things as friends, comments, and recently watched, posted and favorite videos (Kaplan & Haenlein, 2010; Smith, Fischer, & Yongjian, 2012; Weimann, 2015).

The dissemination of video content has long been an important strategy of violent extremists to project their narratives, and YouTube is one content communities that has been used to project them to a broad audience (Conway & McInerney, 2008; Vitale & Keagle, 2014; Weiman, 2015). ISIS in particular has made sophisticated use of YouTube, although it should be noted they are only one of many violent extremist groups that have utilized the platform. Within ISIS’ official media apparatus there are various media wings that have made use of the YouTube platform. They include its ‘original’ one, al-Furqan Media, as well as al-Hayat Media which produces video content which is deliberately designed to speak to non-Arab speaking youthful target audiences. The videos are carefully scripted and are high quality productions, which use a variety of cinematographic techniques, some of which have a ‘video game’ like quality (see for example ‘No respite’¹⁰; Vitale & Keagle, 2014). Perhaps the most (in) famous example of YouTube use by ISIS, is their series of videos of violent executions¹¹. These videos mainly represent efforts to intimidate its foes and Western audiences (Veilleux-Lepage, 2014).

However, the full spectrum of the material that has been posted on YouTube by ISIS is much broader, and among other things, includes form videos that highlight battlefield successes, portray the bloodshed and violence of the enemy, romanticize the daily lives of the group’s fighters, and depicts the idyllic life in and the virtues of the so-called “caliphate”. Examples include the videos ‘The End of Sykes-Picot’¹², ‘The Flames of War’¹³, ‘The Clanging of the Swords I-IV’

¹⁰ Al-Hayat Media (24 November, 2015). *No Respite*. [Video File]. Retrieved on 7th of September 2017, from LiveLeak: http://www.liveleak.com/view?i=553_1448792160

¹¹ These include the execution of captured soldiers from Syria (25 July 2014), journalists James Foley (19 August 2014) and Steven Sotloff (September 2014), humanitarians David Haines (13 September 2014), Alan Henning (3 October 2014), and Abdul-Rahman (formerly Peter) Kassig (16 November 2014).

¹² Al-Hayat Media (29 June, 2014). *The End of Sykes-Picot*. [Video File]. Retrieved on 7th of June 2016, from: http://www.liveleak.com/view?i=d43_1404046312

¹³ Al-Hayat Media (20 September 2014). *Flames of War*. [Video File]. Retrieved on 7th of June 2016, from: http://www.liveleak.com/view?i=5c2_1411222393

^{14,15,16,17,18}, and ‘Upon the Prophetic Methodology’¹⁹, which legitimize the existence of the ‘Islamic State’ and serve as an attempt to lure new recruits (Vitale & Keagle, 2014; Veilleux-Lepage, 2016).

Furthermore, to supplement these videos, the Hayat Media team created a series of short High Definition propaganda videos - so called ‘mujatweets’ – which specifically aim to portray the movement’s softer side and the prosperous day-to-day life in their Caliphate²⁰. Examples include fighters handing out candy to children, Jihadi’s visiting wounded in the hospital, and townsfolk talking about the improvements that ISIS has brought to their local village, town or city. These videos depict life in the Caliphate as stable and normal, while at the same time portraying the popularity, attractiveness and strength of ISIS’ leadership (Vitale & Keagle, 2014).

In addition to being used for its video hosting capabilities, YouTube has also fulfilled an important function as a social networking tool (Conway & McInerney, 2008; Weimann, 2015). Specifically, YouTube comment sections as well its capacity to send private messages, enable those producing and engaging with violent extremist content to quickly identify each other and establish contact (Weimann, 2015).

2.5.2. TWITTER: ONLINE OUT-REACH

Twitter is a ‘blogging service’, or more specifically a ‘micro-blogging service’ with 328 million active users worldwide at the second quarter of 2017 (Statista, 2017). Registered users of these sites can post publicly visible messages – so called “tweets” - on their websites. These “tweets” were originally simple 140-character text messages (and has recently increased this to 280 characters), but can now feature photos, videos, hashtags, Vines, and so forth. Users can subscribe to other users, and automatically receive what they post. In addition, they can follow specific topics using ‘hashtags’ (#), by which posts belonging to a certain group or topic are ‘flagged’. To illustrate, one could use #violentextremism to get updated on tweets concerning the topic of violent extremism.

¹⁴ Al-Furqān Media (2014). *Clanging of the Swords I*. [Video File]. Retrieved on 7th of September 2017, from: https://videos.files.wordpress.com/foglzKon/ssw4_std.mp4

¹⁵ Al-Furqān Media (2014). *Clanging of the Swords II*. [Video File]. Retrieved on 7th of September 2017, from: https://videos.files.wordpress.com/bfPiBnnt/islamic-state-of-iraq-22clanging-of-the-swords-part-2e280b3_std.mp4

¹⁶ Al-Furqān Media (2014). *Clanging of the Swords III part 1*. [Video File]. Retrieved on 7th of September 2017, from: https://videos.files.wordpress.com/FymP4L4W/islamic-state-of-iraq-22clanging-of-the-swords-3-part-122_std.mp4

¹⁷ Al-Furqān Media (2014). *Clanging of the Swords III part 2*. [Video File]. Retrieved on 7th of June 2016, from: https://videos.files.wordpress.com/Li8bJpSI/islamic-state-of-iraq-22clanging-of-the-swords-3-part-222_std.mp4

¹⁸ Al-Furqān Media. (2014). *Clanging of the Swords IV*. [Video File]. Retrieved on 7th of September 2017, from: https://videos.files.wordpress.com/ix3vGg7v/islamic-state-of-iraq-and-al-shc481m-e2809cclanging-of-the-swords-part-4e280b3-en_dvd.mp4

¹⁹ Al-Furqān Media (29 July 2014). *Upon the Prophetic Methodology*. [Video File]. Retrieved on 7th of September 2017, from: http://www.liveleak.com/view?i=181_1406666485

²⁰ For an example see: Al-Hayat Media (2014). *Islamic State of Iraq and al-Shām – “Mujatweets Episode #2*. [Video File]. Retrieved on 7th of September 2017, https://videos.files.wordpress.com/8KHS3xZU/islamic-state-of-iraq-and-al-shc481m-e2809cmujatweets-episode-2e280b3_dvd.mp4

Twitter has become one of the preferred social media platforms for violent extremist organizations (Weimann, 2015). The ability to instantaneously and rapidly send small pieces of information to a potentially unlimited number of people free of charge (including links to other content and pages) has made Twitter an extremely valuable platform for violent extremist groups (Dean, 2016). Although it should be noted that Twitter has recently become more pro-active and successful in its efforts to thwart misuse of its platform (Berger & Perez, 2016).

Once again, it is ISIS that is particularly notorious for its Twitter use. They have utilized the platform for various purposes, including intimidation - directed at both distant foes as well as local enemies - as well as for indoctrination and recruitment (Klausen, 2015; Veilleux-Lepage, 2014; Vitale & Keagle, 2014). ISIS has been particularly successful in leveraging the platform's crowdsourcing function for massive outreach, allowing others to positively contribute to their campaign by adding their own content, retweeting official tweets, and using popular hashtags (Vitale & Keagle, 2014). In the past, Berger and Morgan (2015) have estimated that no fewer 46.000 twitter accounts were used by ISIS supporters. However, most of the activity can be ascribed to small number of hyperactive online supporters, numbering 500 to 2000 accounts, which have tweeted and re-tweeted content in concentrated, high volume bursts.

One tactic these users have relied on to disseminate content is so called 'hashtag hijacking' (also called 'trendjacking'): The practice of utilizing an established (and innocuous) hashtag for the purpose of disseminating one's own content or message. For instance, in 2014, ISIS supporters coopted World Cup hashtags, such as #Brazil2014 or #WC2014. Using this strategy, any Twitter user browsing these hashtags could find ISIS' tweets among regular World Cup tweets (Greene, 2015; Veilleux-Lepage, 2014; Vitale & Keagle, 2014)

To further the group's strategy of crowd-sourced dissemination, ISIS affiliates have also developed an Android application named the 'The Dawn of Glad Tiding'. It was temporarily available in the "Play Store", before Google removed the app for violating their terms of service. This app has been advertised as a way to keep up on the latest news about the group's activities. The application allowed ISIS affiliates to automatically posts tweets - approved by its main communication branch - to the Twitter accounts of application users without upsetting Twitter's spam-detection algorithms. As such, it allowed them to flood twitter with ISIS propaganda (Veilleux-Lepage, 2014). At its height, app posting activity was around 40.000 tweets in a single day (Aly, 2017; Vitale & Keagle, 2014).

The content of ISIS-related messages on Twitter has varied. Twitter has been used to circulate links to official videos (for instance on YouTube) and publish 'video stills' of horrific acts of violence as well as to provide 'Go-Pro footage' from foreign fighters on the frontline. However, it is not merely violence that is depicted. Messages have also portrayed 'the good life' inside the Caliphate. For instance, Western foreign fighters have tweeted about touristy snapshots of the local cuisine, hanging out with friends and their new houses (Briggs & Silverman, 2014; GISS, 2014; Klausen, 2015).

2.5.3. FACEBOOK: SOCIAL NETWORKING AND RECRUITMENT

Facebook falls into the 'social networking' category of social media. It is the biggest social media platform with over 2 billion monthly active users worldwide in the second quarter of 2017

(Statista, 2017). The service enables users to connect and maintain relationships with other members (family, friends and strangers if desired). Users can create public or semi-public personal profiles using their personal details, and can then post 'status updates' on their profile pages or write messages to other users (Dean, 2016; Kaplan & Haenlein, 2010; Smith et al, 2012; Weimann, 2015). People can also subscribe to groups on similar interests, for instance, on the basis of support for a particular organization (Dean, 2016).

Like earlier mentioned platforms, Facebook is also used for multiple purposes (Weimann, 2015). Besides being an important media outlet, Weimann (2014) argues Facebook has been especially important for letting violent extremists identify mainstream Islamic youth "*who may be occasional viewers of jihadist content*" and, in turn, "*link them to the more conspiratorial jihadist forums that primarily attract already hard-core jihad sympathizers*" (p. 6). One of the strategies utilized is creating Facebook interest groups based on seemingly innocent ideals: "*As member numbers for the groups increase, jihadist material can be slowly introduced by members of the organisation to the Facebook group in a way which does not directly condone or encourage jihadist actions, and thus does not constitute a violation of Facebook policy*" (Dean, 2016, p. 235).

Salafi Jihadi extremist have also utilized Facebook groups to identify and make lists of potential devotees. By examining people's profiles and tracing their activities online, Jihadi-Salafist groups can decide whom to target and how to approach each individual or group and tailor their content in a way that resonates or 'sticks'. As such they provide particularly effective tool for 'narrowcasting'. Narrowcasting involves aiming messages at highly-defined segments of the public to better compel recipients. Drawing on available information on audience members, such as their subscription to particular interest groups, demographic attributes and preferences, they cleverly adapt their profiles, images, videos, appeals and so forth to match the profile of a particular group and increase the persuasive impact (Weimann, 2014).

At the same time, platforms such as Facebook have made it far easier for sympathizers to reach out to violent extremist milieus themselves. In part, because those seeking to contact Jihadi-Salafist do not face the previous restrictions that often apply in more insular forums (Torok, 2016). The GISS (2014) indeed asserts that young people that are in a radicalization process actively search for like-minded individuals on Facebook. Moreover, they subsequently post Salafi-Jihadi content on their own pages, thereby potentially influencing their own circle of friends.

In sum, social media have presented extremists with powerful new avenues for strategic communication purposes. But how messaging through those channels in support of Salafi-Jihadi strategic narratives impacted on the process of radicalization into violent extremism? In the next section, we outline the current knowledge on this topic.

2.6. ONLINE VIOLENT RADICALIZATION

The exact nature and extent of the influence of Salafi-Jihadi narrative work on the internet and social media on processes of violent radicalization remain a matter of "*conjecture and hypothesizing*" (Aly, 2017, p. 3). To date, most studies have focused on studying violent extremist social media practices and analyzing the content of their online propaganda material, rather than exploring how people actually engage with online materials on social media during a radicalization process (ibid.). Little is known about how individuals actually experience extremist content they

find on social media as well as what impact it has on them, either in the short or long term (Davies et al., 2016; Edwards & Gribbon, 2013). What is however emphasized is that exposure to extremist narratives and messages does not explain radicalization. The process of radicalization into violence is complex, dynamic and multifaceted, and may involve a variety of influences, including, but not limited to, grievance, beliefs, and group / social dynamics (Neumann, 2015).

Several ideas on how the exposure to extremist materials on the internet and on social media may be implicated in the processes of violent radicalization have been put forward. First, it is argued that extremist material found online may produce some sort of ‘awakening’ for individuals who are becoming aware of issues around the world for the first time. For instance, Muslims in Western countries may be introduced to events and circumstances on the ground in areas such as Syria, Iraq and Palestine, which are framed by extremist groups as situations that urge revenge. For those that are already sympathizing with or supporting extremist groups, such content may serve to further harden their opinions, beliefs and attitudes (Davies et al., 2016). Related to this, exposure to extremist content, such as emotional arousing videos that show atrocities and moral violations in conflict zones, can elicit a strong sense of moral outrage, which has been described as an important trigger to engage in extremist action (Sageman, 2008). Furthermore, it is suggested when individuals are immersed in violent or graphic imagery for extended periods of time, the amplified effects may catalyze emotional desensitization (Bartlett & Reynolds, 2015; Davies et al., 2016; Pyszczynski, et al., 2006; Neumann, 2013a). For instance, Pyszczynski et al. (2006) argue that continuous exposure to death-related discourse and imagery can result in ‘mortality salience’ (i.e., existential threat as a result of reminders of one’s own mortality) which can encourage allegiance to good versus evil worldviews and may lead individuals to become more accepting of violent actions or even willing to engage in violent actions themselves.

In a recent study into the relation between exposure to (online) violent extremist settings and (self-reported) political violence, Pauwels and Schils (2016) argue that active engagement with extremist content on social media – as opposed to passive and accidental consumption - is of vital importance. They also posit that there is a correlation between offline differential associations with like-minded peers and self-activated online engagement with political violence. In a study by the Rand Corporation (Von Behr, Reding, Edwards & Gribbon, 2013) several assumptions about online radicalization in the literature were tested on the basis of primary data of 15 radicalized individuals, nine of whom were convicted under the terrorism legislation of the UK. They found empirical evidence to support the assumption that the Internet creates more opportunities to become radicalized, arguing that for all 15 cases, the Internet was a “*key source of information, communication and of propaganda for their extremist beliefs*” (p. xii). However, assumptions that the Internet accelerates the process of radicalization or that it promotes self-radicalization without any form of physical contact were not supported.

The online milieu in which individuals are exposed to propaganda appears to play an important role. Social media can facilitate the development of so called ‘echo-chambers’: Settings in which the same opinions and arguments are continuously repeated and reinforced (e.g., von Behr et al., 2013; Davies et al., 2016; Geeraerts, 2012; Neo et al., 2016; Neumann, 2013a; Pauwels et al., 2014). As a consequence, extreme voices become amplified while moderate voices die out (Davies et al., 2016; Geeraerts, 2012; Neumann, 2012, 2013; Pauwels et al., 2014). This may result

in individuals becoming more extreme in their views. Or as put by Neumann (2013a), it leads individuals to “*acquire a skewed sense of reality so that extremist attitudes and violence are no longer taboos but—rather—are seen as positive and desirable*” (p. 436).

Beyond exposure to propaganda, the interactive properties of these environments allow for the formation of interpersonal bonds and reinforce a sense of community, thereby enabling further socialization into extremist ideologies. This may, in turn, lead individuals to deepen their involvement with the extremist cause (Davies et al., 2016; Neo et al., 2016). As described by Neo et al. (2016): “*The interactive features of the online platforms create the required social environment, which aim to cultivate partisanship and facilitate the assimilation of new social conduct, routines and behaviours. Furthermore, these interactions may create a sphere of influence where violent extremist groups socialize their online members to radical worldview thereby creating a ready-made antecedent for more radical ideology*” (p. 8).

In sum, while online messages on social media, or via direct tools such as Telegram apps, that project Salafi-Jihadi narratives are in and by themselves unlikely to lead to violent radicalization, immersion in such material and online extremist settings can potentially contribute to the process. However, more research in this area is needed.

Unsurprisingly, governmental and non-governmental actors have sought to tackle online radicalizing influences. They have done so, in part, through negative measures, suppressing extremist activities for example by account suspensions, take-downs, filtering, censorship etc. (Davies et al., 2016). While it has been found that measures such as account suspensions can resort effect (see for example Berger & Perez, 2016), other authors argue that the effectiveness of such measures is at best limited (e.g., Davies et al., 2016). It is suggested that the amount of extremist content that can be found online is simply too huge to sift through, it may be difficult and expensive to implement such measures; and there are challenges when it comes to determining what exactly constitutes extremist content (Davies et al., 2016). Hence, alternative approaches have been sought. One of the strategies is to project other narratives through online and offline strategic communication efforts in order to displace or counter extremist narratives. Here the term counter-narratives comes into play.

2.7. COUNTERING THE NARRATIVE

So what are counter-narratives? As noted by Ferguson (2016), while the term counter-narrative is used frequently in academic and policy circles “*a common understanding of this relatively new lexicon has not emerged*” (p. 7). This was also noticeable in our focus groups, in which there was considerably debate about what a counter-narrative actually is (or should be and do). This is arguably related to the many different conceptualizations of narratives across disciplines. In line with the above given conceptualization of narratives, we found the description of Grossman (2015) useful. She argues that a counter-narrative essentially refers to a “*variety positional or relational discourse, at once overtly constructed and implicitly normative, that seeks to disrupt, dismantle, or speak back to other narrative trajectories that exert discursive power*” (p. 74). They are storylines designed and nurtured to “*resist, reframe, divert, subvert, or disable other stories and other voices that vie for or already command discursive power*” (ibid.).

As such counter-narratives are always strategic narratives, they are designed and promoted with specific purpose in mind. For instance, taking the perspective of extremist groups, one could argue that they put forward a counter-narrative which actively aims to reposition those who subscribe to other narratives. For example, a narrative about the pitfalls of a democratic system, might be construed as a counter-narrative to a narrative in which democratic rule is being put forward. Within the context of countering violent extremism, however, they are constructed and projected to “*actively reposition or reclaim the allegiance of those who embrace narratives that justify or extol violence...*” (ibid.).

Alternatively, the concept counter-narrative has been used more broadly to refer to a variety of strategic communication programs or ‘messaging’ activities that serve counter-extremism purposes and may involve a variety of actors (Briggs & Feve, 2013). Such efforts can focus on different audiences across the radicalization spectrum, from upstream (i.e., few or no signs of radicalization) to ‘downstream’ (i.e., more signs of radicalization) audiences (Silverman, Stewart, Birdwell, & Amanullah, 2016). For instance, the Institute for Strategic Dialogue (ISD) has described a ‘counter-narrative spectrum’ that includes three types of messaging activities: (1) counter-narrative messaging (e.g., activities that directly or indirectly challenge extremist narratives); (2) alternative narrative messaging (e.g., activities that aim to provide a positive alternative to extremist narratives), and (3) strategic communication by the government (e.g., activities that provide insight in what the government is doing) (Briggs & Feve, 2013; see also RAN, 2015). Strictly speaking, the term counter-narrative only fits one variation of messaging activities (van Ginkel, 2015). However, counter-narrative programs often make use of a variety of formats in which these counter-narratives are cast. Such formats may regard direct messages, Facebook campaigns or public advertisements or warnings against jihadism and radicalization. They may also include more indirect, sophisticated programs or even groups and organizations that are financed to develop approaches towards specific target audiences (family members, friends, peers of radicalized youngsters) that receive support and service to help them cope with the radical ideas and behavior of their relatives/children/friends. These programs combine both real life sessions (discussion rounds, trainings) with social media and internet information.

Based on the above notions, we conceptualize counter-narratives as strategically constructed storylines that are projected and nurtured through strategic communication (or

messaging) activities with the intention to undermine the appeal of extremist narratives of violent extremist groups, either directly or indirectly. Having conceptualized counter-narratives, we now examine the different forms of counter-narrative activities that have been suggested by the ISD (Briggs & Feve, 2013) more closely.

2.7.1. COUNTER-NARRATIVES

The first category, counter-narrative messaging, involves reactive strategic communication efforts that “*directly or indirectly challenge extremist narratives either through ideology, logic, fact or humour*”. These programs are targeted towards audiences that are already further along the path of radicalization, ranging from sympathizers, passive supporters to those more or less active within extremist groups (Ibid.). They may aim to achieve a number of aims, including “*de-radicalization of those already radicalized to the seeds of doubt among ‘at risk’ audiences potentially being exposed to or seeking out extremist content*” (Briggs & Feve, 2013, p. 16). They involve more targeted and tailored efforts, including person-to-person interventions. Such narratives may be projected by many different actors. It is suggested that some may be better positioned to project a particular counter-narrative than others. For instance, religious scholars could potentially be well-positioned - if they have religious authority and credibility to project a ‘religious’ or ‘ideological’ counter-narrative, whereas this may be difficult for the government (we will explore this difficult position further in Chapter 4). One example of a counter-narrative messaging effort offered in the literature is the ‘Say No To Terror’ campaign (see Box 1; other examples include “Moonshot”²¹, perhaps “Je suis Charlie”, if we consider this a counter-narrative, it certainly became influential²², same for “Je suis Ahmed”, the policeman who died in the streets of Paris in front of the Charlie Hebdo office²³).

Box 1: ‘Say No To Terror’ campaign

The ‘Say No To Terror’ campaign provides an example of an online counter-narrative. It comprises a dedicated website²⁴ that provides links to The ‘Say No To Terror’ campaign (Twitter²⁵, Facebook²⁶, YouTube²⁷; Weimann, 2015). The website, which is written entirely in Arabic, hosts information content as well as videos, posters, and forums. As outlined by Aly,

²¹ Moonshot cve. (c2017). *Moonshot CVE. Countering violent extremism through data driven innovation*. Retrieved on 7 September, 2017, from <http://www.moonshotcve.com>

²² Devichand, M. (3 Januari 2016). *How the world was changed by the slogan 'Je Suis Charlie'*. Retrieved on 7 September, 2017, from <http://www.bbc.com/news/blogs-trending-35108339>

²³ Bentley et al. (8 January 2015). *'He died defending the right to ridicule his faith': France unites behind #JeSuisAhmed on Twitter in tribute to Muslim officer slain by fanatics as he begged for his life*. Retrieved 7 September, 2017, from <http://www.dailymail.co.uk/news/article-2901681/Hero-police-officer-executed-street-married-42-year-old-Muslim-assigned-patrol-Paris-neighbourhood-Charlie-Hebdo-offices-located.html>

²⁴ Say No To Terror. (ca. 2010). Retrieved on 23th of June 2017, from: <http://www.sntt.me>

²⁵ Say No To Terror. (March, 2011). [Twitter timeline page]. Retrieved on 23th of June 2017, from: <https://twitter.com/SayNoToTerror>

²⁶ Say No To Terror. (ca. 2010). [Facebook timeline page]. Retrieved on 23th of June 2017, from: <https://www.facebook.com/saynototerror>

²⁷ Say No To Terror. (ca. 2010). [Youtube channel]. Retrieved on 23th of June 2017, from: <https://www.youtube.com/user/saynototerror>

Weimann-Saks and Weimann (2014), it presents a counter campaign that aims to challenge the validity of terrorism, the legitimization of violent jihad, as well as the authority of the leaders of Jihadi extremist groups. It aims to do so by communicating incentives and benefits for resisting terrorism that appeal to both the cultural and religious values of the target audience. In contrast to the Al-Qaeda narrative that imposes on Muslims a religiously sanctioned moral duty to wage a violent Jihad, the ‘Say No To Terror’ campaign draws on Islamic cultural history to impose a religiously sanctioned moral obligation to protect themselves and their communities from extremism and terrorism. The salient themes in the content of campaign comprise a narrative that emphasizes the criminal nature of (witting and unwitting support for) Jihadi terrorist and extremist groups, highlights the hypocrisy of their narrative (among other things, by highlighting Muslim victimization terrorism), and depicts the terrorist leaders as liars and manipulators, whilst, at the same time, reinforcing traditional Islamic leadership models²⁸. As astroturfing is being utilized, the origins and motives of the campaign are unclear. However, according to Aly et al. (2014), content and distribution strategies suggest the source is either affiliated with or sympathetic to the Saudi Arabian government.

2.7.2. ALTERNATIVE NARRATIVES

The second category is ‘alternative narratives’ and constitutes online strategic communication efforts that aim to undercut violent extremist narratives by focusing on “what we are for” rather than “what we are against” (Schmid, 2014, January). As put by Briggs & Feve (2013), these efforts attempt “*to influence those who might be sympathetic towards (but not actively supportive of) extremist causes, or help to unite the silent majority against extremism by emphasising solidarity, common causes and shared values*”. The main goal is to project a positive storyline, which highlights values such as democracy, freedom, rule of law, equality and respect for human rights (Dafnos, 2014). Unlike counter-narratives, they are not explicitly intended to directly address extremist narratives, although they can have secondary effects, which may undercut or displace them (Beutel et al., 2016). Like counter-narratives, these narratives may potentially be projected by many different actors. The campaign ‘Dare to be Grey’ provides an example of an alternative narrative projected by students (see Box 2).

Box 2: ‘Dare to be Grey’

A recent Dutch example of an alternative narrative campaign is ‘Dare To Be Grey’. This initiative aims to amplify the voice to grey ‘middle ground’, which is “*being drowned out by the extreme voices of today*”, and aims to “*put a stop to the polarization that is dividing [Dutch] society through promoting an open debate*” (Dare to be Grey, 2017). It facilitates an online platform against polarization through our different social media channel, including Facebook²⁹,

²⁸ For an English subtitled example, see: Saynototerror. (2 April, 2015). Say No To Terror English subtitles [video file]. Retrieved on 20th of January 2017, from <https://www.youtube.com/watch?v=08oLPfCVuS0>

²⁹ Dare to be Grey. (2016). [Facebook timeline page]. Retrieved on 20th of January 2017, <https://www.facebook.com/DaretobeGrey>

YouTube³⁰, and Instagram³¹ and a dedicated website³². They intend to raise awareness by disseminating videos, images, photos and online articles as well as offering a platform for people on which people can tell their 'grey' story. Their campaign also has offline components, Dare to be Grey organizes regular events. They are also developing an educational program and magazine. According to one of those involved who took part in one of our focus group, rather than directly challenging extremist narratives, they wanted to encourage and inspire the grey mass (on all sides of the political spectrum) to put forward a strong alternative message that can compete with those on the fringes. The initiative was developed by a group of students from the University of Utrecht, who took part in the 'Peer 2 Peer: Challenging Extremism'³³ contest, a competition that is co-sponsored by the U.S. Department of State and Facebook. The program encourages college students from universities from all around the world to develop an online social media campaign and digital products that contribute to countering extremism (EdVenture Partners, 2017).

2.7.3. GOVERNMENT STRATEGIC COMMUNICATION

The third and final category is government strategic communication, which is, as the term implies, inherently the domain of the government. It involves strategic communication efforts that ensure that government policies and positions are clearly articulated and directed towards the appropriate audiences that help amplify attempts of the government to build relationships with key constituencies, and, occasionally counter misinformation about government policies and actions. For instance, this can be achieved through public awareness activities (Briggs & Feve, 2013; van Ginkel, 2015). A recent government campaign in France 'Stop Jihadism' is one example (see Box 3).

Box 3: 'Stop Jihadism'

One example of a government strategic communication initiative is the #StopDhjihadisme³⁴ campaign that was launched by the French Government shortly after the January 2015 terrorist attacks on the office of the satirical magazine Charlie Hebdo and a kosher supermarket in Paris. Besides sharing information on Twitter³⁵ and Facebook³⁶ with the aim to

³⁰ Dare to be Grey. (2016). [Twitter timeline page]. Retrieved on 20th of January 2017, from: <https://twitter.com/DaretobeGrey>

³¹ Dare to be Grey. (2016). [YouTube channel]. Retrieved on 20th of January 2017, from: <https://www.youtube.com/channel/UCzcVDSSF1QB8If9FTkVWRog>

³² Dare to be Grey. (c2017). #Dare to be Grey. Retrieved on 7th of September, 2017, from www.dtbg.nl/en

³³ EdVenture Partners. (c2017). Peer to peer. Retrieved on 7th of September, 2017, from <https://edventurepartners.com/peer2peer>

³⁴ Stop-Djihadisme. (c2017). *Stop-Djihadisme*. Retrieved on 7 September 2017, from <http://www.stop-djihadisme.gouv.fr>

³⁵ Stopdjihadisme. (January, 2015). [Twitter timeline page]. Retrieved on 7th of September 2017, from: <https://twitter.com/stopdjihadisme>

³⁶ Stopdjihadisme. (2015). [Facebook timeline page]. Retrieved on 7th of September 2017, from: <https://www.facebook.com/Stopdjihadisme.FR>

dissuade would-be jihadists from joining Jihadi groups, they also produced and disseminated a YouTube video³⁷ that depicts the heinous reality of daily life in conquered territories by ISIS. In addition, the French government has launched an educational and informative website for the general public, and more specifically: *“those close to young people on the path of radicalization (teachers, associations) as well as young people themselves”*. According to the French Government, the website should allow citizens to *“better understand the issues and the means to combat terrorism”*. Among other things, the website presents information on how to “decrypt” propaganda and manipulation techniques used on the internet by Jihadist recruiters and displays an infographic that is supposed to help identify potential jihadists. It also conveys information about the government’s resources and actions concerning the fight against terrorism (Gouvernement.fr, 2015).

2.8. SUMMARY

In this chapter, our aim was to provide answers to research questions 1 and 2, namely (1) how can we conceptualize narratives and counter-narratives?; and (2) how are narratives and counter-narratives used via social media?

We conceptualize narratives as strategically constructed storylines that are projected and nurtured through (online) strategic communication activities by state and non-state actors in attempts to shape how target audiences feel about or understand events or issues, and ultimately, guide their behavior in a manner that is conducive to their aims and goals. When applied to Salafi-Jihadi narratives, three core elements appear to be central: (a) they project a basic grievance paired with an assignment of blame - commonly the West; (b) they offer an alternative vision of the good society - the caliphate; and, (c) and offer a promising pathway to realize this vision - violent Jihad. On the internet, we see that strategic narratives are being projected through messaging via various online channels, not only by Salafi-Jihadi groups themselves, but also by sympathizers and supporters. Specifically, we have outlined how they have been projected using YouTube, Twitter and Facebook.

In the context of countering violent extremism, we conceptualize counter-narratives as strategically constructed storylines that are projected and nurtured through strategic communication (or messaging) activities with the intention to undermine the appeal of extremist narratives of violent extremist groups, either directly or indirectly. As we have illustrated, on social media, such activities can be categorized in three domains: (1) counter-messaging (e.g., activities that challenge extremist narratives head on); (2) alternative messaging (e.g., activities that aim provide a positive alternative to extremist narratives); and (3) strategic communication by the government (e.g., activities that provide insight in what the government is doing).

³⁷ Gouvernement.fr. (28 January, 2015). *Stop-Djihadisme: Ils te disent...* [Video File]. Retrieved on 7th of September 2017, from: <https://www.youtube.com/watch?v=ke3i9-7kkQM>

3. DE-RADICALIZATION THROUGH ONLINE COUNTER-NARRATIVES CAMPAIGNS?

3.1. INTRODUCTION

In the previous chapter, we have described the process of Jihadi-radicalization and how online narratives can play a role in this process. As such, we have dealt with research question (1) How can we conceptualize narratives and counter-narratives?; and research question (2) How are narratives and counter-narratives used via social media? We now turn to research question (3), namely “to what extent is it possible to use counter/alternative-narrative programs via social media to prevent radicalization or de-radicalize individuals?”. We start by presenting a model of the radicalization process (Section 3.2) and discuss the theories that articulate how people are less likely to be influenced by counter-narrative attempts to the extent that they identify strongly with their group and their cause (Sections 3.3 to 3.7). In Section 3.8, we summarize the main conclusions that we draw from this analysis.

3.2. THE NARROW FOCUS OF A RADICAL PERSON

Many models of the radicalization process articulate certain phases, from moderate to extreme. For example, Moghaddam (2005) has introduced a staircase model to terrorism that describes people going from the group floor with grievances, to the fifth and final floor, in which they commit an act of terrorism. In line with this model, Doosje et al. (2016) present a process model of radicalization that involves three steps: (1) a sensitivity phase, (2) a group membership phase and (3) an action phase (see Figure 1).

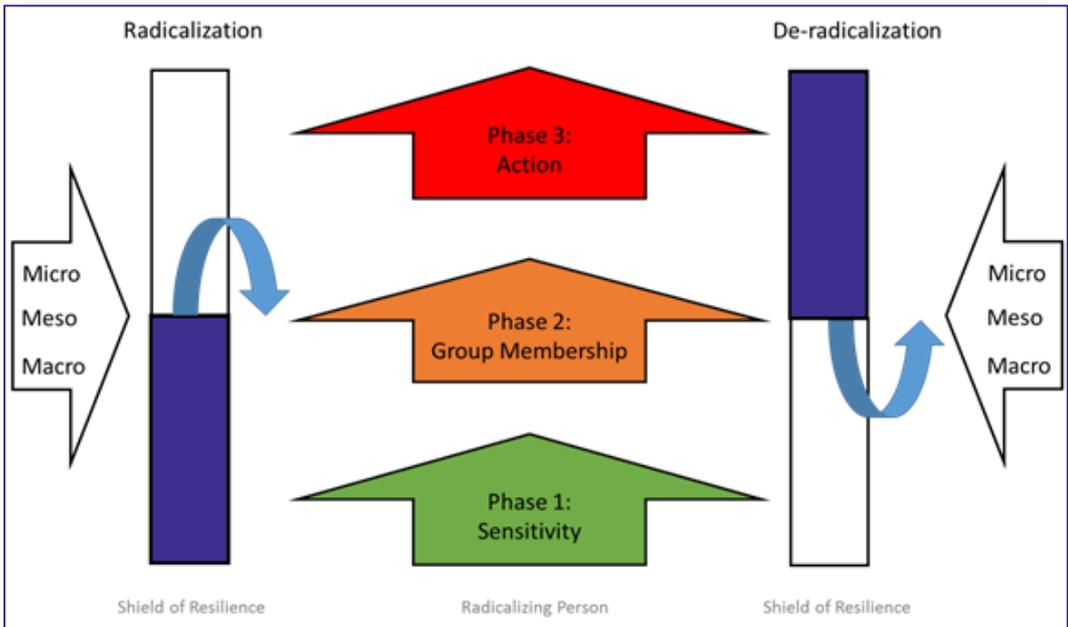


Figure 1: The (de)radicalization process and its determinants. From Doosje et al. (2016).

According to this model, the majority of people have a shield of resilience against extremist influences (left side of the figure), but there are micro (personal level) factors, meso (group level) factors and macro (societal level) factors, that may threaten the resilience shield and make people vulnerable for extremist influences. At this side of the model, prevention efforts might strengthen the shield of resilience, making people more likely to be able to resist persuasive attempts by radical groups. These prevention efforts might include presenting people with counter-narratives.

Equally important for the current question about the potential use of counter-narratives, according to this model, it is shown that members of extremist groups have another shield of resilience, namely against de-radicalization forces (right side of the figure). Generally speaking, this shield enables them to resist any influencing attempts from moderate sources. This makes it less likely that counter-messaging efforts or any other de-radicalizing forces from outside their group will have effect, to the extent that the group has become really important to the individual and has prepared the individual to counter-argue the persuasion attempts. This argument is in line with various literatures, as well as our interviews and focus groups.

However, it should be noted that, in some cases and under some circumstances, this shield of resilience of a radical person may fall apart, allowing a process of de-radicalization to start (i.e., rejection of attitudes and beliefs that underlie support for a violent ideology). This is usually preceded by a process of 'disengagement', which is characterized by a change in behavior (i.e., stop using violence and leaving the radical group). Yet, it certainly does not ensure it. There are many examples of people who are disengaged, but not de-radicalized (e.g., see Horgan, 2008). Indeed, disengagement without de-radicalization appears to be the rule rather than the exception (Schmid, 2013).

Important for the current discussion, Doosje et al. (2016) also argue that a (radical) person's focus becomes narrower with each consecutive step in the radicalization process. More specifically, it is predicted that to the extent that a person identifies strongly with the group (and the group with the individual), there are strong forces to become a strong and cohesive group. As such, this group will be motivated to isolate itself from outside influences that may undermine their narrative.

In the subsequent sections, we first describe five influential theories that predict that when people reach a certain level of extremity, they are less likely to be influenced by counter-messaging efforts. In addition to these influential theories, in Box 4, we present the main comments about this issue from our interviews with experts and our focus groups.

3.3. THEORY 1: COGNITIVE DISSONANCE THEORY

The first theory to back up our argument that counter-narratives programs may face substantial challenges when aiming to influence those that already committed to extremist groups and the views they espouse, is the theory of cognitive dissonance (CDT; Festinger, 1957).

In essence, CDT postulates that individuals seek consistency among their cognitions. Cognitions can be thought of as elements of knowledge about behaviors, perceptions, attitudes, beliefs, and feelings (Harmon-Jones & Mills, 1999). When two or more cognitions that are relevant

to each other are inconsistent, individuals will experience an unpleasant mental tension or discomfort, called cognitive dissonance. To use a classic example of Festinger (1957), knowing that smoking is detrimental for your health is likely to cause dissonance if a person smokes. CDT asserts that people have a motivational drive to reduce dissonance, as well as a tendency to avoid situations and information which are likely to increase it.

When dissonance reaches a sufficient level, people will engage in dissonance-reduction work. The amount of dissonance is determined by the number of dissonant and consonant elements, and the importance of those elements. To reduce dissonance individuals can change one of the dissonant cognitions, add consonant or remove dissonant cognitions to reduce the overall level of inconsistency, or decrease the importance of the elements involved in the inconsistency (Harmon-Jones & Harmon-Jones, 2007; Simon, Greenberg & Brehm, 1995). Typically, dissonance-reduction efforts focus on the cognitions that are less resistant to change.

To illustrate using the example of the smoker, people could stop smoking, which would indicate a change their behavior (consonant with the belief that smoking is bad for health). Yet, it may be easier to change their cognitions about the effects of smoking, for instance by seeking out information that questions the harmful effects (reducing dissonant cognitions). Alternatively, individuals could seek for the positive effects of smoking (adding consonant cognitions). The smokers may also come to adopt the position that the dangers of smoking are negligible in comparison to the joy they get from it (altering the importance of cognitions) (Harmon-Jones & Mills, 1999).

Individuals often experience dissonance when they do something that conflicts with their prior attitudes or beliefs. When there is no obvious external cause for that behavior, people have the tendency to adjust their attitudes and beliefs to correspond more closely with their behavior. Even when they know this behavior is wrong. This is because knowledge about behavior is usually quite resistant to change (Harmon-Jones & Mills, 1999; Harmon-Jones & Harmon-Jones, 2007). As noted by Harmon-Jones & Harmon-Jones (2007) “...if a person behaved in a certain way, it is often very difficult to undo that behavior” (p. 8). In a similar vein, people often start to believe what they say (de Wolf & Doosje, 2010). Applying this to radicalization, the more often people make statements that are more extreme than their actual views, the more they start to believe in those statements (Veldhuis & Staun, 2009). De Wolf and Doosje (2010) explain that this process is often leveraged by radical groups when grooming new recruits. New members are encouraged to voluntarily and publicly make ‘moderate’ statements and promises in support of the radical group and its ideology. When these requests are honored, they align their attitudes and beliefs with their actions, which make them more amendable for future, more significant requests. This mechanism has been referred to as the ‘foot-in-the-door-principle’.

Additionally, CDT proposes that people experience dissonance when they engage in an unpleasant activity to obtain some desired outcome. The dissonance will be greater, the greater the unpleasant effort required (Harmon-Jones & Mills, 1999). One way to reduce the dissonance is by evaluating the reasons for engaging in that activity more positively (i.e., add consonant cognitions) (Harmon-Jones, Harmon-Jones & Levy, 2015). People who want to become part of an extremist group typically need make major sacrifices (e.g., in money, time, behaviour, energy, socially), by demanding these investments from (potential) members such groups discourage ‘free

entry' and 'free exit' (Demant, Sloodman, Buijs, & Tillie, 2008). For example, new recruits are often encouraged to 'break bridges' with family and friends and adopt an ideology that may conflict with their original beliefs and attitudes (De Wolf & Doosje, 2010; Doosje et al., 2016). CDT suggests that the more people (have to) sacrifice, the more they will believe that these sacrifices were worth it. This will lead radicalizing people to become even more committed to their extremist group and views (De Wolf & Doosje, 2010; Veldhuis & Staun, 2009). After all, in case of a change of opinion or disengagement from the group, the costs would all have been for nothing (Demant et al., 2008).

Particularly important for our current undertaking, CDT suggests that people will generally be motivated to resist influence attempts when they lead to incongruent or conflicting cognitions (Tormala, 2008). Arguably, this may hold especially for those that have committed themselves to the extremist groups and ideology, because the actions of those individuals and their investments in justification and rationalization will significantly increase the cognitive-dissonance costs of deviation from the group and their way of thinking. Dalgaard-Nielsen (2013) suggests that 'hardcore' members, who are part of the inner core of extremist networks or groups who have made substantial investments in the name of their commitments will be particularly reluctant to go through these cognitive revisions: "...*having committed crimes, served time in prison, broken with friends or relatives, and/or submitted to various forms of hardship in the name of their extremist commitment, [extremists] will be highly resistant to embark on the supposedly rather fundamental cognitive revisions required, if they were to admit to themselves that they had been misguided*" (p. 107). She therefore argues that "*as a point of departure, we should expect that external attempts at influencing attitudes in order to promote exit will face substantial challenges*" (ibid.).

Importantly, arousing cognitive dissonance, for instance through counter-narrative efforts, can be a strategy to get people to reconsider their positions. However, it should be noted that it may be quite difficult to effectively expose committed extremists to such dissonance arousing information. Dissonance research indicates that people are more inclined to examine information that confirms their positions, and often avoid information that is contradictory (Harmon-Jones, 2012), phenomena termed respectively 'selective exposure' and 'selective avoidance' (Fransen, Smit, & Verlegh, 2015). Evidence suggest that this bias may be more pronounced when people hold strong attitudes towards an issue (e.g., Brannon, Tagler, & Eagly, 2007) and when they attach personal importance to attitudes (i.e., a strength-related attitude feature; see Howe & Krosnick (2017) for a review of theory and evidence regarding attitude importance).

Also, one should be careful, as the outcome of such dissonance arousing efforts may well lead to results that were not intended. Messaging efforts that aim to influence attitudes by introducing dissonance arousing information, can instigate a process that actually reduces the cognitive dissonance. If the information presented in the message manages to trigger sufficient cognitive dissonance, but is not persuasive or convincing enough to change a person's views or attitudes, this may lead the person to intensify his original attitudes and behavior - a phenomenon also referred to as a 'boomerang effect' (Jervis, 2017).

In sum, CDT suggests that those undertaking counter-narrative efforts will have a hard time influencing those who have invested heavily in extremists groups and ideology. Not only will they be reluctant to expose themselves to information that may arouse dissonance, they will be

highly resistant to embark on fundamental cognitive revisions. Ultimately, according to this cognitive dissonance perspective, such efforts may even back-fire as they may serve to strengthen the original beliefs and attitudes.

3.4. THEORY 2: PSYCHOLOGICAL REACTANCE THEORY

The second theory, Psychological reactance theory (PRT; Brehm, 1966; Brehm & Brehm, 1981) is a well-known framework for understanding why people may be inclined to resist influence attempts. While the theory may be more general, it is of no less relevance for our specific endeavor. PRT is based on the assumption that people generally value their perceived freedom to act, think and feel as they want. In short, the theory proposes that when individuals perceive a third party is threatening or constraining their freedom, they experience 'reactance'. Reactance is an unpleasant motivational state, comprised of anger and negative cognitions (e.g., counter arguing), which motivates individuals to engage in cognitive and behavioral efforts to reassert their freedom (Steindl, Sittenthaler, Traut-Mattausch & Greenberg, 2015). To provide what may be a recognizable example, when parents tell a child not to wear particular cloths to school, but the child beliefs that he or she is free to decide what to wear. The child is likely to experience reactance and will be motivated to restore his or her freedom. In consequence, he or she may decide to wear the cloths to school anyway (Miron & Brehm, 2006).

With respect to persuasive communication, the theory assumes persuasive messages may be perceived as a threat to behavioral or attitudinal freedom, and, as such, may arouse reactance. For example, on the attitudinal side, people may want to feel free to hold a particular position on issues (e.g., pro or con), alter their position, or not adopt a position. Any persuasive message that attempts individuals to adopt certain positions may be construed as threats to such attitudinal freedoms (Eagly & Chaiken, 1984). Once reactance is aroused, individuals will be driven to restore their freedom, which may cause a 'boomerang effect' similar to the one described before. Namely, individuals may change their attitudes in an opposite direction of the advocated position (Burgoon, Alvaro, Grandpre, & Voulodakis, 2002; Miller, Burgoon, Grandpre, & Alvaro, 2006). Other freedom-restoring responses that have been suggested include observing others restoring the threatened freedom, subjectively increasing the attractiveness of the threatened freedom (and decreasing the attractiveness of the alternative option) (Miron & Brehm, 2006). A particularly damaging response, may be source derogating, which may be accompanied by hostility or aggression towards the source. This is because it will undermine the reference power and credibility of the influencing agent, which may have a long term effect, as it can lead to diminished influence over the long term (Miller, Lane, Deatrck, Young, & Potts, 2007).

The magnitude of experienced reactance depends on the importance of the threatened freedom, the proportion of freedoms threatened and the perceived magnitude of the threat (Burgoon et al., 2002). Furthermore, it has been found that reactance, may differ according to the traits of the target group. Arguably relevant for our current endeavor, research indicates that adolescents and young adults are likely to be particularly reactance prone, and are more strongly inclined to reject persuasive messages perceived as attempting to control them in any kind of matter (Hong, Giannakopoulos, Laing, & Williams, 1994). Grandpre, Alvaro, Burgoon, Miller, and Hall (2003) suggest that this may be attributed to characteristics related to transitional stage of adolescence.

In addition, message characteristics have been found to affect the magnitude of reactance. Generally speaking, all aspects of a message that overtly force a certain behavior or attitude upon someone (i.e., that obviously show persuasive intent) may arouse reactance; Franssen, Smith & Verlegh, 2015). Studies on the use of language indicate that messages using explicit, forceful, dogmatic, controlling, and threat-to-choice language are more likely to be perceived as threatening freedom, and will increase the amount of reactance, often resulting in an undesirable boomerang effect (Dillard & Shen, 2005; Miller et al., 2007; Quick & Considine, 2008; Quick & Stephenson, 2008). For example, Dillard and Shen (2005) found that phrases such as “No other conclusion makes any sense” and “There is a problem and you have to be part of the solution” will arouse reactance. Also, studies indicate that messages using imperatives such as you “ought to”, “need to”, “should” etc., can increase perceptions of a threat to freedom (Steindl et al., 2015).

Taking the above in account, arguably, the main takeaway from reactance theory is that counter-narrative messages - especially those that overtly and forcefully attempt to challenge (or ‘counter’) extremist attitudes and behaviors and encourage others - may well lead people to engage in freedom-restoring responses. Not only may it lead people to reject the message, it may well cause a boomerang effect.

3.5. THEORY 3: SOCIAL JUDGEMENT THEORY

The third theory that predicts little effect of a counter-messaging among already radicalized people, is the Social Judgment Theory (SJT); (SJT; Sherif & Hovland, 1961; Sherif, Sherif, & Nebergall, 1965). The central tenet of this theory is that the effectiveness of persuasive communication efforts will depend upon the way in which the receiver evaluates a persuasive message (O’Keefe, 2015). They will not do so purely on the merits of what is advocated in the message, rather people assess the advocated position and compare it with their own attitude. Subsequently, they will determine whether they should accept what is advocated (Perloff, 2010).

Of course, on any given issue there are likely to be a variety of positions or points view available. Take for instance the issue of ‘abortion’. One may hold that abortion should be illegal or alternatively think that women should be permitted to have an abortion whenever they request it. One may also hold a variety of intermediate positions that vary in the amount of restrictions. A person is likely to assess various positions differently. One may have one preferred position, but find others more or less acceptable or tolerable. There may also be positions one finds objectionable or is ambivalent about (O’Keefe, 2015). SJT aims to specify these reactions by proposing that people have different ‘latitudes’ (i.e., ‘judgmental categories’) that exist on a continuum. The ‘latitude of acceptance’ consists of the range of positions that a person finds acceptable, including his or her most preferred position (the ‘anchor’). The ‘latitude of rejection’ contains a range of positions with which the person finds objectionable, including the position that one finds most objectionable. In between there is a ‘latitude of non-commitment’, a range of positions to which the person has no commitment or is indifferent (Perloff, 2010). With respect to persuasive communication, whenever a position that a message advocates is judged to fall within one’s latitude of acceptance or non-commitment, attitude change in line with that position is likely to occur. An advocated position that is perceived to fall in the latitude of rejection yields little or no effect, or may lead to attitude change away from the advocated position (i.e., ‘boomerang effect’) (O’Keefe, 2015).

The SJT further proposes that somewhat ambiguous messages that advocate positions that belong to a person's latitude of rejection may be contrasted. This means that they are perceived as even further away from a person's own position (i.e., 'contrast effect'). Here, the 'anchor' serves as the main reference point. In similar fashion, when the advocated view is in the latitude of acceptance, the message may be perceived as closer to the person's own position (i.e., 'assimilation effect'). In the latitude of non-commitment, either the assimilation or contrast effect may occur. The location of the boundary point at which assimilation effects stop and contrast effects begin is not clear (O'Keefe, 2015). However, O'Keefe (2015) suggests that this may be closer to the latitude of rejection than the latitude of acceptance.

SJT suggests that attitude change is likely to occur when there is a certain discrepancy between the position advocated in the message and the receiver's position. However, this is true up until the point the message falls in the latitude of rejection. Importantly though, influence the structure of a person's latitudes on an issue is presumed to be influenced by 'ego-involvement'. Ego-involvement is often understood as the degree to which an individual is 'involved in an issue'³⁸, which often goes hand in hand with holding an extreme position on an issue. When people are highly involved in an issue, their latitude of rejection increases, which results in a decrease of the size of the latitude of acceptance and non-commitment (O'Keefe, 2015). Hence, it follows from this that they will find fewer positions acceptable. Highly involved people will also contrast even mildly disagreeable positions more frequently. Furthermore, they will only assimilate ambiguous messages when the advocated positions are generally consistent with their attitudes. As such, people that are highly involved in an issue are generally harder to persuade (Perloff, 2010).

When involvement in an issue is very high, people's thinking on issues can take on an extreme, black-or-white quality, in which there may effectively be only two judgmental categories. There is no middle ground, things are either good or bad (O'Keefe, 2015). People who go through a process of radicalization may come to adopt a black-and-white, absolutist worldview that structures their thinking on whole range of matters. This process is stimulated by socialization in a radical group (e.g., through mechanisms of indoctrination and isolation) (de Wolf & Doosje, 2010). SJT suggest that these individuals will be highly resistant to just about any message that is not adjacent to their own view on the issue. Those undertaking an influence attempt may be able to advocate safely a position that is only moderately discrepant with the person's preconceived attitude and falls within the latitude acceptance. When a message ends up in the latitude of rejection, it will be contrasted, and is perceived to differ more sharply from one's attitude than it actually is (ibid.)

Once more, the implication is that those undertaking counter-narrative efforts need to be careful when targeting these individuals, as they have very little wiggle room. Given that the theory suggests that advocated positions cannot veer too much from their pre-existing positions, ideally, one has a clear picture these individuals' latitudes of acceptance and rejection before one

³⁸ According to O'Keefe (2015) who provides a review of the concept of ego-involvement, a person can be said to be involved with an issue "when the issue has personal significance to the individual, when the person's stand on the issue is central to his or her sense of self (hence ego-involvement), when the issue is important to the person, when the person takes a strong stand on the issue, when the person is strongly committed to the position, and so forth" (p. 22).

undertakes influence attempts. The theory raises substantial doubts as to whether influence efforts that aim offer a direct rebuttal of rigidly held positions can be successful. According to the theory, in all likelihood, messages that do so are rejected outright.

3.6. THEORY 4: SOCIAL IDENTITY APPROACH

The fourth theory that can explain why it is difficult to de-radicalize radical individuals (via on-line counter-narratives or in any other manner) is the social identity approach. This approach refers to two related theories, namely Social Identity Theory (SIT; Tajfel & Turner, 1979; 1986) and Self-Categorization Theory (SCT; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). Before we address the relevance for this particular research question, we first provide a sketch of the basic tenets of this approach (for a broader outline see Hornsey, 2008), because it articulates the importance of group membership, which is highly relevant in the context of radicalization and de-radicalization (Doosje et al., 2016).

The social identity approach asserts that the (social) identities that we derive from our group memberships form an important part of our self-concept. Social identity refers to a definition of the self in terms of a shared social category based on intergroup comparisons. The concept of social identity can be distinguished from personal identity, which refers to definitions of self in terms idiosyncratic attributes and interpersonal relationships (Baray, Postmes & Jetten, 2009). The social identity perspective argues that our cognitive representation of self may vary depending on whether the personal identity or social identities are made psychologically salient. When a particular social identity is salient, it becomes the basis for self-conception, in turn influencing how we think (attitudes), feel (emotions) and act (behaviour; Ellemers, Spears & Doosje, 2002).

SIT postulates that the salience of a particular group membership will induce a motivation to positively differentiate that group (i.e., the in-group) from other relevant groups (i.e., outgroups) on valued dimensions of comparison (Reicher, Spears & Haslam, 2010). This motivation is believed to be underpinned by a variety of factors such as the need to establish and maintain a positive identity (Tajfel & Turner, 1979), reduce subjective uncertainty (Hogg, 2000) and achieve optimal distinctiveness (Brewer, 1991). This results in a tendency to favour one's own group over other groups (i.e., in-group bias) which can be reflected in attitudes, emotions and behaviors.

The degree of identification with the in-group is assumed to moderate this tendency. One important prediction derived from this approach is that high identifiers are more inclined to selectively seek out favorable information about the in-group in order to bolster their self-concept. Similarly, they tend to avoid negative information of their group that could threaten their favourable sense of self (Appiah, Knobloch-Westerwick, & Alter, 2013; Knobloch-Westerwick & Hastall, 2010). Or as put by Knobloch-Westerwick and Hastall, *"individuals seek out particular messages that enhance their social identities, essentially by providing favorable social comparisons with out-groups"* (p. 518).

Furthermore, SCT argues that when social identity is salient, people experience a cognitive shift such that perceptions of self and others will become group-based, eliciting a tendency to conform to in-group norms. More specifically, it is argued that, when a social identity

becomes salient, people come to perceive themselves and others less as unique individuals, and more in terms of our group memberships. When this occurs, there is an assimilation of the self to an in-group prototype (i.e., a process termed depersonalization; Smith & Hogg, 2008). A prototype may simply be understood as cognitive representation of group norms (Moran & Sussman, 2014). When a group prototype becomes an internalized part of our self-concept through assimilation, we assign the in-group norm to ourselves, and tend to think, feel, and behave accordingly. We do so even in the absence of surveillance of other group members. In addition, we will judge others in terms of their conformity to the in-group prototype (Smith & Hogg, 2008)³⁹.

Social identity theorists argue that such norm-based influence may have a considerable impact on attitude change attempts. The key point is that when social identity is salient, the in-group norm will come to serve as a standard for validity judgements in relation to persuasive information. As such, it is predicted that in-group messages will be perceived as more valid than outgroup messages, and that people will be more influenced by in-group members than outgroup members (Smith & Hogg, 2008). When individuals identify more strongly with the group, these norms exert more influence (Moran & Sussman, 2014).

Research supports the notion that persuasive messages stemming from an in-group source are generally more effective than those stemming from an outgroup source (e.g., Mackie, Gastardo-Conaco, & Skelly, 1992; Mcgarty, Haslam, Hutchinson, & Turner, 1994). This effect is suggested to be due to people's use of source information, not only as a heuristic cue to accept the in-group's position. It is also suggested that people process in-group information in a rather systematic manner (Wyer, 2010). This holds particularly when the message pertains to issues that are group-defining or group-relevant, and when the source is a prototypical group member (Smith & Hogg, 2008). To the extent that messages are considered to be derived from an out-group source, they are typically not processed deeply (Mackie, Worth, & Asuncion, 1990). This is especially the case when out-group members articulate a critical or anti-normative message (Esposito, Hornsey & Spoor, 2013). According to Esposito, Hornsey and Spoor (2013), "*their message is likely to be rejected regardless of whether it is objectively 'right', well-considered, well-justified, or well-argued*" (p. 394).

Of course, anti-normative information can potentially also stem from someone who is perceived as an in-group member. Research suggests that when someone or a minority within the in-group advocate a position that differs from other in-group members, they are not likely to be influential. As argued by Marques, Abrams, Páez, and Hogg (2008), the "*perception of someone as being a deviant discredits and devalues them, and reduces their persuasive potential*" (p. 401). In fact, the 'black sheep effect' predicts that the other group members may well derogate and reject the deviant in-group member more strongly than out-group members. Furthermore, under certain circumstances, a boomerang effect may occur, thus, they may change their attitudes in the opposite

³⁹ Social influence can also occur under conditions when personal identity is salient. Social influence under social identity salience has been termed 'interpersonal influence' and is fostered by interpersonal bonds. When personal identity is salient, social influence may also be driven by individuals' striving for distinctiveness from their communication partners as well as fulfilment of other personal needs (Sassenberg & Jonas, 2009).

direction than that which the deviant advocates (Bazarova, Walther, & McLeod, 2012). The black sheep effect has been demonstrated in a variety of studies (Marques, Abrams, Paez, & Martinez-Taboada, 1998; Marques, Yzerbyt, & Leyens, 1988). Moreover, also been shown that the black sheep effect is stronger when identification is relatively high (Coull, Yzerbyt, Castano, Paladino, & Leemans, 2001).

Particularly important for our current endeavor, highly entitative social groups that display strong ideological premises (e.g., extremist groups) strengthen socio-psychological processes that encourage a salient group identity and enforce assimilation into prototypes and group norms (Harris, Gringart, & Drake, 2014; Harris, 2011). Harris et al. (2014) put forward that “[t]hese groups emphasise their distinctiveness and impose boundaries between themselves and the mainstream, which enhance the strong connections between members and fosters the ‘us and them’ mentality” (p. 4). The cohesive nature of these groups is argued to make members highly resistant to any form of disruptive influence (Ibid.).

Thus, the Social Identity Approach indicates that group membership is crucial in explaining how people process information: People are generally more likely to be persuaded by someone from their own group than by a member of an out-group. This has serious implications when considering the source of an online counter-narrative program: when people perceive the message stemming from an out-group, they are not easily persuaded by the content, no matter the quality of the materials and the arguments. Even when a message is considered to stem from an in-group member. To the extent that it is perceived as against the norms of the in-group, this member is usually less persuasive and in fact may fall from grace and become a black sheep.

Equally important from the current perspective, both tendencies (i.e., to discredit information stemming from an out-group source and from a deviant in-group source) are particularly strong for people who identify strongly with their group. This is due to the fact that they are more likely to perceive the world in terms of their group membership. At the same time, they are motivated to keep a “healthy distance” to out-groups and anti-normative in-group members. This effect will be especially pronounced when these are highly entitative groups with strong ideologies such as extremist groups. For these reasons, from a Social Identity Approach, presenting an online counter-narrative to these highly identified people may fail to fall on fertile soil.

3.7. THEORY 5: SIDE MODEL OF DEINDIVIDUATION EFFECTS

The fifth theory also argues that people are less likely to be persuaded when they feel connected to a group, is the Social Identity Model of Deindividuation Effects (SIDE) model. This is a well-established theory on computer-mediated-communication (CMC) grounded in the social identity approach (see previous section). We treat it separately, because this theory predicts that the effects of norm-based influence may be even more pronounced in online groups. It is argued that the scarcity of individuating cues in CMC further depersonalizes social perception of others and the self. This is believed to heighten group identification and perceived in-group homogeneity, in turn, enhancing conformity to the in-group norm (Lee, 2006; Spears, Lea, Postmes, & Wolbert,

2011)⁴⁰. That is, provided the social identity is initially salient and the group norm is known or can be inferred (Tanis & Postmes, 2003). Under these circumstances, it is argued that the in-group-norm will be more influential in CMC in comparison to face-to-face communication (Sassenberg, Boos, & Rabung, 2005).

The predictions of the SIDE model have received considerable support (for a review, see Spears, Lea, & Postmes, 2001). For instance, Postmes et al. (2001) conducted an experimental study pertaining to the conditions under which individuals conform to group norms in an online environment. Groups were requested to discuss a policy dilemma of a hospital concerning efficiency versus patient care priorities. It was observed that anonymous discussants, unwittingly primed with a particular norm, were more likely act consistent with (and socially transmit) that norm, than were non-anonymous discussants. This effect grew stronger over time.

Besides predicting that anonymous online interaction can enhance conformity to group norms, SIDE theorist have also suggested that it enhances the group polarization process that we referred to in chapter 2. Group polarization refers to a consistent finding in social psychological research that following frequent interaction within like-minded others in closed groups, people tend to endorse more extreme positions in the direction to which the group was already leaning (Lee, 2007). Some authors suggest that online settings such as those provided by social media provide ideal environments for this phenomenon to occur, as they allow individuals to pro-actively select groups to join (and which not) and seek out likeminded (and avoid dissimilar) others. Within these groups (i.e., echo chambers), participants face both normative and informational influences that move them further towards more extreme and inflexible positions. As explained by Geerearts, (2012), “[b]ecause participants in homogeneous groups share similar perspectives, they do not express opposing views. Furthermore, as a normative influence, participants might adjust their opinions to the expectations of other group members who are more extreme” (p. 26). In turn, this make people less open or even hostile to dissimilar others and counter-attitudinal viewpoints.

The SIDE model suggests that a lack of individuation information in online social interaction may further enhance this process. It assumes that group polarization is a function of identity salience and occurs as group members converge towards a polarized in-group norm (Postmes, Spears, Lee, & Novak, 2005). That is, people move toward a more extreme position in the direction of the in-group norm, even beyond the average position, in order to maximally differentiate themselves from (implicit) out-groups (Lee, 2007). The SIDE model predicts that anonymous online social interaction will foster group identification and conformity to group norms, resulting in greater group polarization (Spears et al., 2011). Several studies provide evidence in line with this prediction (e.g., Lee, 2007; Spears, Lea, & Lee, 1990). For instance, in a CMC experiment, Lee (2007) found that de-individuated group members exhibited stronger identification with their anonymous group members and were also more likely to polarize their opinions than individuated counterparts.

⁴⁰ Under conditions of personal identity salience deindividuation is assumed to obstruct interpersonal influence, as it hinders the development of interpersonal bonds that are the basis for interpersonal influence [see also the previous footnote] (Sassenberg, Boos & Rabung, 2005).

BOX 4: INTERVIEWS & FOCUS GROUPS

In addition to the literature, we also examined the answers by the experts from the interviews and the focus groups (see Appendix A for details about the method). Specifically, they were asked the following question: Is it possible to utilize counter-narratives via social media for prevention of radicalization or de-radicalization? Overall, these experts agreed that it is important to make a distinction between prevention of radicalization and de-radicalization. Most of them were convinced that using a counter-narrative to de-radicalize a person might not be very useful:

“As a general response, I would say yes, it can make a contribution, but I would see some differences between the two goals, either preventing versus de-radicalizing people, who are already prone to extremism. For de-radicalization, I would say that it’s more difficult to get them even in a social media context because extremists tend to put them into bubbles where they try to cut them from normal media influence so they won’t see any neutral messages anymore.” (Diana Rieger)

“...radical people will never even come across that [counter-narrative] message. Because, information doesn’t circulate randomly, but circulate across networks and so, simply that kind of messages doesn’t get into the networks. ... It [presenting counter-narratives to radical people] would be a complete waste of time ... it’s just a waste of resources [...] I am not saying that communication is not important, but it’s a waste of time, to put it very bluntly, to message extremists.” (Christina Archetti)

“People who are vulnerable for radicalization are generally more susceptible for certain images, for certain arguments than people who are radicalized already” (Maarten van Donk)

“And it depends, if they’re so far radicalized that they are completely isolated, it’s not gonna work” (Anne Aly)

“At the moment that the youth are seeking and they are in a process of radicalization, there are other things needed and much more tailored work and personal contact” (Marije Breedveld)

In some cases, people warn that certain counter-narrative campaigns may even backfire and produce the opposite results than intended:

“It is very difficult, to directly counter-argue. There have been campaigns, to show the disillusion. There have been campaigns that showed how it was in Raqqa. And that has been a complete failure. That has backfired, because the images of violence still were shown. So, especially by this campaign, these people saw that violence and they became radical all the same, because such a counter-campaign did provide that romantic image. It is very risky to counter this in this manner.” (Participant 4, Focus Group Amsterdam Municipality)

“I think that when you look at the final phase of the [radicalization] process, it is counter-productive ... You need to be very careful even: Aren’t we fueling this process, to make things worse?” (Michel Kok)

However, it is important to note that some experts argue that counter-narratives can play a certain role in the *prevention* of radicalization:

“So, a lot of international work has been done on countering messages that have to do with the consideration phase. Actually, everyone seems to sort of miss the point which is it surely more appropriate to get into the pre-consideration phase. Cause if you can get people at the pre-consideration phase, they are never going to get to consideration and in fact if the world counter

terrorism forces put more effort into that, I think squash ISIS' and any other form of messaging much more effectively.” (Sven Hughes)

“But preventing unwanted behavior, if you can frame it that it concerns everyone. That does not concern a specific group. ... In that case, you reach a much broader target people ... That’s the reason why I focus so much on polarization” (Abdullah Pehlivan)

“When can you be effective with your interventions? And that is mainly in the pre-phase [of radicalization].” (Marije Breedveld)

“I think you have to look carefully at the target group. Because for whom are these videos meant? These videos are meant for youth that experiences certain relative deprivation in society. But it is also possible to look before that happens. So it’s possible to have various strategies. And then it can certainly work.” (Arjan de Wolf)

Thus, these experts argue that it is possible to use online counter-narrative campaigns in a very early phase of radicalization. However, once people have adopted a more radical position, counter-narratives may not be effective at all, and they can run the risk of being counter-productive, in the sense that they can fuel radicalization.

3.8. SUMMARY

In this chapter, we have addressed research question (3), namely “to what extent is it possible to use counter-narrative programs via social media to de-radicalize individuals?” To summarize, from five theoretical perspectives (cognitive dissonance theory, social judgement theory, psychological reactance theory, social identity approach and the SIDE-model), as well as our expert interviews and focus groups, we argue that to the extent that people have become an adherent of a violent extremist group and its ideology, there are less likely to be influenced by (online) counter-messaging efforts. This is especially the case when they follow a strategy of rebuttal. They may be hard to reach as well as difficult to move. In fact, they may adopt positions in the other direction than intended as a reaction to this persuasive attempt (i.e., a boomerang effect). Thus, counter-messaging campaigns may not be a very effective tool to de-radicalize those that are already further along the path of radicalization.

Having articulated the clear limitations of any on-line counter-narrative campaign in terms of reaching those already further along the path of radicalization, we argue that, if any success is to be expected at all from such efforts, the payoff is likely to be more significant among those who show some curiosity or sympathy for extremist groups and the views they espouse, but are not yet staunch supporters (i.e., when they still are in the sensitivity phase of the model by Doosje et al., 2016).

In addition, despite the fact that there are good and solid theoretical reasons (and indications by the experts we interviewed) to expect that counter-narratives may not work for (mostly highly identified) radicals, it is an interesting question whether or not such efforts may have some effect when some individuals from radical groups show a glimpse of doubt about their group and its goals. In terms of the model by Doosje et al. (2016) presented earlier, for some individuals, the shield of resilience may show signs of holes or weak spots. May a counter-narrative fall on fertile grounds for such individuals? Therefore, it is an important question how to design such a counter-narrative campaign (mainly for preventive purposes and for the individuals who

show signs of doubts about their group). Which aspects of a counter-narrative campaign need to be considered? These questions will be addressed in the next chapter.

4. WHAT MAKES A COUNTER-NARRATIVE CAMPAIGN EFFECTIVE?

4.1. INTRODUCTION

In this chapter, we address research question 4: What are the pre-requisites for a counter-narrative program for it to be effective? Thus, we examine the key principles that may guide the development and implementation of social media counter-narrative program addressing potential radicalization, given the limitations discussed in the previous chapters. These limitations include that a counter-narrative program might not be effective among those who already passed the first stage of radicalization (see Chapter 3). The current chapter will therefore address research question 4 mainly in a preventive sense. However, as argued earlier, despite the fact that members of radical groups generally speaking will not be inclined to read materials from outside their sphere, still some individuals who show signs of disengagement with the group, might be possible targets for a tailored campaign as well.

Overlooking the literature, it is clear that effect studies for counter narrative programs are sparse and it is hard to draw conclusions whether these programs can actually be effective. We will briefly discuss the state of the art below (Section 4.2). Subsequently, the rest of the current chapter will draw upon insights from the broader communication literature and supplement this with the findings provided by counter violent extremism (CVE) experts and practitioners as well as assessments of current social media initiatives aimed at addressing violent radicalization. General theories from mass communication have gradually evolved from one-directional sender-message-receiver models toward more dynamic and interactive models including feedback loops and agency of both sender and receiver. The latter do more justice to contemporary mediated communication processes including social media. In the following, we try to outline the key principles that can be derived from many of such smaller theories in perspective of a counter-narrative program. In brief, key principles are outlined in subsequent sections, and we follow the general steps in setting up a communication campaign:

1) Research and planning (including the use of theory; determine target audience, and lessons from (extremist) groups; Section 4.3).

2) Program development and design (including defining goals and objectives; assessment of available resources; development of effective messages, and selection of credible messengers; Section 4.4).

3) Implementation (including dissemination channels, disseminating the message, and risks; Section 4.5).

In a final Section (4.6), we summarize the findings from this chapter.

4.2. PRINCIPLES OF AN EFFECTIVE COUNTER-NARRATIVE CAMPAIGN?

At the outset, it should be noted that the field is still in its infancy (Briggs & Feve, 2013), and that some authors content that it is doubtful whether counter-narrative programs as they are currently conceptualized can actually be effective at all (see Davies et al., 2016; Ferguson, 2016; Hemmingsen & Castro, 2017, for relevant discussions). 'Best practices' suggested the counter

violent-extremism (CVE) literature often appear to be based on anecdotal evidence and expert opinion rather than thorough case studies and evaluations. In fact, Ferguson (2016) states that “*the absence of methodologically robust monitoring and evaluation (M&E) practices with regard to CVE counter-narratives is striking*” (p. 8.).

An underlying issue is that important foundational questions regarding online violent radicalization still need to be addressed, and that only little substantive empirical research has been undertaken to do so. As noted earlier, there is for example very limited knowledge about the relationship between the consumption of extremist content and (online) violent radicalization. Conway (2017) clearly sums up the inherent challenge: “*Even if one accepts, in the absence of sound research [...] on the basis of common sense, anecdote, and the like that violent online radicalization is actually occurring, how can effective counter strategies be developed, if we do not know **how** it’s occurring?*” [Emphasis in the original] (p. 83).

Furthermore, the measurement of outcomes and impacts of such programs is far from a straightforward matter. In the field of CVE in general, practitioners and programmers are confronted by a variety of methodological and practical challenges that makes the impact of these initiatives extremely difficult to empirically assess. This holds certainly for preventive strategic communication efforts, since, if successful, the result is per definition a ‘nonevent’. Put differently, the result is an outcome that did not occur (radicalization) (Davies et al., 2016; Romaniuk, 2015; Romaniuk & Fink, 2012; Vidino, 2010; Lindekilde, 2012). In addition, (de)radicalization processes are complex and many variables may confound outcomes. Hence, it is difficult to identify and draw a causal link between changes in attitudes and/or behaviors and counter-narrative efforts (RAN, 2015). Put differently, if any positive changes can be observed, it may be difficult to pinpoint whether they are in fact due to the effort undertaken.

Above mentioned issues make it a challenge to reach any evidence-based research findings from which valid pre-requisites for an effective CVE strategic communication programs can be drawn. However, the field of strategic communication has a long history and is not the limited domain of CVE. It is inherently multidisciplinary and draws on literature from a wide variety of subfields including, but not limited to health promotion, (social) marketing and public relations. The broader literature is rich in descriptions of practices and approaches that are *associated* with more successful communication efforts. However, it should be mentioned that programs delivered via social media have only recently caught the attention of scholars. The following will provide an overview of what can be learnt in this respect from the broader communication literature.

4.3. START A CAMPAIGN WITH RESEARCH AND PLANNING

From the literature in communication science, it is clear that in the starting phase of a campaign conducting research is essential. It is vital to develop a thorough understanding on the topic at hand, define the problem and the audience, gain insight in program strategies that may prove fruitful, develop an outline and global planning of how the program will be designed. In the following, several important aspects in this phase are described along the following subsections: use of theory (4.3.1), define and know the audience (4.3.2), learn from (former) extremist narratives (4.3.3), analyze available recourses (4.3.4.) and determine goals and objectives (4.3.5.).

4.3.1. USE OF THEORY

A first important principle that can be drawn from broader communication research is that the application of theory (or rather theories) is a *sine qua non* for the design of effective communication efforts (e.g., Rossmann, 2015; Egner, 2009; Korda & Itani, 2013; Noar, 2006; Noar, Palmgreen, Chabot, Dobransky & Zimmerman, 2009; Perloff, 2010). What distinguishes effective from ineffective communication efforts is that the former “*reflect painstaking application of theoretical principles*” whereas the latter “*are based on ‘seat of the pants’ intuitions*” (Perloff, 2010, p. 332). Indeed, one of the most comprehensive meta-analyses of online health promotion interventions demonstrated that programs based on theory generally had greater effect than those that were not (Webb et al., 2010).

Theories can inform programs in a variety of ways. For instance, communications programs that are generally more effective than others are grounded in thorough analyses that define the problem or situation to be addressed in the context in which it occurs. In addition to empirical evidence, theory can help to provide insight into the nature of a problem and its causes, determinants and mechanisms, and assists in identifying the range of factors that program developers might seek to modify through their effort. Furthermore, theories suggest ways to drive change in attitudes and behaviors and can guide program planners in the development and implementation of program elements (Green, 2000). Beyond drawing on or borrowing from existing theory, it is also argued program planners do well to articulate a sound, program specific ‘theory of change’, which specifies how a program hopes to cause a desired outcome (Waldman & Verga, 2016). Such a theory explicitly spells out “*each hypothesized logical [and theoretical] relationship[s] linking program design to eventual program effects*” (Egner, 2009, p. 331), and is, among other things, valuable for evaluation and monitoring purposes.

Sound theoretical underpinnings are important. as without a full rational appraisal of the issue or problem in a given context as well as potential solutions, one easily falls in the trap of focusing on “*wrong or inappropriate variables (i.e., miss the target completely)*” or tackling “*only a proportion of the combination of variables required to have the desired effect (i.e., hit only a few of the total number of possible targets)*” (Green, 2000, p. 126). To illustrate, a program that is designed to dissuade audiences from engaging in particular behavior, for instance smoking, is highly unlikely to succeed if it does not effectively address its known determinants. One could argue the same for a program that aims to address radicalization, whether it is focused on cognitive or behavioral radicalization.

Given the lack of solid empirical studies on both (online) radicalization and counter-radicalization processes, a sound theoretical foundation seems particularly important for strategic communication initiatives addressing violent radicalization. As argued by Davies et al., (2016), those undertaking counter-narrative efforts should at a minimum take into account what is known about the factors that are purported to animate radicalization processes. While they recognize that there is no ‘grand theory’ of radicalization and that the reasons why individuals radicalize into violent extremism are not yet fully understood, they also point out that there is growing consensus that a wide range of inter-related factors may be implicated in (violent) radicalization. Indeed, various studies point towards a complex interaction between factors at the micro, meso and macro level (see also Borum, 2011; Doosje et al., 2016; Ranstorp, 2010; Schmid, 2013 ; Vidino, 2010).

These authors found this complexity to be add odds with how counter-narratives are currently conceptualized. They reviewed eight counter-messaging initiatives, and found that all applied an understanding of particular extremist or terrorist groups and the ideology they conveyed in their narratives to the construction of counter- or alternative messages. Yet, had little interest in social processes and push/pull factors that are often involved in radicalization. As such, they wondered whether there was any utility in these programs. According to these authors, “*given their incomplete theoretical underpinnings of these campaigns, it would be difficult for these programs to meet their desired aims*” (ibid. p. 78).

In sum, use of theory is an essential strategy in order to enhance the effectiveness of communication efforts. Theory-driven approaches are more likely to produce results, and return on investment, if any can be expected.

4.3.2. DEFINE AND KNOW THE AUDIENCE

Audiences are not monolithic groups and audience members can differ substantially in their responses to different communications. Moreover, it is unlikely that all members of the audience can be reached utilizing the same approach. Hence, different communication strategies and messages will be necessary for different people. Within the broader communication literature its therefore generally recommended that program planners carefully define and segment their target audience(s), and customize their communication strategy accordingly (e.g., Boslaugh et al., 2005 ; Egner, 2009; Kreuter & Wray, 2003; Noar, 2006; Noar, 2011; Noar, Harrington, Aldrich, 2009; Slater, 1996; Slater, Kelly & Thackeray, 2006; Snyder, 2007; Thackeray, Neiger, Hanson, MacKeznie, 2008).

The basic idea of audience segmentation is dividing a heterogeneous audience into relatively homogenous subgroups (i.e., segments) on the basis of (theoretically) meaningful attributes which are either known (or presumed) to be related to desired outcomes (Boslaugh et al., 2005). This process is considered to be a prerequisite for developing messages that are responsive to the circumstances, predispositions, life experiences and other relevant characteristics of the audience (Kreuter & Wray, 2003). Likewise, segmentation will largely drive the selection of other aspects of communication strategy (e.g., channels, and credible sources (Noar, 2011; Slater, 1996). The importance of audience segmentation cannot be underemphasized. A failure to adequately segment is may render a campaign ineffective. For example, Berger (2016) notes that one of the reasons why many anti-drug campaigns of the past have largely failed to produce intended results, is that they targeted to wide an audience.

There is no overriding segmentation strategy for every situation. Differences in objectives, populations, contexts, and circumstances will necessitate different segmentation decisions. In addition to such considerations, segmentation choices will depend upon the available resources. However, as a general rule, the more narrow and homogenous the segment, the more targeted strategies and approaches can be and the more likely that they will be effective (Noar, 2011). Thus, ideally you want to have a perfectly tailored strategy for each person. While the digital era offers unique opportunities for individual tailoring, in practice, this may however not always be feasible (Slater et al., 2006).

There is an almost infinite number of variables on which audiences can be segmented. Arguably, the simplest segmentation approach is to subdivide audiences on the basis of demographic variables (e.g., age, gender, race or ethnicity, social class, religion) and geographic location (e.g., city, region) (Boslaugh et al., 2005; Egner, 2009). This approach is fairly common in other fields, such as health promotion (Boslaugh et al., 2005). However, it is frequently argued that segmentation on the basis of geodemographic homogeneity alone may only be of limited use for developing meaningful strategies and messages. Among other things, because it tends to lead to subgroups with great intra-segment variance (e.g. Boslaugh et al., 2005; Rimal & Adkins, 2003). Hence, rather than using geodemographic variables alone, its recommended to use more sophisticated strategies and give consideration to cultural psychographic, attitudinal, and behavioral variables and/or combinations of these and other variables (Boslaugh et al., 2005).

Related to our research group, geographic and demographic variables have been used for targeting (online) prevention efforts towards supposed vulnerable communities, typically Muslim communities (Harris-Hogan, Barrelle & Zammit, 2016). Critics have argued these initiatives inherently end up with too broadly defined audiences. As a consequence, such efforts do not reach those most in need of attention and largely draw in the wrong crowd, namely mainstream Muslims (Richardson, 2013 ; Harris-Hogan, Barrelle & Zammit, 2016) rather than those to whom the intervention should be specifically addressed. Moreover, it is asserted that such initiatives may label entire Muslim communities ‘suspect’ or ‘a security threat’, and may well alienate Muslims who feel they have unrightfully been singled out solely due to their religion and/or Arab background (Berger, 2016; Richardson 2013). In addition, it has been pointed out that targeting to broad an audience could potentially increase curiosity in extremist groups and their ideologies (Berger, 2016). As such, rather than countering (violent) radicalization, such efforts may be counter-productive. Thus, it is recommend to avoid targeting large cross-sections of the general population and direct one’s efforts to smaller and narrowly defined audience segments. For instance, as we suggested in the previous chapter, perhaps those individuals who show some curiosity or sympathy for extremist groups and the views they espouse, seek to engage with extremist content, but are not yet active supporters.

Of course, one can only do so to the extent target audiences are thoroughly understood and can meaningfully be identified. Unsurprisingly, conducting rigorous audience research is widely considered to be vital for a successful campaign (e.g. Egner, 2009; Fink & Barclay, 2013; van Ginkel, 2015). For instance, van Ginkel (2015, p. 9) suggests that it is important that those undertaking a counter-narrative campaign gain insight in: *“the profile of the target group, the question of with whom its members are in contact, the background that its members come from, the level of education or professional experience that they have, their (former) interests, and the sources used to find information”*. Other authors have put forward one should consider *“(1) how information reaches and flows through communities, (2) how the credibility of the information is determined, (3) whether formal or informal power structures are of greater importance in disseminating the message and shaping its interpretation, and (4) which individuals or groups have the greatest potential to influence the behavior of the campaign target group”* (Fink and Barclay, 2013, p. 33). The main point, however, is that one needs to learn as much as possible about the target audience, and that decisions related to the target audience should be informed (as much as possible) by data instead of presumptions (Noar, 2011). It has become increasingly possible to harness the power of social

media data to gain insight in the audience. We will explore some of these and other strategies in Chapter 5.

It should be noted that it may not always be possible to conduct a very thorough audience analysis, as it requires initiatives to have sufficient resources (e.g., finances, time, and technical know-how). Moreover, not all information can be found with open search research (Van Ginkel, 2015). As pointed out by van Ginkel (2015), it might even be necessary to create legal powers for certain authorities to conduct particular kind of analyses thoroughly. Furthermore, it may necessitate proper funding and cooperation to facilitate the exchange of information and knowledge. This may not always be feasible for small-scale efforts.

Thus, in the planning phase of the campaign, defining the specifics of the target audience is an important step to take into account when designing a counter-narrative campaign. In the previous chapter, we already concluded that counter-narrative campaigns may be best targeted in a preventive manner. However, from this literature, it is clear that a too broad scatterplot approach may run the risk of being ineffective and counterproductive. Thus, ideally, a campaign should aim to focus on a tailored and specifically targeted number of individuals or segment of a group.

4.3.3. LEARN FROM (FORMER) EXTREMIST NARRATIVES

Especially when one aims to develop counter-narratives, various authors mention that it is important to analyse and develop an in-depth understanding of the mixed layers and themes present in the narratives and supporting messages of the extremist group(s) (e.g., Ashour, 2011; Braddock & Horgan, 2016; Hedayah & ICCT, 2014; van Ginkel, 2015; Zeiger, 2016). As noted in chapter 2, among others, extremist narratives contain political, moral, religious, and socio-psychological dimensions. There is no 'one-size fits counter-narrative' that can address all the layers and themes contained in Salafi-Jihadi narratives. Rather, counter-narratives should be tailored to address specific components (van Ginkel, 2015). Hence is important to gain insight in the type of narrative used by the violent extremists and their messengers in order to better tailor and focus on appropriate counter or alternative narrative.

Beyond looking at extremist narratives, it is also important to consider why people in a certain context are drawn to these narratives and why they might resonate (Davies et al., 2016; Tuck & Silverman, 2016). Studies that explore the motivations of those that are drawn to Salafi-Jihadi extremist groups may provide some general guidance. For instance, Feddes, Nickolson and Doosje (2015) have distinguished four key motives⁴¹: (a) search for identity; (b) search for meaning/significance; (c) search for justice; and (d) search for adventure/sensation. A clear *identity* derived from group membership can be a source of self-esteem and it can reduce feelings of uncertainty (Doosje et al., 2013; Hogg, 2014). Related to this, some people feel attracted to radical groups as they can provide them with a meaning, purpose and significance in life (e.g., Kruglanski et al., 2014) . For others, the search for justice for their group is a stronger motivation. They are

⁴¹ In part, their work draws on Venhaus (2010), who conducted an extensive examination of the motivations of youth that joined al-Qaeda. On the basis of interviews and personal histories of 2,032 "foreign fighters" he identified four groups of 'seekers'. (1) *revenge seekers* who seek an outlet for frustrations; (2) *status seekers* that long for some form of recognition; (3) *identity seekers* looking for belonging to a group; and (4) *thrill seekers* who seek adventure.

primarily ideologically driven. Finally, for some people, the attraction of radical groups lies in the adventure, sensation and excitement associated with the violence nature of the group. Importantly, each of those 4 motivations may potentially be triggered by radical propaganda. Likewise, a counter-narrative campaign could pay attention to any of these 4 potential motives that people might have for joining an extremist group. Thus, a counter-narrative aimed at significance seeking individuals may look quite different from a counter-narrative aimed at people seeking for adventure and excitement in the radical group.

Similarly, it is possible to distinguish between different types of narratives. For example, on the basis of testimonials of 58 ISIS defectors, Neumann (2015) distinguished between three types of 'recruitment narratives'. These narrative provided rationales, justifications and incentives that contributed to these people joining ISIS. The first and most prominent narrative relates to the Syrian conflict, in particular the atrocities that have been committed by the Assad regime. The study points out that ISIS defectors often perceived this conflict in sectarian terms, and many believed that (Sunni) Muslims in the region faced genocide. It explains that "[c]onfronted with the notion of an 'existential threat' [...] this helped create a strong sense of obligation based on humanitarian instincts and their (Sunni) Muslim identity" (Ibid., p. 10). The second reoccurring narrative relates to faith and ideology, and was most prominent among those who were already part of extremist milieus and committed to Jihadi ideology long before the declaration of the so-called caliphate. They became convinced that the ISIS' so called Caliphate reflected a perfect and just Islamic State, which every Muslim had a duty to support. The third narrative that is distinguished relates to the personal and material needs. Some of the defectors mentioned they were attracted by the promise of material rewards (e.g., luxury goods, having debts paid off) or by notions of adventure, brotherhood, fighting, and heroism. According to Neumann (2015), those responding to these lures were typically not very religiously literate, and rarely put forward a strong sense of religious identity or obligation. Of course, identifying *what* attracts people in extremist narratives is not the same as explaining *why* they are attracted to them, hence it is imperative to give consideration to the local push and pull factors that may be implicated in processes of radicalization.

'Formers' can also inform the development of counter-narratives in a different manner, namely by studying the motivations of those 'who have left terrorism behind' (i.e., 'formers', 'defectors' or 'drop-outs') (e.g., Bjørge & Horgan, 2008, Jacobson, 2010; Weimann, 2015). The reasons why extremists leave organizations can potentially be leveraged in counter-narrative messages. In this regard, Neuman (2015) identifies four key narratives of ISIS defectors which mirror those that convinced them to join, in the sense that all defectors tended to argue that their (political, religious, or material) expectations were not met once they became part of the group. The most persistent narratives are (1) the degree to which the group is involved in fighting other (Sunni) rebel groups, and their restraint in fighting the Assad regime and helping those targeted by it; (2) the group's involvement in brutality against fellow (Sunni) Muslims; (3) Corruption and un-Islamic behaviors within ISIS ranks (oftentimes commanders and 'emirs'); and, (4) disappointment about the harsh living conditions and the quality of life. Related to this, defectors also pointed out that experiences in combat was not at all what they thought it would be, and did not meet their expectations in terms of action and heroism.

Jacobson (2010) provides some guidance as to how such aspects could potentially be leveraged. He suggests counter-narrative campaigns could incorporate elements that: (1) undermine the authority of and legitimacy of leaders of terrorist groups; (2) emphasize the targeting of fellow Muslims and civilians by terrorist attacks; (3) portray terrorists as criminals who do not live up to Islamic principles; and (4) demonstrate the reality of the hard life as a terrorist.

Thus, when developing a counter-narrative, it is important to develop a thorough understanding of the extremist narrative(s) one aims to counter. In part, it might be helpful to also consider why the target audience in question might have been attracted to these narratives in the first place. Insights gained from ‘formers’ could also be valuable in this regard. They can also serve to inform what elements to outline in a counter narrative.

4.3.4. ANALYZE AVAILABLE RESOURCES

Another recommendation is that program planners should think carefully about their available resources. In comparison to other approaches, it is argued that social media provide a relatively cost-effective means to reach target audiences (Freeman, Potente, Rock & McIver, 2015; Gold et al., 2012; Hanna, Rohm & Crittenden, 2011). While social media tools are available for free or at limited costs, running a campaign on social media evidently requires resources. In fact, several studies in the field of health promotion indicate that the development and implementation costs may be higher than often expected (Evers, Albury, Byron & Crawford, 2013, Freeman et al., 2015; Gold et al., 2012). Developing and implementing a social media program can be time and labor-intensive, especially when the chosen messaging strategy requires continuous content creation and responsive interaction and feedback (Freeman et al., 2015). Importantly, this is not to say that an engaging social media content strategy necessarily always needs to involve excessive creative development and production costs (Hanna, Rohm & Crittenden, 2011).

Furthermore, familiarity with social media from personal experience is typically not enough to run an efficient program in a social media setting. In fact, it may require a multidisciplinary team with broad range of skills and knowledge (Gold et al., 2012). Time and resources may be required to foster skills and competencies necessary to communicate effectively in social media (Evers et al., 2013; Freeman et al., 2015; Gold et al., 2012). Team members may for instance require training on the functionalities of different social media platforms, how users engage and interact in social media environments, the development of shareable content, and social media analytics (Freeman et al., 2015; Gold et al., 2012). Gold et al. (2012) also note that one should be aware that building and maintaining multidisciplinary collaborations itself may require (time, financial) investments.

In the case of insufficient resources it may be valuable to partner up with other organizations and stakeholders order to expand them (Weinreich, 2011). Even if this is not the case, this may be valuable. In particular, it is suggested that soliciting meaningful participation of relevant local communities in campaign planning, design and implementation may benefit the effectiveness of campaign efforts (e.g. Egner, 2009; GCTF, 2013; Snyder, 2007; MacNair & Frank, 2017; RAN, 2017). Among other things, it is argued that civil society groups and/or individuals may be best positioned to highlight the specific locally persuasive narratives that refute or

undercut extremist narratives and may be more effective and credible in delivering an alternative or counter-narrative than governments and statutory organizations (GCTF, 2013; RAN 2017; also see section 4.4.2.). Of course, soliciting community participation may require continued investment in building relationships and trust.

Thus, even though running an online counter-narrative campaign may be cost effective, developing and implementing a social media strategy may require a team with diverse theoretical and practical knowledge and technical expertise and skills, which is associated with relatively higher costs. Therefore, it is important to carefully consider available recourses. Also it is fruitful to build connections and enlist involvement of organizations and people at the grass root-level.

4.3.5. DETERMINE GOALS AND OBJECTIVES

An often mentioned and rather self-evident (yet, not always thoroughly implemented) best practice in the CVE and broader communication literature is that clarity of goals and objectives is essential (e.g., Beutel, 2016; Briggs & Frennet, 2014; MacNair & Frank, 2017; Snyder, 2007 ; Zeiger, 2016). There are many distinct targets for counter narrative efforts, which may require different types of campaigns, strategies and tactics (RAN, 2015). Determining clear goals and objectives is vital as it will guide the development of appropriate approaches and will help determine whether the program was a success. Goals tend to be formulated in more abstract terms and identify desired end-states to which an effort is directed. Whereas objectives are more narrowly defined descriptions of achievements that are needed to arrive at the desired outcomes(Veldhuis, 2012).

The Institute of Strategic Dialogue (Reynolds & Tuck, 2016; Tuck, 2016) provides some general pointers in regard to objective setting. First, they should be well defined and specific about the desired effect. What kind of changes will occur? Second, they should be established around measurable (or observable) criteria, *“with campaigners confident before the inception of the campaign that they will be able to discern, from available metrics and evaluation activities, whether or not they were achieved”* (Reynolds & Tuck, 2016, p. 10). Lastly, they should be realistic, taking in account matters such as the time-span and the intended audience of a campaign, budget and other resources.

4.4. DESIGNING YOUR PROGRAM

In developing a communication campaign, and thus also in developing a counter-narrative campaign, a more specified campaign strategy will be made on the basis of initial research and planning. In this phase of program development and design there are several important aspects to consider: developing messages (Section 4.4.1), selecting credible messengers (Section 4.4.2), and determining dissemination channels (Section 4.4.3).

4.4.1. DEVELOPING THE MESSAGE

Another factor that should be considered in the domain of counter-narratives involves the messages itself. As suggested earlier, there is no silver-bullet approach to developing effective messages. In fact, campaigns might be most successful when they combine a diverse array of messages tailored to different audiences while taking in account the given context. Yet, it is also argued that messages should be interrelated and support a consistent overarching narrative.

Otherwise may appear ad hoc and confusing at best or contradictory at worst (Ingram & Reed, 2016).

What are the important message factors to take into account? Studies of social media campaigns often do not explicitly specify message factors such content, evidence and appeals that are utilized in campaign (Shi, Poorisat & Salmon, 2016). Hence we relied primarily on the broader communication literature combined. While we are aware of the fact that the literature on effective messaging is very broad, we discuss four important factors: (1) the content of the message; (2) message format (2) message sidedness; (3) emotional appeals. We will discuss them below in relation to counter narratives, drawing on insights from the CVE literature.

Before we do so, it is important to point out that studies on social media campaigns indicate that the nature of messages is more complex in social media in comparison to traditional media channels (e.g., Shi, Poorisat & Salmon, 2016). Social media allow users to comment on and reshape messages, which may interfere with or undermine the intended effects of campaign messages. For instance, user-generated comments typically appear alongside an original message. Some research indicates that this may impact on the effect of the original message (e.g. Walther et al., 2011; also see Box 6, section 4.4.2.). This is particularly relevant in the context, of counter-narrative efforts, which may very well spark a negative or abusive responses (e.g., see Ernst et al., 2017). More research is warranted how interactions of audiences with a message interact to influence individuals.

4.4.1.1. THE CONTENT OF THE MESSAGE

Obviously, the content of the counter-narrative message is very important. Generally speaking, it is important to tailor the message to the audience. This can be achieved via various techniques related to the content of the message.

First, it is important to acknowledge that there is not only one counter-narrative message possible, but that there are several contents to consider. In line with the motivations for joining a radical group outlined earlier (i.e., justice seekers, identity seekers, significance seekers and sensation seekers; e.g., Feddes et al., 2015), it might be fruitful to consider drafting counter-narrative messages that are relevant for people with certain motivations (similar to the observation that ISIS creates messages catering for people with different types of motivations). This might imply that some counter-narratives may be more suitable for some people (e.g., sensation seekers), while other messages may be more applicable to others (e.g., significance seekers).

For example, in one of our interviews, Rieger⁴² suggested that counter-narratives messages should provide a sense of belonging and a positive identity (related to the identity seekers type). She argues that an important motivation of those who are drawn to extremist groups is that they provide them with a *“sense of belonging, a sense of we all belong together, relatedness, which is even more important for individuals who feel out of society already, or out of their peer groups. So, counter messages would, in my opinion and in the opinion of our research group, make a good point*

⁴² Diana Rieger, interview, 19 May 2016

if they also showed ‘hey, there is a place for you in our society, you can belong to other groups that are not extremists’. Give this sense of social groups being together instead of being alone”.

As a second technique with regard to the content, reframing the original radical propaganda might work better than denying the side of the radical propaganda all together (Beutel et al., 2016). For example, it is possible to create counter-narratives in which, on the one hand, the grievances of the group are acknowledged, but on the other hand, offer alternative (and legal) routes to dealing with the situation, in contrast to the violent solutions presented by extremists. This strategy, in which people are taken seriously in terms of their concerns, most likely makes people less inclined to counter-argue the counter-message.

A third technique in terms of content might be to present an alternative narrative (and not so much counter their narrative directly). For example, Aly et al. (2014, p. 36) argue that counter-narrative messages should provide “*an alternative vision to which one opts-in; a storyline that gives meaning to the actions it is requesting of the subscribed*”. This is in line with a narrative for people motivated to seek significance.

Thus, we conclude that, in terms of content, counter-narratives may focus on different sorts of motivations that people might have for showing interest in or associating themselves with radical groups. Specifically, motivations experienced by justice seekers, identity seekers, significance seekers and sensation might need different counter-narrative messages in order to be effective. This also means that it might be difficult, if not impossible, to create counter-narratives that will address each and every motivation of all the members of a target group.

4.4.1.2. NARRATIVES AND NON-NARRATIVE MESSAGES

The relative effectiveness of messages that are presented in a narrative format as compared to a non-narrative format has attracted considerable attention from researchers. A non-narrative messages include expository and didactic forms of communication, providing reasoned arguments and factual evidence in support of an advocated position (Kreuter et al., 2007). Loosely speaking, a narrative message provides “*a story that contains information about setting, characters, and their motivations*” (Braddock & Dillard, 2016, p. 1). This may for instance be an anecdote or testimonial. It should be noted that both message types are not necessarily mutually exclusive alternatives and evidence for their effectiveness is mixed. A non-narrative message may well integrate a narrative example. Similarly, a narratively structured message may contain non-narrative information such as arguments (Bilandzic & Busselle, 2013). Researchers have sought to determine the relative effectiveness of narratively structured messages in comparison to other, non-narrative forms of communication. However, it appears difficult to compare the different message types and often elements are included that may provide alternative explanations. Yet, research does show that narratively structured messages can influence “*recipients’ beliefs, attitudes, intentions, and behaviors such that they move into closer alignment with viewpoints espoused in those narratives*” (Braddock & Dillard, 2016, p. 18). Furthermore, of interest for our current endeavor, narrative messages are believed to offer several persuasive advantages when resistance is to be expected.

The power of narrative messages stems in part from their unthreatening nature (Weimann, 2015). Persuasive intend is often less obvious in narrative messages as compared to non-narrative

messages. As a result, recipients may be unaware that a message is trying to persuade them (Hoeken & Fikkers, 2014). In this regard, a relevant feature is the context in which the narrative is presented. For instance, in so called entertainment education, narratives are embedded in entertainment programming to convey a message (Slater, 2002). In practice, these are often health messages directed at reducing risk behavior (e.g., HIV prevention; Moyer-Gusé, 2008) . When people do not expect or are unaware of persuasive intent, various forms of resistance to persuasion are less likely to occur (Dal Cin, Zanna & Fong, 2004; Moyer-Gusé & Nabi, 2010). It is also an effect of the narrative format itself (Cohen, Tal-Or, & Mazor-Tregerman, 2015). In narrative messages standpoints and arguments are often implied (i.e., ingrained in the story and portrayed through events and characters) as opposed to explicitly stated, which is believed to make it harder to argue against them (Dal Cin, Zanna & Fong, 2004; Hoeken & Fikkers, 2014; Kreuter et al., 2007). Because narrative messages present experiences of other people, it is also suggested that they may be more difficult to refute to begin with (Dal Zin, Zanna & Fong, 2004). Hence, entertainment education and narrative messages in entertainment format appear useful for projecting a counter-narrative.

It is also proposed that the processing of narrative messages differs in important ways from the processing of non-narrative messages. In traditional dual process models such as the Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1986) , ability and motivation to engage in more effortful elaboration impacts on how a message is processed. Perhaps the most frequently mentioned factor that induces motivation is personal relevance of an issue (i.e., issue involvement). When motivation and ability are high, people process a message via a central route, in which the relevant arguments in favor and against are carefully weighed. When motivation and ability are low they do so via the peripheral route, in which people pay more attention to non-message related cues, such as the identity of the source (e.g., a celebrity advertising a product). It is argued and demonstrated that more durable attitude change is happening when a persuasive message is being processed via the central rather than the peripheral route. However, when confronted with narratives, involvement with the topic or issue addressed is assumed to be less relevant, and other factors such as involvement in the narrative, identification with story characters, and emotions evoked or emotional involvement are believed to be more important. For this reason and several others (see Igartua and Jair Vega, 2015)), dual process models are assumed to be less suitable for explaining narrative persuasion processes.

Researchers have used different labels to describe narrative involvement (e.g., transportation, engagement, absorption). Despite the different terms, the shared idea is that narratives are potentially capable of inducing a state where one “*primarily [engages] with the storyline while one’s own real-world environment or real life becomes less prominent*”, which is accompanied by an “*increased cognitive and emotional response to the unfolding events in the narrative*” (Moyer-Gusé & Dale, 2017, p. 2). Perhaps the most commonly used term is ‘transportation’ which has been defined as “*an integrative melding of attention, imagery, and feelings, focused on story events*” (Green, 2004, p. 248). Several studies found that transportation is associated with enhanced persuasion (e.g., Escalas, 2004; Green, 2004; Green & Brock, 2000). One explanation suggests that when individuals are thusly carried away, they are less able and motivated to contradict story points. Consequentially, when an individual is sufficiently transported, he or she is less likely to counter argue, in turn, making it more likely that narrative-consistent attitudes are accepted (Moyer-Gusé & Dale, 2017).

The degree of identification with characters has also been found to be positively associated with narrative effects (e.g., de Graaf, Hoeken, Sanders, & Beentjes, 2012; Hoeken & Fikkers, 2014; Igartua and Jair Vega, 2015). However, identification has been conceptualized in many different ways by different researchers across various studies. For instance, it is commonly used to refer perceived similarity to and liking of the characters (Konijn & Hoorn, 2005). Another conceptualization describes identification as a process whereby an individual loses self-awareness and temporarily comes to adopt the perspective, feelings and goals of a specific character with whom he or she identifies (Cohen, 2001). There are several explanations for the role of identification in narrative persuasion. Among other things, it is suggested that messages conveyed by characters with whom one identifies are more likely to be attended to, remembered, and learned. It has also been put forward that identification can facilitate the process whereby individuals become involved in a narrative (Cohen, Weimann-Saks & Mazor-Tregerman, 2017), as suggested, this may reduce their ability and motivation counter-argue.

In each of these constructs, emotions evoked or emotional involvement is an important underlying mechanism. In general, narrative messages are believed to be particularly powerful when they evoke strong emotions. Results of experimental studies (Konijn, Walma van der Molen, & Van Nes, 2009) provide evidence that emotions and being emotionally involved, overruled the knowledge that some narrative was fictional or ‘fake’ (even though this was explicitly stated beforehand). Participants in an emotional state attributed more realism to what was presented as fictional narrative and subsequently attributed more information value than participants who were not in an emotional state. This is explained by the lower and higher pathways in processing information in our brains in which emotions take control precedence (see Konijn et al., 2009). Thus, when a counter-narrative connects emotionally to the recipient, particularly when connecting to relevant needs, desires, and goals related to the emotions evoked, this increases the likelihood that it will be effective (see Section 4.4.1.3).

In sum, several benefits are associated with applying a narrative (entertainment) format that appears potentially useful for a counter-narrative strategy. It should be noted, however, research on narrative persuasion has predominantly focused on narratives advocating a clear, pro-social and consensual message which do not need to overcome strong counter attitudes (Cohen, Tal-Or, Mazor-Tregerman, 2015). Clearly, in case of counter-narratives addressing (violent) radicalization, strong counter-arguing can be expected. Nevertheless, the emotion evoking strength of (entertainment) narrative formats in particular, seems to be promising in reducing cognitive reflection and adding persuasiveness through enhancing perceived realism and information value of the message content.

4.4.1.2. MESSAGE SIDEDNESS

While some authors question the effectiveness of an approach that relies too much on logic and argumentation, others assume such an approach may still prove effective in relation to those who are in the initial phases of radicalization. For instance, Schmid (2015, p. 3) argues that “*While the fanatical extremists of ISIS might no longer be open to rational, persuasive arguments, many of those not yet fully radicalised might still have open minds. They can be confronted with facts and rational reasoning and might then be able to see ISIS for what it is...*”. He provides a series of argumentative elements that can be utilized to counter a dozen claims of ISIS which form key parts

of its narrative. Others have suggested such an approach may perhaps be fruitful in an earlier stage, in order to build resilience against radicalizing influences (e.g., Mann et al., 2015).

Taking this approach, one may want to consider how much attention should be directed to opposing viewpoints or perspectives. The degree to which messages consider opposing viewpoints is called message sidedness. Messages that only present those arguments in favor of a particular proposition or only from the perspective of those undertaking the influence attempt are commonly referred to as one-sided messages, while a two-sided message presents the arguments in favor of a proposition but also considers the opposing arguments and acknowledge the existence of opposing views (Allen, 1991). Two-sided messages often include arguments pro and con as well as refuting the latter. The question whether one-sided or two-sided messages are more effective has garnered much attention in the communication literature, albeit typically in relation to non-narrative messages containing solid arguments and factual information.

Allen (1998) and O'Keefe (1999) conducted meta-analyses of research regarding one- and two-sided messages, and both reached the same conclusion: two-sided messages are more likely to influence audience members than one-sided messages, but only when they refute opposing viewpoints. Two-sided messages are believed to gain their persuasive advantage by enhancing the credibility of the source (see Section 4.3.2.) as well as providing cogent arguments for why opposing views are wrong (Perloff, 2010).

Furthermore, it is suggested that the recipients' prior knowledge about and stance towards an issue may influence the effectiveness of both types of messages. One-sided messages have been argued to be more effective when the message recipients are uninformed about an issue. In contrast, two-sided messages are more likely to exert effect when the audience members are knowledgeable. One-sided messages are also suggested to be more effective if audience members are already favorably predisposed towards message assertions, whereas two-sided messages are more effective when message recipients are predisposed against message recommendations (Keller & Lehmann, 2008). Thus, in view of a counter narrative strategy, two-sided messages seem more likely to be effective.

Mann et al. (2015) suggest that one approach to two-sided messages referred to as 'inoculation' can potentially be used to build and strengthen youth's 'resilience' against extremist and radical propaganda on the internet. The inoculation theory (McGuire, 1964) holds that, analogous to the process of vaccinating individuals against a virus by administering a low dose of that virus, it is possible to inoculate individuals by exposing people to messages providing a weakened argument against an attitude they hold (Banas & Rains, 2010). As explained by Banas & Miller (2013), an inoculation message contains two key components: threat and refutational preemption. The threat component provides an individual with the motivation to bolster his or her attitudes. One way to elicit threat is through forewarning of an impending persuasive attack. This calls attention to the vulnerability of an individual's existing attitudes, thereby motivating persuasion resistance. The refutational preemption component provides arguments or evidence the individual can use against the impending attack. This component serves two purposes. It provides individuals with means to counter the persuasive attack and allows them to practice defending beliefs through counter arguing.

Research indeed indicates that inoculation messaging might be an effective strategy for conferring resistance to influence attempts. A meta-analysis by Banas and Rains (2010) shows that inoculation strategies are consistently effective in enhancing people's resistance to attitude change. Moreover, they found that inoculation messages were more effective than messages that provide mere support for individuals' initial attitudes. The efficacy of inoculation has been demonstrated in a wide variety of contexts (e.g., health, politics, marketing; see Compton (2013) for a review). For instance, inoculation has been shown to be a helpful strategy for protecting children to engage in smoking cigarettes (e.g., Pfau & Van Bockern, 1994; Pfau, Van Bockern, & Kang, 1992) and drinking alcohol (e.g., Godbold & Pfau, 2000). It has also been used successfully to confer resistance among supporters of political candidates to attack messages from opposing candidates (e.g. Pfau & Burgoon, 1988; Pfau, Kenski, Nitz, & Sorenson, 1990). However, their usefulness for strengthening resilience against extremist messaging has not been studied (Mann et al., 2015). Banas and Miller (2013) did show that inoculation can be applied to confer resistance against propagandistic conspiracy theory messages. Their study provides preliminary evidence that brief, cognitively focused inoculations can reduce the effectiveness of comparatively longer and emotionally charged persuasive messages.

4.4.1.3. MESSAGES DESIGNED TO AROUSE EMOTIONS

An often made recommendation in the literature on counter-narratives is that messages should concentrate on arousing emotion rather than rely on facts and logical arguments. In line with this suggestion, studies in political science have shown that people remember information better when it is presented in a manner that arouses emotions (e.g., fear, enthusiasm, anger) than when it is presented in a neutral manner (e.g., Civettini & Redlawsk, 2009). However, just as there are many emotions, there are a wide variety of emotional appeals which could be leveraged in messaging. Interestingly, the utility of different 'emotional appeals' has only received very limited attention within the specific context of counter-narratives. However, the broader literature on emotional appeals in persuasion provides insights. This literature overwhelmingly focused on fear appeals, but some attention has also been given to other emotions, such as humor and regret. In this section, we provide a short outline of this literature.

Before we do so, it is important to note, however, that a message may be designed to evoke a particular emotional state, but may not be successful in doing so. For instance, what may be considered fearful for those designing a campaign, may not be considered fearful by message recipients (O'Keefe, 2015). This seems particularly the case with adolescents who may, for example, consider the gory graphics on cigarette packages kind of humor rather than fearful (Konijn, 2008). Likewise, warning messages may create boomerang effects among this age group (Bijvank, Konijn & Bushman, 2009; Veldhuis, Konijn, & Seidell, 2014).

Furthermore, an appeal may not only arouse the intended emotion but also other emotions. To illustrate with an example from the field of health promotion, Dillard et al. (1996) reviewed thirty-one AIDS prevention fear appeals and found that thirty evoked change in more than one emotional state. Moreover while most research focuses on the impact of a certain emotional appeal after messages are viewed, more recently, it has also been suggested that message recipients may experience a flow, or evolution, of emotional experiences over the course of a message with could impact its persuasiveness. Although this still needs further study, it is

suggested that presenting information with carefully patterned emotionally evocative sequences may enhance the effectiveness of messages (Nabi, 2015) .

With this in mind, in the following, we discuss fear and humor appeals, which have attracted some attention in the literature on counter-narratives. Both aim to arouse an emotional state in order to achieve persuasive effects. A related way to harness emotion for persuasion is through the anticipation of emotional states, which we discuss subsequently.

Fear appeals

Fear appeals are commonly used to confront people with the negative consequences of not adopting message recommendations, under the assumption that this will motivate people to follow or perform the recommendations (Rogers & Mewborn, 1976). As explained by Ruiter, Kessels, Peters, and Kok (2014), in theory, effective fear appeals provide two types of information. First, they present threat (e.g., lung cancer) to which a person is deemed to be susceptible (e.g., smoking cigarettes may cause lung cancer), and which is severe (e.g. lung cancer may be deadly). Both the elements of severity and susceptibility are believed to be necessary to arouse fear (however, these are often implicit assumptions, hardly tested). Second, they need to provide a prospect to avert the threat by providing recommendations for a protective response. This acceptance of the recommendation is promoted by presenting the suggested response as effective in negating the threat (i.e., response efficacy) and by boosting a person's perception that he can effectively execute this response (i.e., self-efficacy). When the recommended response is perceived to be effective and feasible to an individual, he or she is likely to adopt the recommendations. Importantly, though, when this is not the case and people think they cannot avert the threat, this may prompt defensive and even counter-productive responses. Although fear appeals have been widely used in various fields and a number of studies have shown their effectiveness, their utility is not uncontested. Studies on health promotion (Peters et al. 2013, Ruiter et al., 2014, and crime prevention (Petrosino, Turpin-Petrosino, & Buehler, 2003) indeed indicate that fear appeals are often ineffective and may even backfire (cf. Konijn, 2008).

Limited attention has been paid to fear appeals in counter-narrative work. Jacobson (2010) suggests that fear appeals could potentially be an effective message strategy for certain counter-radicalization purposes. For instance, in order to “*bring home the reality*” why individuals should be afraid to engage in violent activities (in this regard he uses the example that raising fear may deter those that want to engage in suicide attacks) . However, recently, Beutel et al. (2016) argued that, given their dubious record in various fields, fear appeals are best to be avoided in this context. However, Beutel et al. (2016) add that this does not mean that there may be some utility in messages that have some “shock value”. Given the above and the specific target group, we conclude that the approach of fear appeals might not be a very effective strategy for designing counter-narratives in the context of (violent) radicalization.

Humor appeals

Humor-based messaging has been utilized in various domains, including commercial advertisement, entertainment, and health. However, findings concerning the persuasive effectiveness of humor appeals have been mixed and contradictory. Yet, there is evidence that humor “*has the potential to enhance psychological states associated with persuasion*” (Nabi, 2016, p.3). Then again, studies also raise several concerns that warrant careful consideration.

For instance, a meta-analysis in the field of advertising (Eisend, 2009) found that humor can increase attention to the message, and positive affect, attitudes toward the brand and add, and purchase intentions. However, contrary to what is suggested in earlier reviews, no evidence was found that humor impacts positive or negative cognitions as well as liking of the advertiser. In fact, it was found that humor may even detract from the credibility of the source (see section, 4.4.2.).

Another line of research has focused on the use of humor in entertainment (-education) programming. In this context, studies (Moyer-Gusé, Mahood, & Brookes, 2011; Nabi, Moyer-Gusé, & Byrne, 2007) have found that while humor-based entertainment content may lead to a reduction of counter-arguing, which could benefit persuasion. It also stimulated participants to trivialize (Moyer-Gusé, Mahood, & Brookes, 2011) and discount the message more (Nabi, Moyer-Gusé, & Byrne, 2007). Elsewhere, it has also been argued that the fun or laughter evoked by the humor appeals of a message may overrule the content of the message that can then not be remembered (Konijn, 2008). Thus, humor-based content could potentially fall flat.

Studies on humor and persuasion typically do not differentiate between different types of humor. Yet, authors have put forward that various types of humor can be differentiated (e.g., slapstick, clownish humor, irony, satire etc.) (e.g., Buijzen & Valkenburg, 2004). Recent research provides preliminary evidence that different types of humor may impact differently on message recipients. A recent study of Iles & Nan (2017) considered the differential effects of ironic and sarcastic versus no humor appeals in health related-advertising messages. They found that sarcasm has a 'detrimental effect' on persuasion. Irony also reduced the persuasive effectiveness, albeit to a lesser extent. Compared to no humor, the use of sarcasm reduced overall negative affect, increased counter-arguing, and decreased perceived argument strength. Likewise, it was found that irony increases counter-arguing. Their study indicates it may be fruitful to consider different types, instead of treating it as a 'generic' concept. Their study suggests that certain types of humor could not only fall flat, but could potentially be counter-productive.

In relation to counter-messaging the utility of humor has received only limited attention. Bartlett, Birdwell & King (2010) argue that the use of satire could potentially help strip extremist groups from glamour and mystique (although they emphasize that governments should refrain from utilizing it). They point out that sustained satire has historically played an important role in undermining the popularity of extremist movements such as the Klux Klux Klan. Others have pointed out the risks, and argue it should be utilized with great caution. For instance, Goodall, Cheong, Fleischer, and Corman (2012) assert that humor is a weapon with a double edge: *"Just as an off color joke can offend your co-workers or sour a personal relationship, humor has the potential to be divisive and motivating in ways that are detrimental to larger policy goals"* (p. 73).

Beutel et al. (2016) suggest that knowing the target and intention of a humor-based message is crucial. They speculate that if the target of humor-based messages is for example a terrorist leader, with the intent to undermine his or her credibility and have potential recruits no longer look up to him, this may increase the potential of such a message strategy. However, when it is intended to ridicule the potential recruits and their motivations and concerns, this may well be taken as an insult and backfire, eliciting defiance as a response. Likewise, Kruglanski suggests mockery and ridiculing carries the potential to backfire *"they may mock you right back"*⁴³. Taking all of the above in account, we conclude that humor is perhaps best utilized with great care and potential risks should be carefully analyzed from various perspectives beforehand.

That having been said, it should also be pointed out that some studies do indicate that humor-based messages may have a particular persuasive advantage. That is, humor has been

⁴³ Arie Kruglanski, interview, 8 March, 2016

found to be a factor that may enhance interpersonal sharing of messages and interaction. For instance, Campo et al (2013), who studied the effects of a pregnancy prevention campaign, found that campaign exposure and the use of humor (i.e., clownish-humor, surprise, absurdity) “*was a significant predictor of talking with and/or showing the campaign to others*” (p. 4) . Related to our particular subject matter, a recent study of Centre for the Analysis of Social Media (CASM) of the Demos Institute (Bartlett & Krasodonski-Jones, 2015.) examined the extent to which different types of ‘counter-speech’ (i.e., a “*common, crowd-sourced response[s] to extremism or hateful content*”) are produced and shared on Facebook across four countries (p. 5). They found that the most popular tone of posts on counter-speech pages across the three countries was a funny or satirical tone⁴⁴. The popularity of a post was measured by the amount of interactions content generated (i.e., likes, comments, and shares). Usually these posts parodied extremist language on hate-pages related to issues such as immigration, religion, and race. Of course, this does not do away with earlier mentioned concerns. In Box 5, we further discuss the relation between emotions and message sharing.

Anticipated regret

Rather than arousing emotions directly and offering a recipient a way to deal with the aroused feelings, appeals may also focus on drawing people’s attention to positive or negative feelings that are to be expected if a particular recommendation is (not) followed (O’Keefe, 2015). Put differently, messages may prompt people to imagine how they might feel if they (do not) follow a certain course of action, which can influence their intentions and actions. In this regard, studies have often focused on anticipated regret. Regret has been described as a negative emotion experienced when an individual realizes or imagines that the present situation could have been better had he or she previously chosen or acted differently (Sandberg & Conner, 2008). The assumption is that people do not like to experience regret as a result of their choices. Clearly, individuals may anticipate regret, which can serve as a motivation to (not) engage in certain behaviors. It is suggested this may hold in particular when the consequences of a high-risk activity are severe and irreversible (van der Pligt & Vliek, 2016). As such, it has been suggested that encouraging people to anticipate regret may impact their intention to engage in risky behaviors.

The potential of evoking anticipated regret in influence attempts has been evidenced in a variety of studies, in particular in the field of preventative health behaviors. Researchers have found anticipated regret can influence a range of health- and safety-related behaviors such as substance use, safer sex practices, vaccinations, organ donation and road safety offences (see Koch 2014 for a review). A meta-analysis of Sandberg & Conner (2008) revealed a strong relation

⁴⁴ Other message factors they looked into were the format of posts (e.g., video, photos, text) and the content of posts. They found that photos and content in the form of questions and commentary were most likely to generate interactions with users. Yet, they also looked into the most popular content on right wing populist pages. In this context, it was found that the most popular tone was celebratory (i.e., commemorating, war dead or patriotic pride) followed by angry. In three countries the most popular format was photos. Comparing both counter-speech and right wing populist pages, they found the latter were way more effective in sharing content beyond their network of followers and were generally more active (Bartlett & Krasodonski-Jones, 2015).

between anticipated regret and behavioral intentions, and a moderate link between anticipated regret and behavior. Recently, a more inclusive meta-analysis (Brewer, DeFrank, & Gilkey, 2016) provided similar findings. In addition, they concluded that “*anticipated action regret* [i.e., anticipated regret from engaging in an behavior] *had smaller associations with behavioral intentions related to less severe and more distal hazards, but these moderation findings were not present for inaction regret* [i.e., anticipated regret from engaging in behavior]” (p. 1264).

To the best of our knowledge, the utility of strategies that uses anticipated regret has not yet been studied in the domain of prevention of radicalization. However, de Wolf and Doosje (2010) suggest it may have some potential as a strategy to encourage those attracted to radical groups to consider possible negative emotional consequences of their decisions and actions.

In sum, the counter violent extremism literature holds that counter-narrative messages should arouse emotion in the target audience. However, limited consideration is given to which emotions should be aroused. The broader literature on emotion appeals indicates that it may be best to avoid using fear appeals. Certain types of humor might be useful, but humor-based messaging also carries clear risks and should be used with severe caution. Leveraging anticipated regret may have some potential to address (violent) radicalization, but research in the area of CVE is lacking. Dealing with younger age groups (e.g., adolescents) may further complicate suggestions as this particular developmental stage is characterized by increases in aggressiveness, sensation seeking, risk behavior, during which emotional appeals and warnings may easily backfire. Among them, gruesome or horrific messages may, for example, be considered ‘humor’.

Box 5: Emotions and message sharing

The effective dissemination of counter-narratives in a social media context may in part be dependent upon the degree to which a message is shared. Whilst the aim for counter-narratives is not to go ‘viral’, which implies untargeted message dissemination (RAN, 2015), the literature on ‘virality’ and online word-of-mouth does provide some interesting insights in regard to what message features increase the likelihood that a message is retransmitted. Although research in this domain is relatively new, some studies suggest that content that is emotionally-charged and evokes high-emotional arousal is more likely to be shared. This is not only useful to ‘get the word out’. As explained by Nabi (2016) this also creates the opportunity for repeated message exposure for the individual who is doing the sharing. Repeated exposure is associated with greater message effectiveness.

An often cited study in this domain (Berger & Milkman, 2012) is focused on how emotions affect email sharing of articles from the New York Times. Using a data set of nearly 7000 New York Times articles, they investigated how the valence of content (i.e., whether content is positive or negative) affects whether a message is highly shared. They further studied how specific emotions (e.g., awe, anxiety, anger, sadness) evoked by content, and the arousal or activation they induce, impact transmission. Their study found that positive content is more likely to be shared than negative content. However, their study points out that social transmission is more complex than valence alone. Arousal is argued to be key. They found that content that evoked high-arousal emotions - regardless of whether they were positive (awe) or

negative (anger or anxiety) – was more likely to be shared, than content that elicited a low-arousal emotion (i.e., sadness).

The role of emotion has also been examined in relation to social media. Analyzing two data sets of more than 165,000 tweets, Stieglitz and Dang-Xuan (2013) found that Twitter messages that are emotionally-charged tend to be retweeted more often than messages that are neutral. In another study, Nelson-Field, Riebe and Newstead (2013) examined the sharing of 800 videos (commercial and non-commercial) on the Facebook platform. In particular, they looked at the emotional responses these videos evoked, and the likelihood of audiences to share those videos given the evoked emotion. Their findings correspond with those of Berger and Milkman (2012). They found that the degree of arousal is the primary driver of the sharing of video content. They also found that valence plays a role in the sharing of video content. Videos that elicited positive emotions were more likely to be shared than those that evoked negative emotions (no matter the degree of arousal). However, while it was found that valence mattered, it did so to a lesser degree than arousal. While they recommend evoking a positive emotional response over a negative one, they argue that when it comes to sharing *“it is less important that the emotion felt be a positive one, than that it should be strongly felt”* (p. 210).

Similar results were found by Guadagno, Rempala, Murphy and Okdie (2013). They conducted an experiment to explore what videos they were most likely to forward. They found that individuals which reported strong emotional responses to a video reported greater intent to spread the video. Furthermore, their results suggest an “arousal hierarchy”. Videos that elicited positive emotions were most likely to be forwarded. Videos that evoked a diffuse arousal were more likely to be forwarded than negatively arousing or non-emotional videos. Finally, they found that videos evoking negative emotions were more likely to be shared than non-emotional videos. Besides these studies there are several other studies that suggest the importance of emotion in message sharing.

Thus, these studies and others suggest that messages that are emotionally-charged and arousing are likely to be shared (cf., Felten, Taouanza, & Keuzenkamp, 2016). In particular, messages are more likely to be shared when they elicit positive emotions rather than negative ones or no emotions at all. At the same time, messages eliciting some negative emotions are more likely to be shared than messages that elicit no emotion at all. These notions are worthwhile to take into consideration when constructing messages.

4.4.2. SELECT CREDIBLE MESSENGERS

The term ‘source’ is often used to refer to either the sender of a message or the model that appears in the message to deliver the information, demonstrate behavior, or provide a testimonial (Atkin, 2004; but see Box 6). The counter-narrative literature generally holds that even when messages are perfectly crafted, the source (here more commonly referred to as ‘the messenger’) should be perceived as credible in order for the message to be convincing to a target audience (e.g., Briggs & Feve, 2013; Davies et al., 2016; Fink & Barclay, 2013; Braddock & Horgan, 2016; ICCT, 2014; van Ginkel, 2015; Weimann, 2015; Zeiger, 2016). The importance of so called ‘source credibility’ is widely echoed in the broader research literature on persuasion.

What exactly constitutes source credibility? Source credibility is commonly understood as “*the believability of a source*” as interpreted by the message recipient, and is composed of two primary dimensions: trustworthiness and expertise (Metzger & Flanagin, 2013, p. 211). Yet, other dimensions may impact on the believability (or effectiveness) of a source as well, such as attractiveness, likability, similarity and familiarity (Atkin & Rice, 2012).

When a source is perceived as credible, then, recipients are generally more likely to accept the message claims, and in turn, influence attempts are more likely to sort effect. However, the level of ‘fit’ also does matter. For example, a famous athlete will not fit an advertisement for alcoholic beverages. In this respect, the old-fashioned sexy lady to sell cars stems from a classical conditioning approach (Konijn, 2008), but clearly is a misfit in today’s society. Furthermore, it is clear that the assessments of credibility in online environment is often much complex than in previous media contexts (See Box 6).

In the context of counter-narratives, counter violent extremism experts and practitioners suggested various sources that can potentially serve as credible sources for counter- and alternative narratives (e.g., Briggs & Feve, 2013; Briggs & Frenett, 2014; Davies et al., 2016; Braddock & Horgan, 2016; van Ginkel, 2015; Weimann, 2015; Zeiger, 2016). However, which messenger will in fact effective will heavily depend on the context and the audience members in question. As noted by Weimann (2015), different countries, regions, communities and groups, may require different approaches. As suggested above it is also vital to consider the message one wants to relay.

As will be explored further in chapter 6, it is commonly assumed that governments may not be the most effective messengers for counter-narratives, although they can engage in other forms of strategic communication (e.g., Briggs & Feve, 2013; Davies et al., 2016; van Ginkel, 2015; Romaniuk, 2015; Weimann, 2015). Among other things, this is due to “say-do” gap. This refers to a perceived gap between governmental rhetoric and actions (Romaniuk, 2015). On the other hand, it might be argued that “the government” is quite broad and can speak with different tongues. For example, communications from different ministries may distribute different messages, and at the local level, different municipalities can have different approaches.

Still, messages are believed to more likely to strike a chord when they originate from the grassroots level. As argued by Bouwman in one of our interviews, “*it is very important that it occurs from within the group. And that you do not play the expert and say "we're gonna develop all the materials". If you really want to apply a positive strategy, it is very much stemming from their own community.*”⁴⁵ In the subsections below, we discuss some commonly mentioned examples of potentially credible sources, which include: former extremists, victims of extremist violence, youth, family members, and civil society and community actors.

⁴⁵ Martine Bouwman, interview, 15 April 2016

Box 6: Source credibility in online environments

The social media contexts present several challenges for traditional conceptualizations of source and credibility evaluation. One challenge stems from the “*multiplicity of sources embedded in the numerous layers of online dissemination of content*” (Sundar, 2008, p. 74). Sundar & Nass (2001) who have provided a typology, suggest there are three layers of sources in online and social media environments: (1) visible sources (2) technological sources, and (3) receiver sources. Visible sources are “*the sources seen by the receiver to be delivering the message or content*” (Sundar & Nass, 2001, p. 58). They are frequently perceived as the originator of the information, while they may just be information gatekeepers. Technological sources are technological interfaces, which may be perceived as an original source, no matter that they are not independent. Lastly, receivers of information themselves can become a source in social media environments (Hu & Sundar, 2009). To illustrate, in the context of a counter-narrative effort, a visible source may for example be a civil society organization or someone involved in a campaign posting a message. A technological source might be Facebook, Twitter or Instagram. To become a receiver source, the recipient of a message may post a like, add a comment, or do a repost after having received a message (Shi, Poorisat & Salmon, 2016).

As suggested above, communication through these platforms is often not a linear process, whereby a sender directly relays a message to a receiver. Instead, multiple layers of source may exist for a message (Shi, Poorisat & Salmon, 2016). To illustrate, a video containing a counter-message may be originally posted by a particular civil society organization (the original promoter) on YouTube, but someone may pick it up from a Facebook post on a friend’s page who himself picked it up through a blog. Which of these sources is most likely to resonate with the recipient as the source? Whereas one might argue that it is the civil society organization who initially distributed it, another might say it is the friend or the blog - and the latter might be more persuasive than the former (Hu & Sundar, 2009). It appears that the perceived credibility of each of these sources as well as their interaction could potentially impact how audience members evaluate message claims (Hu & Sundar, 2009; Shi, Poorisat & Salmon, 2016). Further study is still required to determine discern the relative impact of layers of sources on message selection and evaluation.

Importantly, as suggested earlier, receiver-sources may potentially add to, change, and repurpose the message as they see fit (e.g., add a comment, change its content etc.). If one further takes into account that sources and content of a message also interact to influence message recipients credibility evaluations one comes to understand how complex the concept of source credibility actually is within contemporary social media environments (Metzger & Flanagin, 2013, see also Shi, Poorisat & Salmon, 2016).

Adding to the complexity, some scholars have suggested that traditional credibility indicators may play a less prominent role in how users locate and make judgements about information they encounter online (e.g., Metzger, Flanagin & Medders, 2010). Traditional approaches may include assessing source credentials and biases, and verifying quality or accuracy of information). However, some studies indicate that information consumers on social media and the internet often do not engage in such effortful evaluation process. Instead, people

routinely invoke simple heuristic cues for credibility assessments of both sources and information (e.g., social endorsements) (Metzger & Flanagin, 2017). How adjacent social cues and audience-generated messages on social media impact on credibility evaluations still warrants further research.

4.4.2.1. FORMER EXTREMISTS

It is commonly assumed that leveraging the voices of former violent extremists (i.e., ‘formers’) and foreign fighters (i.e., ‘defectors’) may prove to be effective (e.g., Ashour, 2011; Braddock & Horgan, 2016; Briggs & Feve, 2013; van Ginkel, 2015; MacNair & Frank, 2017; Neumann, 2015; QIASS, 2013; Waldman & Verga, 2016; Weimann, 2015; Zeiger, 2016). Just as with former criminals, former alcoholics or former drug addicts, they are more likely to be perceived as ‘street credible’, as they have experienced the life as an extremist first-hand. Furthermore, they can probably better relate to the experiences of those individuals attracted to violent extremist groups and ideology, as they have gone through similar processes themselves (Zeiger, 2016). On similar grounds, Rieger⁴⁶ suggests that in comparison to other sources, they may perhaps have “*greater potential to raise doubt in young adolescents or in young adults who are maybe more extremists*”. Talking from direct experience, they may also be able to tell a powerful story about the reality of life as an extremist or jihadi and their disillusionment with the group (Weimann, 2015). In this regard, Kruglanski⁴⁷ points out that stories of formers can show that “*that joining extremist groups does not provide glory and fame like they expect. According to our analysis a major motivation for joining is this quest for fame and glory [...] If people who left extremism can recount their experiences about how they were humiliated, how they were mistreated, how they were given second or third rate jobs, came to serve as cannon fodder just to be killed. These kind of messages can be effective to show they are not treated in glamorous, glorified manner*”. As such, it is argued they formers may represent a powerful voice that could potentially help prevent vulnerable individuals from joining these groups.

Yet, one should also be aware of other challenges and limitations in regard to soliciting the involvement of formers. First, it is quite difficult to find formers willing to speak out against their former group, for several reasons (Hedayah & ICCT, 2014). They may want to leave that part of their lives behind them or may not want to have public profiles (Hedayah & ICCT, 2014; RAN, 2015). In addition, some may fear repercussions or harassment by members of the former group when they find out what they are doing, which could potentially happen and is another point of concern (Briggs & Feve, 2013; Hedayah & ICCT, 2014). In general, careful consideration should be given to the safety of those that engage in counter-narrative work (Hedayah & ICCT, 2014).

Second, formers may not be psychologically healthy enough to engage in such work, having gone through traumatic experiences. Related to this, reliving past experiences by engaging in counter-narrative work may be stressful. Hence, serious consideration should be given to their

⁴⁶ Diana Rieger, interview, 19 May 2016

⁴⁷ Arie Kruglanski, interview, 8 March, 2016

well-being (Briggs & Feve, 2013). Thus, it will require careful vetting and selection, to determine whether a former can be reliably used in the capacity of a messenger (Zeiger, 2016).

Third, given their history of involvement violent extremism, leveraging these voices could potentially give rise to contention among some publics (QIASS, 2013). In other words, people might wonder about the role that the former has played in the past and whether the new role is deserved.

Thus, we conclude that, given some noteworthy limitations, formers may be in a good position to deliver a persuasive counter-narrative, as they can directly relate to their audience and can deliver a message from their own direct experience, although they have not always resulted in clear effects in other domains.

Box 7: Former extremist as source: the 'One2One' project

One example of a social media effort in which formers have been leveraged is provided by a recent pilot project of the Institute for Strategic Dialogue (ISD). It was set up as a partnership with Curtin University and members of the Against Violent Extremism network (AVE). Their One2One program aimed to devise a methodology and test the viability of an approach *“based on directly messaging those openly expressing extremist sentiment online and seeking to dissuade them from following that path”* (Frenett & Dow, 2015, p. 18) . The team made use of the Facebook tool called 'Graph Search' to identify at-risk candidates that openly endorsed and promoted right wing or jihadi extremist narratives online. Once identified, profiles were passed to former extremists to verify their 'at-risk status' and were subsequently contacted by them using the Facebook peer-to-peer messaging system in conjunction with a 'pay to message' functionality in order to elicit a conversation. Results of the early stage effort appeared to be quite promising. Of those individuals that were willing to respond (43% of those considered at risk of falling into what they termed 'violent Islamism'), a majority were willing to engage in a sustained conversation, which they defined as an exchange of more than five messages. They also found that engagement rates were affected by the variables tone and anonymity. When former extremists revealed their true identities, shared their experiences, and used a more casual tone, engagement rates were higher. Anonymous communication and antagonistic and aggressive tones led to lower engagement rates (Ibid., 2015).

4.4.2.2. VICTIMS

A second potential source of counter-narratives are victims of violent extremism. With proper training, coaching and support, victims of terrorist violence may also be able to fulfill a role as messengers (e.g., Beutel et al., 2016; Briggs & Feve, 2013; Davies et al., 2016; Hedayah & ICCT, 2014; RAN, 2016; Schmid, 2012; van Ginkel, 2015; Zeiger, 2016). Victims may be particularly effective messengers because they can speak with the unique moral authority as a survivor or witness (Beutel et al., 2016). In addition, testimonies of victims can potentially serve a strong preventative function by giving a face to and humanizing those that are attacked as well as highlighting how ordinary people can be affected and can cope with these terrible situations (see

Handbook: voice of victims of terrorism, RAN, 2016). Van de Donk⁴⁸ adds that the personal stories of survivors and relatives of victims can highlight the real impact of violence, which can potentially serve as a reminder that violence is not a solution. Importantly, he also suggests that leveraging the voices of victims against already committed extremist may actually be counter-productive. Their might testimonials serve to justify already committed extremists in their convictions. As with former violent extremists, when leveraging the voices of victims careful consideration should be given to their personal well-being and security (Hedayah & ICCT, 2014). Survivors have been used in a variety of projects, one of them being the well-known ‘Against Violent Extremism’ project (See Box 8).

Box 8: Example of victims as source: ‘Against Violent Extremism’ Network

A commonly mentioned example of an initiative in which victims play a central role is the Against Violent Extremism Network (AVE)⁴⁹. The AVE is overseen by the ISD in London, and is the largest existing network of victims and former extremist. It was founded at the Summit Against Violent Extremism in 2011, and is the result of a partnership between the Institute for Strategic Dialogue (ISD), Jigsaw (formerly Google Ideas), and the Gen Next Foundation and. The AVE brings together ‘formers and ‘survivors’ and seeks to use the lessons, experiences and networks of those who have first-hand experience with violent extremism and amplify their voices. In addition, it aims to serve as an incubator for initiatives that contribute to tackling violent extremism, Members are enabled to connect through their website and social media channels and exchange ideas, work together, find potential investors, partners and volunteers, and amplify their message (Against Violent extremism, n.d.-a). According to their website, the network has helped to establish more than 2600 connections and launch more than 80 (counter-narrative) projects.

4.4.2.3. FAMILY MEMBERS

Family members could potentially also be key messengers that can challenge the messages and appeal of violent extremist groups, and can play an important role in stopping the violent extremist narratives from spreading and undermining their appeal (e.g., van Ginkel, 2015; Zeiger, 2015; Saltman & Frenett, 2016; Weimann, 2015). At an interpersonal level, they are well-positioned to detect early signs of radicalization, and engage in dialogue to present an alternative narrative or counter-narrative. Mothers in particular, have been suggested to have a great deal of emotional influence on their children, and messages stemming from them could potentially have a strong impact (Saltman & Frenett, 2016). Yet, the radical person might also be that someone who already drifted away from the family, often with conflicts, and is not responsive to them anymore (Sieckelink & de Winter, 2015). Hence, it is suggested that capacity building and training may be needed before families and relatives can fulfill such a role (van Ginkel, 2015).

⁴⁸ Maarten van de Donk, interview, 19 april 2016

⁴⁹ Against violent extremism. (n.d.). Retrieved on 7 September, 2017, from <http://www.againstviolentextremism.org>

Additionally, it has been pointed out that family members may play an important role in persuading individuals to drop out of extremist and terrorist groups (Jacobson, 2010; Weimann, 2015). For instance, Jacobson (2010) points out that renewed ties and contacts with family members have led some individuals to reconsider their membership in such groups.

Beyond the interpersonal level, it is suggested that these voices can also be leveraged in online campaigns. Some campaigns have aimed to spread the stories of family members of violent extremists (See Box 9). It is suggested their testimonials can potentially provide a powerful narrative that can serve a preventative function (Zeiger, 2016).

Box 9: Example of family members as a source: ‘Open letter’ campaign

One example in which family members play a role, is the ‘Open Letter’ campaign which was launched by Mothers for Life Network. A network coordinated by German Institute on Radicalization and De-radicalization Studies (GIRDS) which brings together mothers who have experienced violent radicalization in their families. The campaign revolves around two open letters^{50, 51} in which counter-narrative techniques are used. These letters were disseminated through social media. Both letters urge sons and daughters in Syria and Iraq to come home, making use of citations from Islamic scripture. The letters were designed by GIRDS in cooperation with mothers from countries around the world. According to the website of GIRDS, the first letter managed to attract quite some attention, it was reported on by news outlets 1,785 times; shared 7000 time on Facebook; translated into eight languages, and responded to by ISIS within three and a half hours after publishing it. Furthermore, mothers from many more countries have since reached out for help and to become members of the network 2016 (GIRDS, n.d.). Their website doesn’t provide any details regarding the second letter.

4.4.2.4. YOUTH

Young people are often seen as audiences of alternative or counter-narrative efforts, leaving unrecognized that they can be key in delivering messages against violent extremism (e.g., Hedayah & ICCT, 2014; RAN, 2015; Richardson, 2013). As argued by Rieger⁵²: *“the key audiences you want to prevent from being radicalized are adolescents and young adults. [...] Having this key audience in mind, it is may be more persuasive or more successful if you launch counter messages that will have the same target group as protagonists, as senders, or as distributors of a message, sending or posting a video or message, creating a hashtag or forwarding a hashtag”*. Kruglanski⁵³ puts forward a similar point, arguing that peer influence tends to be high in the age groups that are vulnerable in radicalization.

⁵⁰ GIRDS. (n.d.). *Open letter to our sons and daughters in Syria and Iraq*. Retrieved on 7 September, 2017, from <http://girds.org/mothersforlife/open-letter-to-our-sons-and-daughters>

⁵¹ GIRDS. (n.d.). *A Second Letter to Abu Bakr al-Baghdadi from the Mothers for Life*. Retrieved on 7 September, 2017, from <http://girds.org/mothersforlife/open-letter-to-our-sons-and-daughters>

⁵² Diana Rieger, interview, 19 May 2016

⁵³ Arie Kruglanski, interview, 8 March 2016

Beyond merely acting as a messenger, young people are likely good at developing content that is relevant to and grounded in the lived experiences of their peers (RAN, 2015). One example of is the earlier mentioned P2P project (see Box 2; Chapter 2), which aims to engage college and university students from all over the world to create alternative and counter-narrative campaigns on social media. Many of these campaigns appear to be focused on positive messaging, for instance, to promote tolerance and understanding (Szmania & Fincher, 2017).

It has also been suggested that youth may be reluctant to act as messengers when campaigns are framed in terms of counter-radicalization or countering violent extremism. As such, it may perhaps be more suitable to solicit their participation in online efforts that have ‘CVE-relevant’ goals, such as those that are more generally related to issues of social cohesion or political activism, rather than ‘CVE-specific’ goals (RAN, 2015). Arguably, this may hold for other potential messengers as well.

4.4.2.5. OTHER COMMUNITY ACTORS

There is a wide variety of other community actors that could potentially bear weight with different target audiences (e.g., valued community leaders and organizations, popular figures such as athletes or artists, respected journalists). In particular, religious leaders and associations have often been identified as potential credible sources (Hedayah & ICCT, 2014; Weimann, 2015; Zeiger, 2016). For instance, van Ginkel (2015) suggests that these actors may carry the religious expertise and authority to provide “*alternative interpretations of the Koran and other religious texts in response to extremist interpretations of Islam*” (p. 11). She further adds suggests that “*these actors can moreover directly counter the extremist jihadist narrative*” and “*could fulfil a role in explaining the values of Western society [...] and the ways in which these values can be respected while still living the life of a devout Muslim*” (Ibid.). Kruglanski⁵⁴ adds that that they could potentially make clear that those who become involved with extremism do exactly the opposite of what their religion requires. However, it is essential these sources carry religious authority and charismatic weight among the target audience. Identifying such sources for those that may sympathize with extremist groups prove to be a challenge, and it may well be the case that such actors are outside traditional structures of religious authority in Islam (Zeiger, 2016).

In sum, messages are more likely to strike a chord when they stem from a credible source. But the level of fit also matters. Furthermore, it is evident that the assessments of credibility in online environment is much more complex than in previous media contexts. Researchers assume that messengers that stem from the grass root are particularly effective. Commonly mentioned candidates include former extremists, victims of violence, peers and family, as well as key members of communities and civil society actors. It is also important to point out, that those well suited to act as messengers, often require training and empowerment to be able to effectively carry out this role (see chapter 6). In the next section, we discuss via which channel a message might best be disseminated.

⁵⁴ Arie Kruglanski, interview, 8 March, 2016

4.4.3. DETERMINE APPROPRIATE DISSEMINATION CHANNELS

Of course, no matter how well a message strategy has been given shape, it is unlikely to resort any effect if it doesn't reach the intended audience. Hence, selecting appropriate channels for message dissemination is generally considered to be of vital importance. However, any attempt to provide a concrete suggestion as to what channels may be most appropriate to reach specific target audiences will likely be outdated before publication, given that the social media landscape is constantly changing and evolving, and user preferences and practices tend to change rapidly (Korda & Itani, 2013, Stevens, 2010). Hence, we will suggest some guiding principles that may inform on determining appropriate channels.

In selecting social media channels – or any other vehicle of communication for that matter - the most important consideration is which channels offer the greatest opportunity to reach the intended audience (Ingram & Reed, 2016; Noar, 2011; Stevens, 2010). It is considered sensible to relay the message through channels that are frequently utilized by the audience one intends to reach (Stevens, 2010). It begs little explanation that in order to identify these channels, gathering data on the specified audience's media preferences and consumption patterns is an imperative (e.g., what social media do they use, how do they use them). Sound audience analysis can provide insight in considering the most appropriate outlets for message dissemination (Thackery, Neiger & Keller, 2012). However, given the continuously changing social media landscape, continuous audits are necessitated (Stevens, 2010). It follows that those undertaking an effort should be flexible and able to adapt. Fortunately, today there are a great deal of tools available to ease this process for digital media.

Furthermore, it is advised that program planners consider the distinct qualities of different platforms when choosing appropriate communication channels, while keeping in mind the communication and content strategy being applied (Atkin & Salmon, 2013; Noar, 2011; Thackery, Neiger & Keller, 2012). Social media encompass a diverse spectrum of tools and technologies. Each individual platform has different attributes and features that can be leveraged for a program. In addition to capacity to reach the intended audience, channels may vary on characteristics such as specialization, interactivity, personalization, meaning modalities, depth capacity, accessibility, decodability and credibility (Atkin & Salmon, 2013). Put differently, using Facebook as a communication channel is not the same as using Twitter, and it is important to consider the specific differences when engaging in an effort (Adewuyi & Adefemi, 2016) .

Some scholars suggest that utilizing multiple complementary social media channels at the same time will be beneficial to a program, as it may contribute to greater reach and exposure and may help reinforce messages (Levac & O'Sullivan, 2010; Korda & Itani, 2013). However, it is important that communication activities across communication channels are aligned with each other so that they complement in each other in message delivery (e.g., Kaplan & Haenlein, 2010; Noar, 2011). Furthermore, it is suggested that expanding the number of channels, it may be more difficult find ways to engage users to visit, participate in, and return to these channels (Korda & Itani, 2013). In any case, it is once again imperative that one takes into account the target audience. Running a program well across a few relevant social media channels through which one can actually reach and engage the target audience makes more sense than taking a broad scatterplot approach (Freeman et al., 2015). Moreover, here it also holds, platforms should suit the

communication and content strategy being applied. Put simply, it makes little sense to add a YouTube channel, when one does not intend to use videos.

In deciding upon the channels to use, program planners should also take into account the available resources. As suggested earlier in this chapter (see section 4.4.3), running a campaign effectively requires adequate technical expertise and skills. Moreover, while social media can often be used for free or against limited costs, running a campaign requires resources that may be costly. Of course, such investments are likely to increase when program planners intend to use multiple channels (Freeman et al., 2015; Gold et al., 2012). Furthermore, social media

Furthermore, while the current report aims to specifically focus on how social media might be used for a counter narrative campaign, the question whether social media is in fact an appropriate channel to reach the audience should be carefully considered. The CVE and broader communication literature indicates that not all potential target audiences towards whom communication efforts might be directed are necessarily engaged with social media (Ingram & Reed, 2016; Levac & O'Sullivan, 2010; Stevens, 2010). An overemphasis on these channels ignores members of the target audience that are not avid social media users or lack access to the internet. It may well be that audience members may be more effectively reached through other means such as traditional media (e.g., television or radio broadcast, print media), interpersonal (i.e., face-to-face) communication or events (Ingram & Reed, 2016; Zeiger, 2016). Hence, various scholars across different fields suggest that rather than employing stand-alone social media campaigns, utilizing a coordinated set of communication activities across different communication channels may be more effective (e.g., Ingram & Reed, 2016; Levac & O'Sullivan, 2010; Noar & Head, 2011; Heldman, Schindelar & Weaver, 2013). Such an approach does not only expand opportunities to expose target audiences to messages, it may also compensate for the respective limitations of any single delivery method in isolation (Ingram & Reed, 2016).

The large amount of messages aired through social media is another point of attention which one may wish to complement with other channels or an event. Some literature explicitly points out that it is beneficial to combine online with offline action (e.g., Taylor, 2012; RAN, 2015). Richardson (2013) argues that including an offline face-to-face component may be especially important when a program targets those most 'at risk' of violent radicalization. As online campaigns are not very likely to steer people away from violent paths in and by themselves. Furthermore, she puts forward that those who are most motivated to engage with online extremist propaganda and settings are unlikely to be active and enthusiastic consumers of information relayed through an online counter-narrative campaign. As such, she argues that interpersonal strategies are likely to be far more effective to both reach and influence these individuals.

To summarize, selecting channels for one's (counter narrative) message includes answering the question of how one's program can best reach and engage the intended audience, taking into account issues of access as well as what one aims to achieve. It is also vital to look at the strengths and weaknesses of different communication channels taking into account the message strategy as well as the available resources. Importantly, social media may not be appropriate in all circumstances, and should perhaps be considered a supplement to other program approaches and channels. Linking online with offline activities is also suggested to be important, in particular in case of 'at risk' groups.

Box 10: ‘Oumnia Works’: A local awareness effort using counter-narrative techniques

A highly interesting case to underscore the importance of the choice of the ‘channels of delivery’ is the program of Oumnia Works. This local organization is a The Hague based, local empowerment program, conceived and organized by a group of volunteers and municipal officials in 2015 to support (Muslim) mothers. The stated aim is “*to provide mothers with additional support and input to enable them to help their children to survive in a high demand society*” (Oumnia Works, n.d.) The program explicitly states the importance of mothers in inoculating their children against falling out of touch with society and embarking on a course of radicalization. Together with a community based NGO “Steunpunt Sabr”, the program has developed an interactive training, carried out by around 150 mothers that have received this training beforehand. In this bottom up fashion, expanding circles of Muslim mothers are being involved in discussions, the said training and in helping each other and themselves with respect to practical, educational problems concerning (fears of) radicalization.

Oumnia Works is a The Hague initiative, but exclusively based in the Schilderswijk, a vulnerable area in the city, in which different cases of radicalization hubs have been detected and disrupted. In the Schilderswijk, through Steunpunt Sabr, local mothers are invited for tea, pastries and for social games; often in connection with (obligatory) integration courses. In this manner, Omnia Works is able to reach mothers, via friends, acquaintances, that sometimes do not speak the Dutch language, and would never have registered for such an education course on their own account.

Oumnia Works engages in three types of activities: The training (including role plays intended to identify recruiters, instructions for dealing with radicalizing children, a smart game, discussions on social media awareness and education, identity and citizenship), a helpdesk (open all day and night), and follow up care (for example direct assistance or referral to professional institutions).

Oumnia works is a local empowerment organization, and not a single issue counter-narrative campaign. However, the training explicitly involves social media awareness, recognition of ISIS-propaganda. It questions and confronts mothers about the way their children deal with extremist narratives and postings. Moreover, Oumnia Works is also active on Facebook, through their own website and via their own twitter account, feeding the participants and trainees with up to date, accurate news rapports, refutations of fake news, and warnings against instances of radical behavior in the community. In assisting mothers in ‘early recognition’ activities of social media recruitment for example, the program fuels them with counter-narrative arguments and texts, that they may use in the education of their children.

Oumnia Works is invited to organize incidental sessions within the scope of local integration programs, it gives presentations, and has, since 2016, also started to train mothers in other cities, to organize similar bottom-up programs there. In this, Oumnia Works is a case in

point of real life, physical encounters and de-radicalization activities in combination with social media attempts to counter terrorist and radical narratives⁵⁵.

4.5. IMPLEMENTING A CAMPAIGN

The phase of implementation of a communication campaign involves disseminating the message (4.4.4.1.). Furthermore, in considering dissemination issues, one should be consciously aware of possible risks and limitations in launching a counter-narrative, in particular, adverse actions from those countered (4.4.4.2). Selecting the proper channels is crucial for the effectiveness of the message or campaign. In the following, we describe what can be learnt in this respect from previous research and the various grey literatures.

4.5.1. DISSEMINATING THE MESSAGE

In addition, to giving consideration to *where* the message will be distributed, program planners should also consider *how* they will effectively distribute messages in order to reach the intended audience through selected (social media) channels. A frequently repeated observation is that the dictum ‘build it and they will come’ does not apply to social media (e.g., Gold et al., 2015; Richardson, 2013; Stevens, 2010). Put differently, simply selecting (an) appropriate channel(s) and establishing a social media presence will hardly be sufficient to get the message out. Given the massive amount of information social media users are faced with as well as the speed with which it changes, one of the biggest challenges for any strategic communication effort on social media will be to cut through the clutter and draw and retain attention. As noted by Richardson (2013), rather one should ‘*meet them where they are at*’ (p. 5). In this regard, it has been remarked that it may be more advantageous to join in communities and conversation, rather than to attempt to set up a new one (Stevens, 2010; Thackeray, Neiger & Kelly, 2012). Some social media enable one to see what topics and hashtags are currently trending, which makes it possible to join or hijack popular discussions (Tuck & Silverman, 2016). Of course, one should also be wary of possible drawbacks (see Section 4.4.4.2.).

When starting out in social media, it may be useful to partner up with people and organizations who have already established a strong social media presence and may be able to reach the intended audience. People and organizations that are well connected with the target audience may be willing to promote and support a program and could potentially influence others to further share content or to connect with the initiative (Freeman et al., 2015; Tuck & Silverman, 2016). Depending on the purpose of the program, it may also be a strategy to contact established online outlets which have a following among the target audience, and ask them to feature the program or share program material (Tuck & Silverman, 2016). News media coverage through both mainstream and online channels could also help to get the word out about a social media campaign (Freeman et al., 2015). Whether it is appropriate to engage the media will of course also depend on the nature and purpose of the campaign.

⁵⁵ Beatrice de Graaf: Interviews with Oumnia Works trainsters, participatory observation of a training, February 2017; Oumnia works (n.d.), Retrieved on 15 September 2017, from <http://www.oumniaworks.nl>.

The timing and volume of messages should also be considered. Timing can play an important role in the reach of a message. If messages are posted one a time at which the intended audience is not active on social media, exposure to a campaign message is likely to be low. In line with this, taking into account when audience members are active on social media could potentially increase exposure to messages (Noar, 2011; Fairlie, 2017). Beyond mere exposure, the frequency of exposure also matters for message resonance. An adequate volume of posts may contribute to reach as well as frequency of exposure. However, post to frequently one may risk annoying the target audience with too much posts. Yet, when one posts to infrequently and a message may not cut through the clutter and reach the target audience (Fairlie, 2017).

It is possible to leverage targeted digital advertising to get the message out (Gold et al., 2011; Tuck & Silverman, 2016). Ads on social media provide opportunities to directly reach out to specific audiences on the basis of selected targeting criteria. Tuck & Silverman (2016, p. 31), who provide a practitioner oriented overview of using advertising for counter-narrative work, set out some advantages of using targeted advertising. Key advantages include: (1) It allows campaigners to narrowly target audiences; (2) it allows one to compete in the same spaces as were extremist content circulates (e.g., by targeting specific channels, videos, keywords, followers of specific users etc.); (3) it can help build momentum for a campaign and contribute to engagement; (4) gives access to more in-depth social media analytics and metrics which can be used for evaluation; and, (5) it allows for the testing of different types of content with comparison trials and control groups (point 4 and 5 are further discussed in Chapter 5). However, these authors also put forward several drawbacks. Perhaps most importantly, audiences tend to be less trusting of advertised content than content that is delivered ‘organically’. Also, it may be time-intensive to start with targeted advertising, and it requires sufficient resources. An interesting example of an initiative that leverages the full capacity of targeted advertising is the ‘Redirect Program’, although the methods applied clearly are beyond the capacities of most actors that engage in counter-narrative work (see Box 11).

Rather than simply broadcasting messages, one may do well to embrace the social component of the social media and actively encourage audience engagement (e.g., Heldman, Schindelar & Weaver, 2013; Lovejoy & Saxton, 2012; Korda & Itani, 2013; Neiger et al., 2012; Taylor, 2012; Veale et al., 2015). Engagement can be understood as an active process of interaction between a campaign or a campaign-generated message and members of the audience, for instance, in the form of liking, sharing, commenting, etc. (Dahl, 2014). As argued by Taylor (2012) *“Engagement is core to campaigning via social media, because it not only encourages followers to move from support to action, but an individual’s involvement is visible on their social network profiles and on the feeds of their connections, which enhances the reach of the campaign”* (p. 14).

The broader literature discusses a variety of techniques which are argued to enhance user engagement with campaigns. It is considered important to seed compelling content, including scheduling regular posts, being responsive to user feedback, engaging directly with members of the audience through individual responses, encouraging interaction and conversations by posing questions, soliciting user-generated content (e.g., sharing personal stories, uploading videos or photos), giving acknowledgement for audience members’ support, and linking online communication with offline activities, among others (Heldman, Schindelar & Weaver, 2013; Taylor, 2012; Veale et al., 2015). Providing messages with a clear ‘call to action’ encourage

members of the audience to do something can also help to stimulate engagement (Freeman et al., 2015; Tuck & Silverman, 2016). Also here it holds, which techniques are in fact appropriate will depend on the purpose and target audience of the campaign (Tuck & Silverman, 2016). Moreover, effective engagement necessitates careful monitoring and listening to what is happening on social media channels (see next section). Also, real-time engagement requires social media efforts to be flexible, adaptable, and ideally, able to respond to (real world) events as they happen (Taylor, 2012).

Box 11: 'Redirect Program'

In September 2016, JIGSAW, a google technology incubator and think-tank (previously called Google Ideas), announced pilot program called the 'Redirect Program', which was established in cooperation with Moonshot CVE, Quantum Communications, the Gen Next Foundation and a team of researchers. The program uses AdWord targeting tools to connect those who are actively searching for ISIS-related extremist material to curated YouTube videos that present a countervailing voice. The basic idea is as follows: targeted text, image and video advertisements appear alongside results that are based on search queries that contain terms and phrases that were identified as being suggestive of positive sentiment towards ISIS. These ads link to Arabic and English-language YouTube channels with themed playlists that feature pre-existing videos that are believed to be capable of debunking the ISIS narrative. The videos in these playlists are not necessarily well-known, and are oftentimes not explicitly designed to challenge jihadi propaganda. Rather, they were selected because program developers believed them to be objective in appearance and credible to the specific audience segment. Examples of videos include testimonials of ISIS defectors, statements of religious scholars that explain the disconnection between Islam and ISIS, and citizen journalist' footage that highlight the suffering of civilians, and the failure of the governance structure in ISIS controlled territory (The Redirect Method, 2016). According to Google, the campaign was quite promising, over the course of two months, more than 320,000 people viewed the anti-ISIS YouTube channels and people stayed on these channels longer in comparison to regular viewing patterns. They also found that the clicked through rate was considerably higher than the click through rate was around 75% to 80% higher than that the click through rate of all ads that ran against similar search term 12 months prior the launch of the pilot (Jigsaw, n.d.).

4.5.2. RISKS AND CHALLENGES

Several risks and challenges related to the use of social media for the delivery of a counter-narrative campaign have already been put forward. However, there are several others that deserve consideration, and should be factored in when developing and implementing a social media effort. Below a non-exhaustive overview is provided.

In general, those undertaking a campaign efforts should be wary of potential unintended harmful side effects. Messages intended for specific audiences often end up being viewed by other audiences as well (Atkin, 2004; Egner, 2009; Fink & Barclay, 2013), and there is limited control over how messages are interpreted (Atkin, 2004). Messages that may be effective with a specific intended audience, may be ineffective or even counter-productive with another, unintended

audience (Egner, 2009; Freeman et al., 2015; Fink & Barclay, 2013). For instance, counter-narrative messages aimed at dissuading people from supporting Salafi-Jihadi groups, may cause alienation or be perceived as offensive when they reach people that are not at-risk of falling in the orbit of these groups (Berger, 2016; Reynolds & Tuck, 2016), or may in fact trigger curiosity about the extremist group and its ideology (Berger, 2016; also see section 4.3.2.)

Also, when choosing social media for a counter-narrative campaign, the risk of losing control over the message should be considered. As has been noted by Heldman, Schindelar and Weaver (2013), in the context of campaigns that leverage social media “*concerns about negative comments, misdirecting and reshaping the message and conversation or “online incivility” are valid*” (p. 10). The authors point out it is important that these concerns are considered at the outset of a program, and are factored into the social media engagement strategy. This point may be particularly important for counter-narrative campaigns, which could well evoke negative responses (e.g. hate speech; see Ernst et al., 2017) and reshaping.

In fact, it may even be the case that entire counter-narrative campaigns are turned around and used by the adversary in a counter campaign (Briggs & Frennet, 2014). Aly⁵⁶ argues that this is exactly what happened to the ‘Say No To Terror’ (See Box 1) and the ‘Think Again, Turn Away’ (See Box 12) campaigns: “*The counter campaign became Say Yes to Jihad. And in fact, the counter campaign messages, and the counter campaign video’s outnumber the Say No to Terror videos. That’s the first risk, that there’ll be a counter campaign. The Think Again, Turn Away had counter campaigns within half a day*”. Conversely, Rieger⁵⁷ puts forward “*we just tried to do an analysis on the hashtag “notinmyname” which was built as a counter-voice. But in Germany at the moment, it is caught up by right wing groups which say ‘all these refugees, this is not happening in our name’, like they all use the hashtag “notinmyname”, it’s covered with right wing extremist messages*”. As such, this may well lead people to come across extremist content, which could potentially increase curiosity into violent extremist groups rather than suppress it”. These examples highlight the importance of thinking about such adverse unintended options and learning from others before one launches a campaign.

Box 12: ‘Think again, turn away’

US State Department’s Center for Strategic Counterterrorism Communications (CSCC; replaced by the Global Engagement Center (GEC) in 2016) is particularly well-known for its now-retired public diplomacy initiative #Thinkagainturnaway. Launched in December 2013, the campaign was designed to challenge Salafi-Jihadi extremist groups’ in their ability to accumulate cultural and political capital and focused on refuting ideological claims (Richards, 2016). The campaign aimed to do so in two ways: by providing counter-narrative material in response to online Salafi-jihadi propaganda; and entering in direct conversations with jihadist accounts (van Ginkel, 2015). The campaign has received substantial criticism. In part, because of its second approach of directly addressing and engaging with jihadi accounts. For instance, Rita Katz (2014), director of the SITE Intelligence Group think tank, has argued that trivial

⁵⁶ Anne Aly, interview, February 25, 2016

⁵⁷ Diana Rieger, interview, May 19, 2016

exchanges with Jihadi accounts have frequently served extremist groups, providing them with legitimacy and a stage on which to air their views.

Related to this, it is vital that appropriate precautions are taken to ensure the personal safety and well-being of those who work on or contribute to an online counter-narrative program. Involvement in counter-narrative work could lead individuals or groups to be targeted by violent extremist groups and receive hate mails. Those engaging in such efforts may for instance face (online) harassment or threats, or may even become targeted psychically (Briggs & Frenett, 2014; Hedayah & ICCT, 2014). It has been suggested that former extremists may well be seen as traitors and could quickly become targets if their safety is not guaranteed (Hedayah & ICCT, 2014). Moreover, victims could potentially face abusive reactions and harassment, leading them to be victimized once more. For both of the above examples, it is also vital to keep in mind that recounting distressing events could also evoke strong emotions and may cause psychological damage (Briggs & Feve, 2013; Hedayah & ICCT, 2014). As such, facilitators should ensure that appropriate support and assistance is in place.

In addition, individuals may be reluctant to engage in counter-narrative work, because they fear they may be perceived as a security threat and become subject of law enforcement investigation when they engage with those that may be supportive of violent extremism (Beutel et al., 2016). Hence, it has been suggested that it is important that governments promote a 'legal space' that allows those interested in becoming involved in such work to do so without being perceived as a security threat (Hedayah & ICCT, 2014).

Given the sheer speed and real-time nature of communication, social media campaigns need to be flexible, responsive, and ready to adapt (Taylor, 2012; Hedayah & ICCT, 2014). However, this may pose a challenge for organizations. Freeman et al. (2015) explain that multi-tier-approval processes in organizations can severely impede the ability to respond to and engage with audience members in a timely fashion. This challenge is further exacerbated when one takes into account, that organizations may ultimately be legally responsible for all content posted on one's social media page, including content and comments generated by members of the audience. This once again points to the importance of having adequate human resources available when engaging in such activities.

Lastly, another difficulty in implementing social media campaigns is to demonstrate the value and effect of such social media efforts. While it is possible to gain insight in the online impact, measuring the off-line impact of a counter-narrative is notoriously challenging (e.g., Adewuyi & Adefemi, 2016; Heldman, Schindelar & Weaver, 2013; Taylor, 2012). The difficulties in measuring effects of a specific social media campaign have been addressed will be examined in more detail in the next chapter.

4.5. SUMMARY

In this chapter, we sought to uncover some of the principles of an effective counter-narrative campaign via social media. In general, a social media counter-narrative program will not be highly effective among those who already passed the first stage of radicalization (see Chapter

3). Therefore, the current chapter addressed research question 4 by focusing on a targeted audience because a counter-narrative program may prevent that the opinions and beliefs become fertile grounds for radical thoughts to take root. Thus, counter-narrative programs may be effective in preventing radical messages to resonate among a targeted audience. However, for a small portion of the radical people in second stage of radicalization, mainly those who have already doubts about the group, counter-narratives could potentially have a limited effect as well.

We examined some of the key principles that may guide the development and implementation of a counter-narrative program as addressed in the extant literature to guide an effective communication campaign. Key principles for such a counter-narrative are therefore derived from the general literature on mass communication, health campaigns and persuasion.

In a first section (4.2), we briefly reviewed several preventive CVE campaigns. While the field of CVE is still in its infancy, it is unknown whether these programs can actually be effective at all. In large part, the literature is based on anecdotal evidence and expert opinion rather than thorough case studies and evaluations of existing programs. A robust methodology to assess effectiveness of such programs is lacking, in particular when it comes to (online) violent radicalization processes. There is only very limited knowledge about the relationship between the consumption of extremist content and (online) violent radicalization.

By and large, from a wide array of theories addressing smaller aspects and elements of communicating a message for attitude or behavior change, we derived several key principles that appear relevant to apply in a counter-narrative program, preferably aired via social media. We derived three stages in developing and implementing a communication campaign, each including several aspects to consider. Each of these (sub) sections describe what can be learnt in this respect from previous research in view of the effectiveness of a counter-campaign.

Stage 1) is “Research and planning” (Section 4.3), which includes the importance of using theory as a starting point. This includes theory about the causes of radicalization, as well as theories of how to change the unwanted behavior. In addition, in order to have clearly defined goals and objectives for the counter-narrative campaign, it is vital that audiences are clearly defined, segmented and understood.

In stage 2) “Program development and design” (Section 4.4), first, the content of the message is important in that it should be tailored to the diverse motivations behind radicalization (i.e., people might be motivated to find identity, significance, justice or sensation via the group). Messages that use the narratives have the advantage of using subtle ways to influence people. Moreover, they may overcome various forms of resistance. In addition, double sided messages (in which one side is being discredited), coupled with a strong emotional appeal may have a persuasive advantage.

In addition, at this stage, it is of great the importance of how to select credible sources as messengers in the development of an effective campaign. Commonly mentioned candidates are former extremists, victims of violence, peers and family, as well as key members of communities and civil society actors. The potential of the government to serve as a credible source might be limited.

Stage 3) “Implementation of a communication campaign” (Section 4.5) includes to consciously consider which dissemination channels to choose and how to disseminate the message as well as being aware of possible risks and limitations in launching a counter-narrative, in particular, adverse actions from those countered (Section 4.5.3). It is important to closely monitor potential risks, challenges and limitations when developing a communication campaign. For example, losing control of the message, ‘stealing’ the message into another counter-campaign by opponents and as such creating unintended adverse effects. Also, possible threats to the safety of those who feature in social media counter-narratives should be guarded.

Selected channels should be frequently utilized by the audience one intends to reach. Using multiple channels is suggested to be beneficial. The strengths and weaknesses of different channels of delivery should also be considered, taking into account the communication and message strategy as well as the available resources. Importantly, social media may not be appropriate in all circumstances, and should perhaps at best be considered a supplement to other program approaches and channels. Linking online with offline activities is suggested to be particularly important.

Consideration should also be given to message dissemination. Given that it is particularly hard to cut to the clutter on social media, one way to do so is to join in already on-going ‘conversations’ (Stevens, 2010; Thackeray, Neiger & Kelly, 2012). Also, it may be beneficial to partner up with people and organizations who have already established a strong social media presence and are popular among the target audience. Conversely, it may also be useful to ask established online (news) outlets to share the message. Furthermore, it may be important to create opportunities for audience engagement. Digital advertisement could also help to get the message out, but also has important limitations.

In general, mass communication theories appear to rely quite strongly on traditional one-directional sender-message-receiver models, each directed at some aspect of the chain sender-message-receiver/audience and oftentimes only addressing just one element of a message that might change the direction of effects (e.g., one- or two-sided messaging). Of course, these have their value and provide important guidelines. However, today’s media environments require more dynamic and interactive models including feedback loops and agency of both sender and receiver to do justice to communication processes via social media. In considering such feedback loops and dynamics, the risks of potential adverse effects (e.g., from opponents) also become more clear and can best be addressed before launching a campaign.

An important limitation of a counter-narrative program via social media is that it is notoriously difficult to demonstrate the effectiveness of such social media efforts. This is the topic of the next chapter.

5. DETERMINING THE POTENTIAL EFFECTIVENESS OF A PROGRAM

5.1. INTRODUCTION

In this chapter, we provide an answer to research question 5, namely “*how can the potential effectiveness of a counter-narrative program be determined?*”. This question thus concerns campaign evaluation. Evaluation has been described as the: “*the systematic application of research procedures to understand the conceptualization, design, implementation, and utility of interventions*” (Valente, 2012, p. 106). This definition is particularly relevant as it highlights the importance of engaging evaluation activities throughout the development and implementation program (not just at the end), and points out that research produces should be used. Furthermore, it indicates that evaluation is not just about measuring effectiveness, but also about gaining insight why and how a program works, which can inform the design and development of future programs (Valente & Kwan, 2012). Generally speaking, the effectiveness of any measure or policy against violent extremism can be extremely difficult to assess (e.g., Noordegraaf, Douglas, Bos & Klem, 2016). This is no different for counter-narrative efforts. Measuring changes in attitudes and beliefs is notoriously difficult, whereas measuring prevention (i.e., measuring a negative) is virtually impossible (RAN, 2017).

In Section 5.2, we describe the types of evaluations that are generally recommended in the communication literature (formative, process and summative types). In addition we will describe the realist evaluation paradigm. In Section 5.3, we present the manners in which it is possible to monitor and analyze social media data or metrics, which are often used as an indicator for campaign effectiveness. Section 5.4 describes other online and offline tools and techniques that can be used in the summative evaluation as well as earlier evaluation phases such as focus groups, interviews, surveys, sentiment analysis and social network analysis. Finally, in Section 5.5 we summarize the most important points from this chapter.

5.2. TYPES OF EVALUATION

As argued above, program evaluation should not be thought of as a one-time activity planned and tackled at the end of a communication effort. As we have argued earlier, in the literature on communication campaigns the evaluation process is commonly subdivided in three distinct phases: formative evaluation, process evaluation (also referred to as monitoring), and summative evaluation (e.g., Coffman, 2002; Macnamara, 2017; Noar, 2011; Valente & Kwan, 2012; Paul, Yeats, Clarke & Matthews, 2015). Formative evaluation occurs at an early stage (or prior to) development and design process. While the campaign is being implemented, process evaluation in the form of monitoring activities is carried out. Lastly, an outcome evaluation is conducted, often on the basis of data collected before, during and after a campaign. In the subsequent sections, we shortly outline each of these phases. While they are presented as sequential phases, they involve a set of interwoven activities that inform each other (Noar, 2011; Paul et al., 2015). Each of them requires some degree of planning at the outset of a campaign in order to execute them properly (Paul et al., 2015).

5.2.1. FORMATIVE EVALUATION

Formative evaluation is the collection of information that can inform program planning and shape campaign design (Coffman, 2002). It involves activities that define the nature and scope of the problem, gather data on potential program strategies, learn about the audience and factors that could potentially interfere with program implantation (Valente & Kwan, 2012). This data enables one to set sound goals and objectives, specify the logic of a campaign and provides baseline values for *ex post* comparison (Macnamara, 2017; Paul et al., 2015). It also informs the communication strategy (i.e. channel selection and message design). Formative research also involves pre-testing concepts, messages and other dimensions of the communication strategy (Macnamara, 2017). Message pre-testing may for instance provide insight in how well a message captures the attention of an audience, is comprehended, is perceived to be relevant, elicits (un)intended consequences and so forth (Atkin & Freimuth, 2012). This aspect of formative evaluation shares characteristics of both process and summative research, as it can be utilized to “*examine both the simulated delivery and the effects of a communication strategy and its tactics*” (Freimuth, Cole & Kirby, 2011, p. 80). As such, it can help to assess whether elements in the mix contribute to program objectives and correspond with the information needs of the intended audience. Of course, it differs in the sense that it is carried out before final production and the implementation of a program (Ibid.).

5.2.2. PROCESS EVALUATION

Process evaluation starts with program launch and is carried out as the program unfolds, in simple terms, involves progressive tracking and monitoring to discern whether milestones are being met or not (Macnamara, 2017). As such, it involves activities to determine whether a program is delivered as intended and to assess output measures (e.g., reach and exposure). Process evaluation is considered to be valuable, as it may help to identify problems in program delivery, reasons that underlie those problems, as well as potential solutions. This allows those undertaking the program to make adjustments (Paul et al., 2015). Process evaluation is a recommended precursor to summative evaluation, as it can be difficult to interpret why a program did or did not produce expected outcomes without some knowledge of the levels and conditions of implementation (Flay, 1987).

5.2.3. SUMMATIVE EVALUATION

Summative evaluation is conducted after the program has been implemented and involves what is arguably most often considered to be 'evaluation' (Atkin & Freimuth, 2012; Coffman, 2002). Summative evaluation consists of activities to assess outcomes that can be attributed or tied to the program and to establish whether the effort has achieved its objectives (Freimuth, Cole & Kirby, 2011; Paul, 2015). This may involve assessments of short-term outcomes (i.e., 'outcome evaluation') as well as long-term impacts, “*measures of the ultimate aggregate results of the campaign's outcomes*” (i.e., 'impact evaluation'; Coffman, 2002, p. 24).

5.2.4. REALISTIC EVALUATION METHOD

Generally speaking, the effectiveness of any measure or policy against violent extremism can be difficult to assess (e.g., Noordegraaf, Douglas, Bos & Klem, 2016). Other authors have noted the difficulty of testing the effectiveness of a certain campaign (or intervention or program) and

have developed ideas how to deal with this concern. An important contribution in this domain is offered by the Realistic Evaluation Method, as conceptualized by Pawson & Tilley (2004). They argue that it may not be possible to determine the overall effectiveness of a campaign. In response to this realization, they have developed a method to derive from theories more specific (rather than general) hypotheses about which aspect of a campaign might work for whom, in which context and how. Thus, they argue that “*Realist evaluations asks not, ‘What works?’ or, ‘Does this program work?’ but asks instead, ‘What works for whom in what circumstances and in what respects, and how?’*” (Pawson & Tilley, 2004, p. 2). As such, they argue that it is possible test specific elements of a campaign, in a specific context, and for specific people. This implies that researchers can no longer claim to test the overall effectiveness of a campaign, but to test a specific part. For example, one might test whether the introduction of a video from ex-members of a radical Islamic group, containing a discussion of their disappointment with their past group membership, may lead young people (age 18-21 years) in The Hague in The Netherlands to decrease their interest in this group.

At the core of this method a series of Contexts-Mechanism-Outcome patterns are being described and tested (Gielen, 2017). To link this to the previous example, a context might be radical youth in The Hague, a mechanism might be to give them a video with ex-members discussing their disappointment, and an outcome can be a weakened interest in this group. Such a specific conclusion from a C-M-O pattern may be linked to the overall effectiveness of an online counter-narrative campaign (in particular when one is able to test multiple specific C-M-O’s), but is arguably more humble in its pretentions.

5.3. SOCIAL MEDIA MONITORING AND ANALYTICS

Digital footprints left by campaign audiences have become popular indicators of the impact of campaigns on social media (Shi, Poorisat, & Salmon, 2016). Such measurements center on the analysis of metrics that can be collected real-time from social media platforms that are used in a campaign (Reynolds & Tuck, 2016; RAN, 2015). Many social media platforms provide associated social media analytics services that allow for the real-time tracking and monitoring of social media data at no costs or limited costs. More professional analytics tools have also been developed by third parties, although their level of sophistication is generally reflected in the price tag (usually a subscription fee) (Neiger et al., 2012). Importantly, though, it is not always clear how these services log and define the metrics they present, thus they can be misleading. Hence, one should make sure to gain insight in what a particular metric represent (Tuck & Silverman, 2016).

There are a wide variety of quantitative metrics that can be analyzed. However, many of these metrics do not help us make decisions or meaningfully gauge performance on key outcomes (e.g., so called ‘vanity metrics’) (Paul et al., 2015). Therefore, it is vital to establish at the outset of a campaign what the program objectives are and what you want to learn, identify the key metrics that are considered to be central for success or can help to answer key evaluation questions, as well as how they should be analyzed (Reynolds & Tuck, 2016). Previous work in this domain identified several metrics which can be used to monitor a campaign and provide insight in its online impact. For instance, the Institute of Strategic Dialogue (Reynolds & Tuck, 2016; Silverman et al., 2016) provide some useful guidance as to which metrics may be worth tracking when one aims to gauge the fidelity and the impact of a program; They broadly distinguish between two categories of metrics: awareness metrics and engagement metrics.

Awareness metrics essentially provide an indication of the amount of people that were exposed to a campaign, viewed campaign material, and provide a crude picture of the demographics and geography of the audience. They offer several examples of commonly measured quantitative metrics related to awareness, albeit it should be noted that the definitions of each can vary between different platforms (this holds for engagement metrics as well). Examples include the number of impressions, reach, video views etc. Awareness metrics can also include aggregate data related self-reported audience demographics and characteristics (e.g., gender, age, language, interests, etc.) and geographic location. The extent to which platforms and services allow you to gain insight in such data varies. There are also privacy related limitations, it is for instance impossible to gain insight into details of individual audience members. Awareness metrics alone are inherently limited in their utility, as they provide no indication whether any substantial and meaningful engagement with a campaign has taken place (Reynolds & Tuck, 2016)

Engagement metrics can provide some additional insights in that regard. Engagement metrics provide an indication of the volume, types and quality of user interactions with a campaign. This includes quantitative metrics such as the amount of likes, shares, viewer retention rates, comments and interactions etc. Generally speaking, actions that require more effort of the audience can be regarded as more substantial forms of engagement than those that require less effort (Reynolds & Tuck, 2016; also see Neiger et al. 2012 and Freeman et al., 2015). Several forms of engagement provide qualitative data which can be monitored and allow for further analysis. When campaigns spark comments and conversations, audience-generated messages can be coded

and analyzed for sentiment or tone, key themes and topics and so forth. Among other things, this can give an indication of what people think or feel about the program and its content, and what themes and topics participants are most likely to engage with (Reynolds & Tuck, 2016). This can be done through manual analysis or more sophisticated approaches using data mining techniques and natural language processing (see section, 5.4.6.).

It is suggested that awareness and engagement metrics can come together to provide an indication of the online, short-term impact of a campaign as defined in the objectives (Reynolds & Tuck, 2016), and can potentially provide useful insights to improve targeting and communications (Kim, Lee, Marguleas, & Beyer, 2016). Of course, engagement in the form of likes, shares, comments does not necessarily indicate that potential extremists have been redirected. That is, whether it actually reaches the intended target audience and is changing people's attitudes or behaviors (ibid., 2016). As argued earlier, establishing an offline impact on attitudes and behaviors is extremely difficult. The anonymity that the internet provides often makes it impossible to establish precisely who engage with the campaign. Furthermore, disengagement and de-radicalization processes are complex, and may take place over a longer period of time. According to the ISD, *sustained engagement* (i.e., longer interactions) or people reaching out to campaigners for assistance may perhaps provide the clearest indication that a campaign is having any campaign, at least, on a personal level (Silverman et al., 2016).

5.4. OTHER EVALUATION TOOLS AND TECHNIQUES

While social media metrics have become popular indicators of the impact of campaigns on social media, developing a better understanding of the relationship between online and offline interaction requires the triangulation of methods (RAN, 2015). There are a wide range of online and offline qualitative and quantitative tools that can help, and may provide valuable information that can enhance a campaign. Below we describe several qualitative and quantitative approaches that are outlined in the literature.

5.4.1. (ONLINE) FOCUS GROUPS & INTERVIEWS

Briefly, focus groups interviews have been described as '*planned discussion led by a moderator who guides a small group of participants through a set of carefully sequenced (focused) questions in a permissive and nonthreatening conversation* (Krueger & Casey, 2015a, p. 506). They are commonly conducted with groups of approximately 5 to 12 respondents (smaller groups are recommended for sensitive topics) which have something in common that relates to the topic or issue of conversation. In relation to counter-narrative campaigns, this might for instance be individuals that are representative of segments of the intended audience or subject-matter experts (also see section 5.4.8.1.). The discussion is facilitated by a skillful moderator who encourages participants into discussing the topic in greater detail using focused questions, keeps the discussion on track, and ensures there is permissive environment in which participants feel comfortable to share their views freely (Krueger & Casey, 2015a; 2015b). They are often used to garner information in the formative evaluation phase, but can be useful in later stages as well. Just to name a few purposes, they can be used to garner insight in the problem situation (e.g., to

understand the audience's perspective on a particular problem or issue) and garner other input that can inform the program design. Focus groups are also frequently used to pilot-test various dimensions of the communication and message strategy, for instance to check whether messages are understandable and appealing (Kreuger & Casey, 2015b; Valente, 2002). However, the utility of focus groups is not limited to the formative evaluation. For instance, they can also be used to conduct when the program is up and running. For instance to gain insight in how the program can be improved and what does and doesn't work. Focus groups can also be setup to discuss the campaign and its effects and can be used to interpret and contextualize quantitative data gathered for summative evaluation (Kreuger & Casey, 2015a; 2015b; Weinreich, 2011).

Focus group sessions have several benefits, they are a relatively cost-friendly, quick, socially oriented and have face validity (Valente, 2002). Moreover, they provide a rather flexible research method, whilst they can be conducted in a laboratory setting, they can likewise be carried out 'on location' (e.g., a convenient environment, or an environment in which participants feels safe), via telephone, and online (Atkin & Freimuth, 2012). However, they also have drawbacks. For starters, results are difficult to analyze and outcomes can be unpredictable (Kreuger & Casey, 2015b; Valente, 2002). The primary risk, however, is that focus groups can provide misleading information. As explained by Valente (2002), the group may be overly influenced by one or several people, people may experience normative pressures which may make them reluctant to deviate from group standards, or the topic of conversation may be too distant from audience members to generate opinions. Furthermore, as with any other described qualitative method (e.g. interviews), results lack generalizability. Hence, one should be wary not to quantify results (Atkin & Freimuth, 2012). Another challenge is that they might be particularly difficult to arrange. This may hold in particular in the context of a counter-narrative effort. Of course, it may be difficult if not individuals targeted by counter-narrative activities may be hard to reach and reluctant to participate in focus groups, for instance, because they want to remain anonymous (Dawson, Edwards & Jeffray, 2014; Reynolds & Tuck, 2016). Furthermore, soliciting participation of the target audience, may not be possible depending where they are situated on the radicalization spectrum (RAN, 2015).

In this regard, online focus groups may offer some advantages over their face-to-face counterparts. Such focus groups can be conducted using synchronous (e.g., instant messaging, chat-rooms, or video-conferencing) and asynchronous (e.g., online discussion boards or forums). Depending on the communication method being utilized, online focus groups allow for greater anonymity than face-to-face groups (Atkin & Freimuth, 2012; Krueger & Casey, 2015b; Morgan, King & Ivic, 2011; Stewart & Shamdasani, 2014). Which could potentially make people more willing to participate, and could make them feel more at ease, which can result in greater disclosure (Morgan et al., 2011). There are several additional benefits. Besides that they are particularly resource-friendly, they also make it possible to interact with hard-to-reach populations, in part, because they do not suffer from geographical restrictions. In relation to online settings that allow for asynchronous communication, it has also been observed that they allow participants more time to reflect and reply at a time that is convenient for themselves (Krueger & Casey, 2015b; Morgan et al., 2011; Stewart & Shamdasani, 2014). When focus-groups discussions are text-based, it also provides an instant transcript (Atkin & Freimuth, 2012).

Of course, there are also drawbacks. Among other things, online focus group sessions typically reduce the in-depth emotional information that can be obtained when one is physically in one location with participants (Atkin & Freimuth, 2012). When the moderator and participants do not see or hear each other, information present in facial expressions and tone-of-voice etc. are lost (Stewart & Shamdasani, 2010). Reliability of the quality of recruited participants may be questionable given the absence face-to-face interactions and a lack of screening (Stewart & Shamdasani, 2014). Due to increased anonymity, it may be difficult to establish who did and did not participate in online focus groups (Morgan et al., 2011). Also, certain text-based communication methods may favor participants who type and reply quickly, allowing some to 'dominate' the conversation (Krueger & Casey, 2015b). Synchronous focus groups in particular, may require access to good bandwidth, which may be a problem in some cases and locations (Stewart & Shamdasani, 2014).

5.4.2. ONE-ON-ONE INTERVIEWS

One-on-one, in-depth interviews with target audiences can be a rich source of information for evaluators. They can be used for similar purposes as focus groups, for instance, to learn about the context in which the program's messages and materials were received and how they were interpreted (Weinreich, 2011). Interviews can be conducted face-to-face, via telephone or through online means (e.g., Skype). During an in-depth interview, an interviewer essentially asks a series open-ended questions about a particular topic utilizing a more or less structured format. These interview questions are open-ended in order to give respondents freedom in answering questions, allowing them to express views that relate to the topic of interest (Valente, 2002).

It is suggested that interviews can be particularly useful to discuss sensitive topics people might feel uncomfortable to discuss in a group or that require deep probing (Atkin & Freimuth, 2012). In regard to face-to-face interviews it is suggested that they can potentially increase access and can make it easier to build rapport with respondents. In part, because they can be conducted on locations that are convenient and comfortable to them dynamics. Of course, in the context of counter-narratives, interviews face may still face similar challenges as focus groups when it comes to soliciting the participation of target audiences.

A benefit of interviews in comparison to focus groups is that they avoid earlier mentioned challenges associated with normative pressures and other undesirable influences related to group dynamics (Freimuth, Cole & Kirby, 2011) . However, at the time it should be noted that group dynamics do help people to explore and clarify views in ways that would be less easily accessible in the context of a one-on-one interview (Kitzinger, 1995). Some of the drawback of interviews include that they may be time-consuming and resource-intensive (Atkin & Freimuth, 2012).

5.4.3. (ONLINE) SURVEYS

Surveys are a useful and efficient means for gathering formative evaluation data for multiple purposes. Among other things, they can be used to obtain a quantitative estimate of important (theoretical) variables (i.e., current attitudes, perceptions, behavior, etc.), establish baseline data that permits the measurement of the effects (or lack of effect), and to pre-test concepts and rough messages (Atkin & Freimuth, 2012; Paul et al., 2015). They can also be used

in later evaluation stages to gain insights in the proximal effects of a campaign (Reynolds & Tuck, 2016). A survey can contain both closed and open-ended questions and can be used to gather both quantitative and qualitative data (Atkin & Freimuth, 2012; Reynolds & Tuck, 2016).

Surveys can be conducted in various ways, each method has various advantages and disadvantages. One way is through online means. They commonly make it more easy and relatively cheap to produce and distribute a survey, can be used to reach and gain insights from geographically dispersed respondents, and have the *potential* for fast data-collection and turn around (Atkin & Freimuth, 2012; Sue, & Ritter, 2012; Austin & Pinkleton, 2015; Wright, 2005). During the implementation stage of a program, they can also be easily promoted through links embedded in material, allowing for data-collection as the program unfolds. Furthermore, social media also allows for the delivery of surveys through social media advertising, which enables targeting particular audiences (Reynolds & Tuck, 2016; Austin & Pinkleton, 2015).

However, online surveys also have well-documented limitations (e.g., see Bethlehem, 2010; Eysenbach & Wyatt, 2002; Austin & Pinkleton, 2015, Wright 2005). For instance, one limitation is that online surveys frequently suffer from a self-selection bias. Many online surveys rely on respondents selecting themselves for a survey, which means they may not be representative for the members of the audience (Bethlehem, 2010; Reynolds & Tuck, 2016, Wright, 2005). To illustrate, it may be the case that those who respond may be particularly inclined to do so, and may generally have more favorable attitudes towards the topic interests. While others may be unwilling to share their views on a topic as sensitive as counter-narrative campaigning, or may be inclined to respond negatively. Those who are further along the pathway of radicalization, may arguably be difficult. In addition, in relation to counter-narrative campaigns it may well be the case that potential respondents may not want to share any personal information (Reynolds & Tuck, 2016). Of course, one could offer the option to fill in surveys more or less anonymously. However, this may also increase the likelihood of hoax responses (Eysenbach & Wyatt, 2002).

Low response and high non-completion rates are another well-known concern when conducting online surveys (Eysenbach & Wyatt, 2002; Austin & Pinkleton, 2015; Wright, 2005) and can lead a sample to be unrepresentative of the audience. While participation in a survey can be incentivized– for instance through a price draw (Reynolds & Tuck, 2016), incentives may however also impact on the representativeness of the sample (Eysenbach et al., 2002). Furthermore, efforts to increase participation may also increase the amount time and resources necessary to conduct a survey, which will be reflected in additional cost (Austin & Pinkleton, 2015).

In sum, while online surveys can potentially provide a useful tool to gathering a greater depth of data during different stages of evaluation, they pose a variety of methodological problems. Consequently, online surveys should be used carefully and results should be interpreted with care (Austin & Pinkleton, 2015).

5.4.4. EXPERIMENTS

Essentially, an experiment is an examination of a cause-and-effect relationship between an independent and a dependent variable. The basic purpose of an experiment is to determine causation. In a conventional type of experiment the influence of an independent variable on a dependent variable is typically examined by making an explicit comparison between at least two

similar groups (of individuals): one group (the treatment group) is exposed to an experimental stimulus, while the other group receives no stimulus or perhaps a meaningless one (the control group). The impact of the stimulus can then be assessed by comparing what happens when it is present to what happens when it is not present (Ackland, 2013; Baxter & Babbie, 2004).

Experiments can roughly be classified into two categories depending on whether they are conducted in the laboratory or in the field. For example, in a laboratory experiment, one group of people could be subjected to a counter-narrative message coming from a former radical, while another group of people do not receive a counter-narrative at all. Subsequently, people's radical attitudes are assessed. This design can test whether presenting a counter-narrative message or not (the independent variable) has an effect on the radical attitudes (the dependent variable).

In order to be able to draw causal conclusions, two elements are crucial: first, people have to be randomly allocated to one of the conditions and secondly, it is important that the independent variable differs only on the dimension of interest, and does not co-vary with another variable. Thus, for example, it is not possible to draw causal inferences from a comparison between a message sent by former radical male and a message sent by a female victim of a terrorist attack, because the properties of this message differ on 2 dimensions at the same time (i.e., gender and "status": former radical vs. victim).

In field experiments, researchers use existing groups and test differences. For example, people in Amsterdam receive a different message than people in Rotterdam. You can still test whether people react differently to the two messages, but the strength of the conclusion is hampered by the fact that people from Amsterdam and Rotterdam may have differed from each other before the message already. Thus, the ability to draw strong causal conclusions from a field experiment is limited.

Interestingly, it is also possible to do experiments online. Several computer programs offer this option (e.g., Qualtrics, M-Turk, etc.). In addition, some research companies offer online samples via "panels", sometimes even representative of a more general population (for a discussion of the quality of the data see Goodman, Cryder, & Cheema, 2013; Mason & Suri, 2012; Paolacci, & Chandler, 2014).

This has several advantages for the current question. It makes it possible to test materials online, for example inspecting the number of views of a message, or by asking people's responses to certain materials (i.e., pilot testing of materials).

The down side of online experiments is that the researcher has no (or limited) control over the situation in which the respondent answers the questions. For example, it might be possible that people get distracted during the study (e.g., their telephone rings, they make a cup of coffee, etc.). In addition, people might be doing the study together with someone else who might influence their responses and thus the rule of independence of the answers is being violated. For these reasons, it is usually advised to have higher number of participants in on-line studies than in off-line studies.

5.4.5. NETNOGRAPHY

'Netnography', a contraction of Internet and ethnography, is essentially ethnography adapted for the analysis of social media and allows in-depth study of smaller online communities (Bartlett & Reynolds, 2015; Kozinets, Dolbec, & Earley, 2014). Netnography utilizes elicited and – more frequently – non-elicited data, which is typically obtained through observation of and/or participation with people as they interact in regular online and social media environments and activities (Kozinets, Dolbec, & Earley, 2014). However, ethnographies may also employ a range of other data collection methods - including many of those outlined in this chapter (although manual methods are most commonly used and numerical measurement or quantification of social media data is typically avoided) (Bartlett & Reynolds, 2015 ; Kozinets 2015). Data collected and analyzed may take a wide variety of forms, ranging from words, images, audio, video etc. (Kozinets, 2015). In marketing and consumer research, the field of origin, ethnography is typically used in formative evaluation to develop a deep understanding of online groups and communities. For instance, it can gain insight in attitude formation, behaviors, conversations, and interaction and discursive styles of specific online consumer groups. Such information can subsequently inform strategy development and message design (Belk & Kozinets, 2017; Kozinets, 2015). It could potentially also be useful in later evaluation states, for instance to gain insight in how audiences respond to communications (Belk & Kozinets, 2017).

Netnography shares several characteristics of traditional ethnographic research, such as being immersive and naturalistic, however, it is less time-consuming, simpler, less expensive and treats online communities as a primary fieldwork site (Kozinets, 2010). Netnography is claimed to have various merits. For instance, it is argued to have the *potential* to be less obtrusive than other qualitative data-collection techniques such as focus groups and interviews as it is conducted within a context that is not fabricated and because it allows for the collection of data without making ones presence visible (Kozinets, 2002; Kozinets, Dolbec & Early, 2014). Therefore, ethnography is regarded by some authors to be particularly useful for studying personally and politically sensitive or controversial topics and illegal acts discussed in online communities by users who prefer to conceal their personal identities (Costello, McDermott, & Wallace, 2017). Perhaps, therefore, netnography is increasingly viewed as a valuable method to gain insight in extremist communities (Bartlett & Reynolds, 2015; Conway, 2017). A study of Hegghammer (2014), which explores the effects of trust issues on jihadi internet discussion forums, provides one example in which this method is applied.

Of course, there are also drawbacks. For instance, Kozinets (2002) asserts that “*the need for researcher interpretive skill, and the lack of informant identifiers present in the online context that leads to difficulty generalizing results to groups outside the online community sample*” (p.3). Furthermore, netnography may draw on data that is not given explicitly and in confidence to netnographers. In some cases, this may raise ethical and legal concerns. Hence, it important to give consideration regulatory and legal requirements and ethical related to issues such as privacy, anonymity, informed consent and risk of harm. Of course, in general it is important to consider potential ethical and legal issues resulting from evaluation design decisions. Perhaps in particular when it involves social media research (see Moreno et al., 2013; Zimmer, 2010).

5.4.6. SENTIMENT & CONTENT ANALYSIS

Storage and retrieval of information on social media, along with datamining techniques, makes analyses of a huge amounts of audience-generated messages possible. Automated sentiment analysis techniques can be used to analyze such data and infer the sentiments of social media users towards some topic. In a nutshell, sentiment analysis, also referred to as opinion mining deals with “*the computational treatment of opinion, sentiment, and subjectivity in text*” (Lee & Pang, 2008, p. 1). It makes use of computational linguistics, natural language processing (NLP) and other methods of text analytics to automatically detect and extract people’s sentiments or opinions towards certain entities (e.g., issues, events, individuals, groups, organizations) and their attributes from text sources (e.g., in the context of social media, comments, tweets, blog posts, etc.). While still in its infancy, it has become an increasingly popular way for companies to determine what sentiments and opinions are expressed regarding their products, services, brands, and campaigns (Cohen, Johansson, Kaati, & Mork, 2014). Among other things, sentiment analysis can provide insight into how a campaign is received and perceived by those who engage with it, adding quantitative insight to other data that has been gather as part of the evaluation process (Reynolds & Tuck, 2016).

A basic purpose of sentiment analysis is to identify opinions, identify the sentiments they express, and then classify polarity or opposing views (Medhat, Hassan, & Korashy, 2014). That is, to classify any given text as either positive, negative or neutral, the software needs keywords to search for relevant text and assess such differentiation. In general, before starting the automatic analyses, specific analyses of the messages’ content are done manually to identify those keywords and check coverage and classification. This process is iterative as it needs improvement to reach good inter-coder reliability as well as reliability once automatized (e.g., precision and recall; Riff, Lacy, & Fico, 2014) . Recent options for sentiment analyses also go beyond polarity, for instance, it may be employed to detect and analyze emotions emotion carried by a text. Sentiment analysis can be conducted at different levels of granularity, including the document, sentence, word, and aspect/feature level) (Medhat et al., 2014). Of course, social media contain many forms of multimedia, and sentiment may arguably not only be conveyed through words. For instance, it could be expressed in photos or emoticons. More research is being conducted on how to analyze sentiment that is not directly being expressed in texts (Hui, 2016).

As an example serves an analysis of how Muslims were portrayed in the news for which seven Dutch newspapers, fourteen news broadcasts on television, and six relevant websites were coded for positively and negatively toned content. Several thousands of messages have been searched by means of automatic content analyses using hundreds of search terms, in several steps, starting with mapping the past years for news about immigration and integration (from 1998 until 2008). The results showed how the focus had shifted over time since the attack on September 11 in 2001. Then, the debate took a turn and hardened the tone. Terrorism became the leading item on the media agenda and became associated with the Islam. Interestingly, since 2006 the attention for terrorism decreased but from 2007, the media attention for political anti-immigration parties in The Netherlands increased considerably. Furthermore, findings of comparing seven digitally available newspapers (from October 2006 until December 2008) showed that Islamic residents and immigrants were often related to terrorism, criminal acts, war violence, and ideological extremism, in those news messages. In addition, results of a comparison of six web fora (popular among primarily Moroccan Muslim youth) showed that the topics of these web fora clearly follow each

other in showing quite some overlap. The discussions on the web closely followed the debate in Dutch news media. Finally, (manual) analyses of 14 television news and information programming (focusing on a specific Muslim-related event during seven weeks) showed how a specific selection of facts colored the news within a specific frame (e.g., problematic youth ‘terrorize’ the neighborhood; ‘street terror’) showing a negative tone of the debate.

Although automated sentiment analysis is a hot research topic, it is still at an early stage of development. There are several important challenges. Perhaps most importantly, there are currently few proven methodologies for extracting and analyzing social media content. Although technological know-how is improving, extracting relevant texts for analysis from huge amounts of data remains a significant challenge. In general, there are several operational and practical challenges and limitations when it comes to collecting social media data, this includes issues related to matters such as volume, access, and, as earlier mentioned, ethics (see Hui, 2016 for a general overview). The informal nature of online communication also poses difficulties. For instance, some problems with sentiment analysis are related to ‘noisy’ texts (e.g., slang, spelling, grammatical errors, problematic and missing punctuation) and the presence of implicit sentiments (e.g., sarcasm, irony), and identifying entities and context (ibid, 2016) .

Still, despite the difficulties, some efforts have been made to establish a method via sentiment analysis to find radical people on-line. For example, Scrivens, Davies and Frank (2017) present what they label a ‘Sentiment-based Identification of Radical Authors’, which entails a combined score for a user’s percentile average sentiment score, volume of negative posts, severity of negative posts, and duration of negative posts. Another example is provided by, Bermingham, Conway, McInerney, O’Hare, and Smeaton (2009) they use sentiment analysis alongside social network analysis (next section) to explore online radicalization. Given the fact that these techniques differ in their focus, it still is challenging to directly combine them, rather than to treating them as two separate outcomes of two separate analyses. Furthermore, more thorough research is needed to establish the predictive value of such sentiment analyses to find radical people via their on-line activities.

Related to sentiment analysis is to conduct more complex content analyses, for example, to study the dominant ‘frames’ in a discussion on social media, or to get an impression how a campaign is discussed by its followers. Such analyses provide richer detail about the content of their responses than merely positive or negative content. A relevant example is an extensive content analysis of blog posts of a Dutch website where Moroccan youth and (young) adults often discuss Islam-related issues (i.e., Marokko.nl) (Mann, Doosje, Konijn, Nickolson, Moore, Ruigrok, 2015). In applying four types of qualitative and quantitative content analyses (i.e., bottom-up and top-down, both manually and automatically), posts could be identified as either representing an ‘extremist radical frame’ or a ‘counter-frame’ based on its features referring to a specific ‘frame’. In the initial stages (mainly by manual content analyses), seven frames were identified, four of which characteristic for extremist messages/posts (e.g., ‘hate against the West frame’; a pro-Islamic State frame; a ‘fight for Syria’ frame) and three frames were characteristic for a counter-message of offering resistance against radical and violent accounts (e.g., a democracy frame; an unacceptability of ISIS-acts frame).

The results of the analyses showed, among others, that the posts on the web forum Marokko.nl in large part contained messages within a counter-frame relative to posts expressing a radical-extremist frame (in an overall corpus of 149.394 posts from 15 September 2013 to 15 September 2014; Mann et al., 2015). Based on the relative percentages of extremist and counter-frames, a so called resilience index could be calculated, per week or per month, to show changes over time (see Figures 5.8. and 5.9 in Mann et al., 2015). It shows, for example, that the extremist voices clearly increased in July 2014, while the counter-voices showed more fluctuations throughout with a dip during the last week of July 2014. In this month, the resilience index showed a reversed pattern in the debate; the amount of extremist posts was larger than the amount of posts from counter-frames. Results like these need more follow-up research to allow more solid interpretations as well as to increase the reliability of the automatic content analyses.

For the automatic content analyses, Mann et al. used the Amsterdam Content Analysis Toolkit (AmCAT, van Atteveldt, 2008). Although this software has been extensively tested, for example on news texts, intensive manual and qualitative analyses are still needed for such 'loose' language use as expressed on social media (e.g., including a lot of slang, typo's, unusual terminology) including many repetitions that should be filtered out. A drawback of the automatic content analyses, specifically with such materials, is its lower reliability (cf. standards in Stryker, Wray, Hornik, & Yanovitzky, 2006). To increase reliability of results, deleting mere repetitions, but also separate older posts from newly added responses and posts would help. However, this requires time and human resources, and thus financial resources. Related, analyses can be restricted to the newly added posts alone. Alternatively, one might select the most active individuals and specifically analyze their posts over time as well as how others respond specifically to this individual. Manual and automatic analyses might further be applied in selecting those who mostly post within extremist frames in comparison to persons typically posting within a counter-frame. Such a dialogue may provide valuable input for designing a counter-narrative on social media.

5.4.7. SOCIAL NETWORK ANALYSIS

Social network analysis (SNA) is all about relationships, and relationships are at the core of social media. In short, SNA refers to a set of theories, methods and techniques that can be used for the collection and analysis of network data (i.e., relational data), which has been described as "*information on the connections and relationships among and between entities*" (Valente & Pitts, 2017, p. 104). Correspondingly, basic units of analysis are 'entities' - also referred to as 'nodes' or 'actors' - and 'ties' which represent connections or relationships between entities (Wasserman & Faust, 1994). These entities can be people, but they may also be groups, organizations and other connected entities. Ties can represent friendship, common interest, transaction, interaction, affiliation, among others (Wasserman & Faust, 1994). Network data can be gathered online, offline, and combined. Advances in computing technology and the rise of social media have increased opportunities for social network analysis and have made it easier and cheaper to derive social network data-sets at scale (Hui, 2016; Valente & Pitts, 2017). Social network analysts use numerous mathematical techniques and a variety of tools to provide quantitative and visual representations of the properties of a network and the entities that comprise it (Bartlet & Reynolds, 2015). Various measures have been developed over time, for example, to represent interaction

among actors, identify how influential each actor is, detect subgroups and communities, examine strong or weak ties and find the network topology and strength (Choudhary & Singh, 2015).

A commonly examined notion is ‘centrality’ which gives insight in: “*the extent a person occupies a prominent or important position in the network*” (Valente et al., 2015, p. 4). Several measures have been developed to arrive at descriptions of centrality, including, ‘degree centrality’, ‘betweenness centrality’ and ‘closeness centrality’ (Choudhary & Singh, 2015). ‘Degree centrality’ counts the number of direct links of an entity to other entities. Put differently, it helps identify the persons with the greatest number of connections, which often are key influencers in a network. ‘Betweenness centrality’ measures the proximity of an entity to other entities in a network. High-betweenness nodes may indicate the person is an important gatekeeper or ‘bridge’ that connects disparate regions of a network, and may therefore have a high degree of control over the flow of information users in a specific part of the network. Finally, ‘closeness centrality’ is measured as the mean length of distances between an entity and all other entities in a network. Low-closeness may indicate it may be difficult to communicate (Bartlett & Reynolds, 2015; Choudhary & Singh, 2015).

It is suggested that social media SNA may be a promising tool for variety of evaluation purposes. To give a glimpse, SNA techniques can be used alongside other methods to identify and develop an understanding of online communities that are supportive of violent extremism and how they evolve (Fairlie, 2017; Marcellino, Smith, Paul, Skrabala, 2017; van der Hulst, 2008). They can be used to gain insight in how and what information flows among people in a network, and help reveal the structure through which influence is exerted (Waldman & Verga, 2016). The use SNA may also provide information that can help distinguish within the target audience, for instance to separate those that more engaged with extremist groups from those that are merely curious in extremist material, which may benefit targeting (Berger, 2016). Likewise, SNA techniques may help to discern the most influential members in a network. In this regard, a relevant example is provided by Carter, Maher, and Neumann (2014). They used an approach grounded in SNA to identify the most prominent sources of information and influence among Western foreign fighters. Similarly, it could be used to identify those most capable of swaying supporters away from extremist groups and to discern other roles in social networks such as ‘bridges’ (Fairlie, 2017; Marcellino, Smith, Paul, Skrabala, 2017).

SNA is an emerging field and applying it to social media presents a variety of challenges. One of these challenges is defining the boundary of the network: “*Analysts need to make informed decision in determining which links, seed accounts or websites are to be targeted for analysis, for this is where the nature of the network in question is hinged*” (Hui, 2016, p. 333). Furthermore, as suggested in the previous section, there are several other operational and practical challenges and limitations when it comes to collecting relational data from social media (ibid.).

5.4.8. OTHER EVALUATION METHODS

5.4.8.1. EXPERT AND GATEKEEPER REVIEWS

Eliciting input and judgement from subject-matter experts and professionals in the field as well as intermediary organizations. Experts could for instance include those with considerable

knowledge of the intended audience and the subject matter of violent extremism, those with expertise and experience in fields such as communications and technology, or people who have experience in counter-narrative campaigns. Eliciting expert judgment can take a variety of forms in different phases. One area in which experts can fruitfully contribute is pre-testing. Expert reviews are argued to be particularly beneficial as they often help to shed light on an issue and trouble spots that may not be identified with other forms of pre-testing (Weinreich, 2011).

Eliciting input from gatekeepers may also prove advantageous (Atkin & Freimuth, 2012; Weinreich, 2011). Gatekeepers are those actors that can act as intermediaries towards intended audiences. For instance, because they control distribution channels that are frequently visited by the target audience. As such, their approval or disapproval of messages and materials could potentially impact upon program success. Besides providing further insights, soliciting their involvement in pre-testing may contribute to enhanced cooperation, and hence, may benefit the execution of a program (Atkin & Freimuth, 2012).

5.4.8.2. ANECDOTES AND TESTIMONIALS

The stories and testimonials of individual people can provide insight in effects of a campaign at a *personal level*. For instance, these might come from members of the target audience that directly reach out in response to a campaign. Interviews with ‘formers’ – especially those radicalized in the past few years – could potentially also shed light on the role of counter-narratives narratives in their process of disengagement or de-radicalization (RAN, 2015). In and by themselves, such anecdotal evidence doesn’t proof anything about the program. However, they can be used to give meaning or color to quantitative measures and bring to life the people behind the statistics (Weinreich, 2011; Paul et al. 2015).

5.5. SUMMARY

In this chapter, we have addressed the question how to measure the effectiveness of online counter-narrative programs (research question 5). While in general it is very difficult to establish an effect of a policy or measure in CVE (e.g., Noordegraaf, Douglas, Bos & Klem, 2016), we have specified various methods that can be employed to at least gain some insight in the potential effectiveness of a counter-narrative campaign. In doing so, it is important to distinguish between the formative evaluation (is the program well-designed, opening the black box: how is the program supposed to work and is it theoretically driven?), the process evaluation (has the program been skillfully executed?) and the summative evaluation (does it result in the desired effect?). We also outlined the option to consider a more modest approach, the realistic evaluation model (Pawson & Tilley, 2004), according to which it is possible test specific elements of a campaign, in a specific context, and for specific people.

In terms of online monitoring, a combination of awareness (e.g., reach and views) and engagement metrics (e.g., clicks, likes, shares, comments and emoji responses) can provide some insight into the extent to which a counter-narrative campaign achieved its desired *online* effect as defined in the goals and objectives in terms of reaching and engaging an audience (Reynolds & Tuck, 2016), which can or cannot be related to *off line* behavior.

Several other, more traditional, research techniques can give further insight in the potential effectiveness of a counter-narrative campaign. First, interviews and focus groups can provide important (often qualitative) information about the reception of online materials (thus useful in the formative phase), but can also provide input in the interpretation of quantitative data in the summative evaluation phase. Second, while online surveys may offer quick results, it may not be always possible to reach the intended audience, partly due to the self-selection bias (i.e., only those with a vague interest will participate in a study). Third, experiments (either off line or on line) may provide good insight in the cause-and-effect-relationship (e.g., do people respond differently to message A and message B?). Fourth, “netnography” is a form of ethnography adapted for the analysis of social media and online communities: it makes use of data usually obtained from observations of people in their regular social media environments. This non-obtrusive manner of collecting data can be useful for sensitive issues, such as counter-narratives. Fifth, sentiment analysis is a method to distract the overall evaluation or sentiment towards a counter-narrative campaign (via Natural Language Processing – NLP). A more elaborated content analysis also picks up the specific frames or ideas that were put forward in a particular online setting or community. Finally, social network analyses may give insight into the structure of a group (i.e., who plays a central role?) and potentially how this might change over time, possibly due to a counter-narrative campaign, although it is difficult to reliably relate an outcome to the counter-narrative campaign, as it is always possible that another factor has caused the change in the structure of the social network.

Thus, while at the moment, it is still very difficult to establish with great certainty an effect of an online counter-narrative campaign, we do believe that some of the research methods that we have outlined may contribute to at least a better understanding of how (some elements of) a campaign may be perceived and responded to by (individuals in) an audience.

6. THE ROLE OF THE GOVERNMENT

6.1. INTRODUCTION

In this chapter, we examine the following research question: What can be the role of the government in a counter-narrative program? We address this question by examining the potential role of the government as an organization that produces counter-narrative campaigns (Section 6.2), that streamlines its own strategic communications (Section 6.3), that can consider fostering public, private and 3d-sector partnerships (Section 6.4), that can build capacity (Section 6.5), that stimulates evaluation and research (Section 6.6) and that strengthens digital literacy and consumption skills (Section 6.7). We summarize the main points of this chapter in Section 6.8.

6.2. THE GOVERNMENT AND COUNTER-NARRATIVE CAMPAIGNS

In general, governments should carefully consider their role in campaign efforts. People are often not very receptive to government communications (Renes et al., 2011). It is suggested that this may hold especially for counter-narrative messages. While various governments have engaged in counter-messaging, it is increasingly acknowledged that government actors are not well positioned to play a role as a producer and messenger of counter-narratives, as they are likely to lack credibility and be perceived negatively by relevant audiences (e.g., Aistrope, 2016; Briggs & Feve, 2013; van Ginkel, 2015; RAN, 2015, 2015; Waldman & Verga, 2016). In this regard, there was also general consensus among the experts interviewed for this study. For instance, Kruglanski⁵⁸ puts forward that *“The government has very little credibility. So, you know, to the extent that these videos are presented as coming from the government, this undermines their credibility”*. Based on a similar notion, Aly⁵⁹ asserts, *“It can’t be done by government. The minute government wants to do it, everyone will run away”*. Thus, governments should proceed with caution if they want to play a direct role in the production and dissemination of counter-narratives (Briggs & Feve, 2013).

One of the problems that plagues government’ counter-narratives efforts is that they tend to suffer from a so-called ‘say-do gap’, which is the difference between what governments say and what they do. That is, in practice the authenticity and credibility of government counter-narrative communications are often times undercut by a gap between policies and the every-day realities and perceptions of those they are targeting (Aistrope, 2016; ISD, 2014). Something which may not only render them largely ineffective, but could also lead them to backfire in the sense that they may exacerbate perceptions or grievances Salafi-Jihadi strategic narratives can exploit. This say-do gap is often leveraged in extremist propaganda (Ingram & Reed, 2016).

It should also be noted that some experts have argued that past government attempts to challenge Salafi-Jihadi extremist narratives head on have actually served to embolden and dignify them. In fact, it is even posited that Salafi-Jihadi have been able to leverage these efforts to reach

⁵⁸ Arie Kruglanski, interview, 8 March, 2016

⁵⁹ Anne Aly, interview, 25 February, 2016.

a broader audience. In this regard, the ‘Think Again, Turn Away’ campaign of the US state department has been mentioned as a primary example (See Box 12; Edelman, 2014; Katz, 2014).

Besides these issues, some authors have also argued that, on a more practical level, bureaucratic and operational constraints as well as low risk tolerance also make it difficult for government actors to communicate timely, responsively and assertively, and hence, effectively (Bole & Kallmyer, 2016 ; Briggs & Feve, 2013).

In sum, the ability of governments to effectively directly engage in counter-narrative work is described as limited at best and counterproductive at worst, although, as argued before, governments can disseminate different messages and do not always speak the same language. But generally speaking, governments may not be in the ideal position to deliver counter-narratives due to source credibility concerns. That is, however, not to say that governments cannot and should not engage in communication efforts on this terrain. What are the other options for governments to pursue in this field?

6.3. STREAMLINING GOVERNMENT STRATEGIC COMMUNICATIONS

While it is increasingly acknowledged that government actors may do best refrain from counter-narrative messaging, some authors suggest that there is a role for the government when it comes to streamlining their own online strategic communications (e.g., Briggs & Feve, 2013; van Ginkel, 2015; Waldman & Verga, 2016). Of course, in complex governmental systems, different branches do not always speak with the same voice. Briggs & Feve (ibid.) of the UK’s Institute for Strategic Dialogue argue that governments should strive to develop and implement a coherent and consistent strategic communication policy that addresses the challenge of extremist messaging on social media and the broader internet. In this regard, they encourage governments to consider establishing a specialized cross-departmental unit to oversee and coordinate this work. Such a unit could also provide tools and services for other government actors, as well as external partners. The strategy itself should address key issues of relevance to countering violent extremism. What kind of practical suggestions have been put forward?

Several reports recommend that governments clearly and proactively communicate and explain their own foreign choices in relation to issue-relevant topics such as their involvement in relevant conflict zones (e.g., Briggs & Feve, 2013; Briggs & Silverman, 2014; van Ginkel, 2015; Hussain & Saltman, 2014 ; ISD, 2014; Russel & Rafiq, 2016). Some authors indicate that anger about what is (alleged to be) happening in these conflict zones and empathy with the people being affected can play a role in radicalization processes and be a driver for foreign fighter travel (Frenett & Silverman, 2016; see also Bakker & Grol, 2015; Briggs & Silverman, 2014; Neuman, 2015). As such, it is posited that government’s effective or ineffective response may impact on these phenomena. In this regard, it is suggested that it may be helpful that governments make more visible what they are doing in terms of humanitarian aid and assistance to populations in Syria, Iraq and the likes. Additionally, it is advised that governments should identify and publish practical and legal alternatives for those who are moved to assist (Briggs & Silverman, 2014; Bakker & Grol, 2015; ISD, 2014). In both cases, it has been put forward that it is crucial that information is accessible to those communities that are mostly ‘moved’ by such crises (ISD, 2014). While we believe such efforts are generally valuable and relevant, it should be noted that since the so-called

Caliphate has been declared, 'humanitarian motives' for wanting travel to Syria and Iraq have declined (Schmid & Tinnes, 2015).

Policies and measures adopted to address national security risks should also be brought under attention (van Ginkel, 2015; Russel & Rafiq, 2016). In relation to the foreign fighter phenomenon, there are some studies that indicate that some youngsters expect to find thrill and adventure, but underplay the risks involved with travelling as well as the grim reality of the terrorist lifestyle. Some authors suggests that (national and local) governments should therefore (continue to) communicate and specify what the legal consequences are for travelling, in addition to pointing out the risks involved (Briggs & Silverman, 2014). Another recommendation that can be made, is that authorities communicate that they are committed to tackling extremist violence and terrorism generally, and minimize the perception that counter violent extremism (CVE) policies are exclusively attuned (a cross-section of) Muslim communities. As we have suggested in Chapter 4, such perceptions run the risk of feeding into the narrative that Muslims are a threat, and may contribute to feelings of alienation among the Muslim populations who feel unrightfully singled out on the basis of their background (Berger, 2016).

In the case of targeted community, specific CVE-policies and programs are deemed necessary. Waldman and Verga (2016) suggest targeting strategies should be risk and evidence-based, rather than rely on generalized assumptions or (geo)demographic variables. Additionally, they argue that contextualized placed-based assessments could potentially help community specific bases for resilience, which could be potentially be leveraged in CVE efforts. Related to this, they suggest governments should clearly distinguish programs from each other in their communications. They put forward that engagement and relationship-building efforts at the level of the community may prove valuable as they could enable more tailored and effective CVE programming towards those deemed 'at-risk'. However, it may prove useful to introduce these programs separately from CVE efforts, so as to ensure that they are not viewed as solely aimed at CVE, which could contribute to distrust in communities and add to the perception that communities are 'singled out'.

Additionally, under certain circumstances, governments are advised to undertake efforts to separate facts from fiction, address misrepresentations of government actions and clarify the intentions behind what the government is doing. However, as we have suggested in Chapters 2 and 4, it is of importance that government actors carefully consider the performative effect of their 'actions' in relation to their 'words'. Otherwise they may well play into the hands of those with a violent extremist agenda (Hedayah & ICCT, 2014; van Ginkel, 2015; Waldman & Verga, 2016). Projecting a message is not enough when it is not backed up by deeds, or worse, if deeds actually contradict what is being proclaimed (van Ginkel, 2015).

Thus, while the government might not be suited as a messenger for counter-narratives, it does play an important role in streamlining its own strategic communication. In particular, governments are well-advised to clearly explain its domestic and foreign policies.

6.4. FOSTERING PUBLIC-PRIVATE PARTNERSHIPS

As argued in Chapter 4, civic and grassroots actors and organizations at the community-level are commonly identified as well-positioned to act as agents for messaging activities that can

undermine extremists' agendas and may best be able to inspire and mobilize others to become active in challenging extremist narratives. This is because they are more likely to know communicative and organizational nuances of their audiences. Furthermore, they are commonly well aware of what goes on with (vulnerable) members in a given community and are often better-placed to identify local issues that may potentially contribute to violent-extremism (Waldman & Verga, 2016; See also RAN, 2015). As these actors are, generally speaking, more likely to be perceived as credible and to have authority, Waldman and Verga (2016) also assume this would make them better-equipped to take conversations offline with those individuals identified to be susceptible to violent extremism. In correspondence with our argument in Chapter 4, they also argue that they are more likely to have an impact here.

Unfortunately, these actors often lack resources (e.g., funding and capacity), expertise and competencies to carry out counter-narrative work effectively (Briggs & Feve, 2013, RAN, 2015). Moreover, it has been noted that their counter-narrative efforts are frequently based on noble intentions, but not on solid and rigorous research (Russel & Rafiq, 2016).

While governments may be ill-positioned to engage in alternative and counter-narrative work themselves because of the credibility gap, they do have necessary resources and may be motivated to support the work of such actors (RAN, 2015). Also, there are many academics that possess valuable expertise that could be leveraged for counter-narrative work. These could potentially also give a helping hand in mediating between governments and community actors and organizations. Yet, it has been opined that only few are involved in 'imaginative counter-radicalization work' (Holtmann, 2013). Furthermore, private sector parties, such as technology and social media firms and advertisement and marketing specialists, could lend valuable skills, expertise, funds, services and access to tools that could enable civil society actors to be more effective in their work. However, these parties often lack knowledge of violent extremism and may have (reputational) concerns about working in this particular area. Moreover, their primary focus will be on their core business, which could be reflected in a lack of time or investments in this domain (RAN, 2015).

Given the complementary roles these different actors could potentially play, it is frequently argued that governments can fulfill a particularly valuable function by encouraging and brokering partnerships between civil society and the private sector. Governments may establish an infrastructure to streamline and sustain such partnerships, and perhaps providing funding or other in-kind support to partnership initiatives (Briggs & Feve, 2013; Briggs & Frennet, 2014; Neumann, 2011; RAN, 2015; Helmus, York & Chalk, 2013; Russel & Rafiq, 2016; Waldman & Verga, 2016). Furthermore, governments are important for giving necessary strategic direction to such efforts as well as guidance to ensure effect and efficiency (RAN, 2015; Russel & Rafiq, 2016).

Importantly, though, it is suggested that governments should remain at arms-length, and should try to repress any urge to front such initiatives or influence the content that is developed and relayed through such efforts. Moreover, it is vital that developed funding models and infrastructures do not 'contaminate' any efforts undertaken, as endorsement of (or association with) the government may well serve as a 'kiss of death', in the sense that it may serve to discredit such initiatives (RAN, 2015).

This is a position echoed by various experts that were interviewed for this study. For instance, Archetti puts forward that: *“anything that becomes associated to the government, for example this network of credible voices, then it starts having a negative—starts being seen negatively”*⁶⁰. Kruglanski⁶¹ also argues that the extent to which campaign content is associated with the government will undermine its credibility. While he acknowledges that the government can potentially facilitate with providing an infrastructure as well as support, training, and funding, “the government has to remain in the shadow [...] the sponsorship of the government is not an advantage, it is a disadvantage”. In line with this, the RAN issue report concludes that governments essentially “need to help to create an infrastructure that is community-owned and operated, but that is non-governmental in how it operates” (RAN, 2015). Conversely, Aly⁶² argues: “the best thing governments can do is, to help build the civil society structures... and infrastructure to deal with this, but they have to be prepared to walk away and do it at arm’s length”.

6.5. EMPOWER EFFECTIVE MESSENGERS

Counter-narratives are assumed to be most effective when they stem from credible messengers. In Chapter 4, we have pointed out that some community-level messengers are commonly assumed to be effective. These include former violent extremists, survivors, peers and family members, religious scholars. In line with this, it may be worthwhile to encourage efforts that solicit involvement of these individuals in communication strategies.

Young people in particular are mentioned by the literature and some of our interviewees as actors that may be able to create content that resonates with peers^{63 64}. Hence, facilitating initiatives for youth engagement may be beneficial here. However, as suggested grassroots actors - youth in particular - may feel reluctant to work on campaigns or topics when they are framed in terms such as CVE as well as to associate themselves with the government (RAN, 2015). Hence, it is argued that it would be more suitable to solicit their participation in alternative narrative campaigns that are CVE relevant rather than specific, such as those related to issues like social cohesion or non-violent political activism (which may in fact be government critical).

Furthermore, as argued earlier, it is vital that governments limit their involvement and allow for operational independence. In addition, government’s community and/or youth-generated online initiatives that are already out there and may be popular among relevant target audiences (e.g., Richardson, 2013; Briggs & Feve, 2013; Helmus, York & Chalk, 2014). As we have suggested, one is more likely to reach audiences by disseminating content in existing social networks, than by establishing new online locations. People tend to be loyal to the online spaces they are already visiting and appeal to their interest (Richardson, 2013). Hence, rather than reinventing the wheel and promoting the establishment of new online channels which may have

⁶⁰ Arie Kruglanski, interview, 8 March, 2016

⁶¹ Arie Kruglanski, interview, 8 March, 2016

⁶² Anne Aly, interview, 25 February, 2016.

⁶³ Anne Aly, interview, 25 February, 2016.

⁶⁴ Diana Rieger, interview, 19 May, 2016

a limited (and self-selected) readership, governments do well to engage with and leverage the power of established and up-and-coming social media influencers – who may well be critical of the government - that have initiated relevant online initiatives (Briggs & Feve, 2013; Richardson, 2013). Governments could investigate how they could support such initiatives, without contaminating them.

Also, anecdotal evidence suggests that former violent Salafi-Jihadi extremists, former foreign fighters as well as survivors of extremist violence may be potent messengers delivering counter-narratives. However, as put forward in Section 4.4.2, there are numerous practical and (and in the case of formers, legal) challenges that stand in the way of soliciting their participation in counter-narrative work. For instance, formers and survivors may not want to have public profiles or may want to leave that part of their life behind them (Hedayah & ICCT, 2014). Those that are in fact willing to engage in this work, often do not have the time, competencies or resources to effectively do so. Hence, it has been suggested that governments may want to consider ways in which they can incentivize and facilitate formers and survivors to engage counter-narrative work to a greater extent (RAN, 2015; Briggs & Feve, 2013). For instance, the ISD suggests assisting formers on an ‘ad hoc basis’. For instance, *“through the provision of contacts to expand their reach, help in unlocking independent funds for specific project-based activities, or brokering offers of in-kind support”* (Briggs & Feve, 2013, p. 19). Of course, governments should have necessary checks and balances and accountability mechanisms in place, before deciding upon supporting or funding anyone, especially formers (ibid.). In this regard, it is argued to be of importance that formers are carefully vetted and selected, before they are supported in engaging in any counter-narrative work (Zeiger, 2016). Also it is important these actors themselves are made aware of potential personal risks involved (Briggs & Feve, 2013; Waldman & Verga, 2016). Lastly, governments should take notice that supporting formers may give rise to contention among parts of the general population (QIASS, 2013).

6.6. CAPACITY BUILDING

A barrier that may stand in the way of community-level actors that want to get involved in the production and dissemination of alternative and counter-narrative is a gap in skills and expertise. As such, another frequently mentioned area in which the government could lend support is capacity building, for instance, by establishing or seeding training programs (e.g., Briggs & Feve, 2013; Briggs & Frennet, 2014; Hedayah & ICCT, 2014; Helmus, York & Chalk, 2013; RAN, 2016; van Ginkel, 2015; Waldman & Verga, 2016).

In developing training programs, governments are advised to involve private sector parties such technology and social media firms, marketing and advertising companies, and other parties with experience in relevant domains which could contribute to their value and effectiveness (Briggs & Feve, 2013; Briggs & Frennet, 2014; Hedayah & ICCT, 2014; Helmus, York & Chalk, 2013; Waldman & Verga, 2016). Among other things, it is suggested that private sector organizations could lend help with providing (technical) guidance on such things as search-engine optimization techniques, marketing and crowdsourcing strategies, and digital media-making. These training programs are ideally delivered by community-level organizations in order to ensure that these are optimally attuned to the needs of key audiences, and set up as seeding programs

(Briggs & Feve, 2013). Also, once such training programs have been established, it would be worthwhile that efforts focus on training the trainers (Hedayah and ICCT, 2014).

In addition, we would argue that the government could act as a turntable in collecting and disseminating information and best practices regarding existing programs and campaigns to interested citizens and groups. In this regard, they can point to various interesting online ‘resource hubs’ that house valuable documents and materials that are already out there (see Box 9 for some examples). Governments should be cautious in creating yet more web-platforms that act as such, and are perhaps better advised to translate key resources and get these to relevant audiences (Briggs & Feve, 2013).

Box 12: Online resources

The Counter-Narrative Toolkit

The ISD’s counter-narrative⁶⁵ toolkit, offers a free online step-by-step guide for any individual or organization looking to engage in counter-narrative campaigning. According to the website it is not intended to be comprehensive, yet provides a very useful and accessible beginners guide for those with little or no experience in campaigning. Besides offering general insights in some ‘best practices’, it provides video-tutorials, and a more detailed downloadable ‘counter-narrative handbook’ (Tuck & Silverman, 2016), as well as several case studies to serve as an inspiration. The toolkit was funded by Facebook. It was inspired by an earlier pilot project with Jigsaw (formerly Google Ideas), who also funded some of its video content(Counter-Narrative Toolkit, n.d.).

VOX-pol network of excellence

VOX-Pol Network of Excellence (NoE)⁶⁶. is a European Union Framework Program which main purpose is the “*establishment of a robust partnering, research, training, and dissemination network that has as its core function comprehensive research, analysis, debate, and critique of topics in and issues surrounding Violent Online Political Extremism*” (VOX-pol, 2016). Their dedicated website provides a comprehensive and regularly updated collection of publications related to various aspects of online aspects of violent online political extremism which can be used as an educational and research resource.

Prevent violent extremism research portal

Prevent Violent Extremism & Social Media portal is joint initiative between the SecDev Foundation and SecDev, with support from the Government of Canada’s Kanishka Project⁶⁷. The portal compiles research on violent extremism and social media. Thereby it aims to enable CVE researchers to quickly find current research on threats and potential of social media in the fight against violent extremism.

⁶⁵ Counter-narrative Toolkit. (n.d.) <http://www.counternarratives.org>. Retrieved on 15 September 2017

⁶⁶ Vox-Pol. (c2017). <http://www.voxpol.eu>. Retrieved on 15 September 2017

⁶⁷ Prevent Violent Extremism. (c2017). <https://preventviolentextremism.info>. Retrieved on 15 September 2017

Against Violent Extremism (AVE) network

Earlier mentioned in Chapter 4, Against Violent Extremism (AVE)⁶⁸ is a private sector partnership between the Institute for ISD, Jigsaw (formerly Google Ideas), the Gen Next Foundation and Rehabstudio and is managed by ISD in London. Its official aim is to “prevent the recruitment of ‘at risk’ youths and encourage the disengagement of those already involved”, and is dedicated to countering all forms of violent extremism (e.g., far-right, far-left, Al-Qaeda-linked or inspired, and gangs). It intends to leverage lessons, experiences and networks of individuals who have dealt first-hand with extremism such as formers and victims. Through a dedicated website and YouTube channel Members can connect, exchange information, ideas and perspectives, find potential partners and investors, and project their messages to a wider audience (Against Violent Extremism, n.d.-b).

6.7. EVALUATION AND RESEARCH

While these are all reasonable recommendations, however, they need to be based on better evidence than that is presently available. A wide range of authors have noted that there is currently a near total lack of published evaluations of efforts undertaken, and that present initiatives appear to frequently rely on untested assumptions about what works (e.g., Beutel et al., 2016; Ferguson, 2016, Waldman & Verga, 2016). Thus, while “[t]here is no shortage of literature discussing the need for metrics to evaluate CVE success, but that has not translated into the implementation of actual evaluation on any significant scale” (Berger, 2016, p. 8). In this regard, Von Behr et al. (2013), pointed out that “the absence of a robust, comprehensive evaluation of counter-narrative work online is a concern – not least because it is not clear whether the work is well targeted or effective in changing the attitudes or behaviours of those vulnerable individuals engaging with radicalizing material online” (p. 35). We believe that there is truth in this assertion. Whilst it may be difficult to measure ‘effect’ (see chapter 5), if any new activities in this realm are to be supported it should be encouraged that they are consistently monitored and evaluated. Stimulating those undertaking an effort to conceptualize a ‘theory of change’ (see chapter 4), can contribute to them doing so meaningfully. As has been noted elsewhere, whenever possible, evaluations should be made publicly available in order to facilitate comparison and analysis so they can inform future initiatives (Romaniuk, 2015). Ideally, outcomes are shared in a way that is operationally useful to citizens and groups interested in engaging in such efforts (Waldman & Verga, 2016).

At the same time, it may be worthwhile for governments to support further research and (online) experimentation to identify meaningful measures or proxies of success in changing attitudes and/or behaviors of specific target audiences (Waldman & Verga, 2016), or more generally, meaningful approaches to come to these measures. Also, we believe that Aly (2017) raises a fair point, when she suggests that “[gaining a deeper understanding in] [t]he process by which terrorists’ [or extremist’] audiences make meaning from the violent extremist narrative and the variables – demographic, contextual, individual and group related – that effect these meanings could help ensure that counter-narratives are equally as effective as the violent extremist narrative appears

⁶⁸ Against Violent Extremism. (n.d.). <http://www.againstviolentextremism.org>. Retrieved on 15 September 2017

to be” (p. 26). As such, we would argue that more contextualized research on how audiences engage with, interpret and reproduce extremist messaging online and how this relates to violent radicalization could inform future CVE strategic communication efforts.

6.8. STRENGTHENING DIGITAL LITERACY AND CONSUMPTION SKILLS

While not directly related to online counter-narratives efforts, there is another area in which the government could play an important role. Bartlett and Miller (2010) of the British think-tank Demos argue that: “[w]hile government cannot tell people what to think, they can help teach people how to think” (p. 38). In this regard, several studies and reports point to the importance of strengthening people’s ability to critically appraise the online (sources of) information they encounter (e.g., Briggs & Feve, 2013; Hussain & Saltman, 2014; Neumann, 2013; Miller & Bartlett, 2012; Reynolds & Scott, 2016). It is often times difficult to distinguish between legitimate and false information in the digital world. While young people may be adept in using the internet and social media, they frequently lack the appropriate skills to thoroughly evaluate and judge the content they consume and share in these spaces. This could potentially make them vulnerable to negative online influences (e.g., conspiracy theories, misinformation and scams), including to radicalizing influences (Miller & Bartlett, 2012). In enhancing children’s and young people’s digital literacy and critical consumption skills, schools, youth and community organizations can play an important role. Government efforts could focus on promoting and supporting the development, implementation and evaluation programs. Also, in this case, the government could act as a hub or turntable in collecting and disseminating information and best practices, as well as support academic research in this area.

Box 13: ‘Digital Citizens: Countering Extremism Online’

A report from the United Kingdom think-tank Demos (Reynolds & Scott, 2016), which sets out the findings of a pilot project involving a digital citizenship intervention developed in cooperation with Bold Creative provides some indication that developing young people’s critical thinking and digital literacy skills can prove effective in improving their resilience to online extremist influences. The pilot, which was conducted under the UK’s Home Office’s Prevent Innovation Fund, involved developing, testing, and evaluating new resources to help schools tackle online radicalization.

The design of the intervention was informed by interviews with 11 key stakeholders and a review of nine higher quality evaluations and meta-evaluations of similar CVE interventions from the UK, US and Australia (Reynolds & Scott, 2016). Rather than focusing on countering the ideologies underlie extremism, the intervention sought to increase resilience to online radicalization through a skills-based teaching approach (Demos, 2016). More specifically, it intended to “teach young people how to recognise online propaganda and manipulation, understand how social media change how we communicate, and develop a sense of responsibility over their online social network” (Reynolds & Scott, 2016, p. 12).

Four schools took part in a pilot of the intervention. The post-pilot evaluation revealed it had produced statistically significant improvements on all three focus areas mentioned above. Moreover, the intervention was conceived of favorably by both teachers and participants.

Although the results are promising, the report also points out that participants and comparison group members were not selected randomly and that the sample size was small. As such, they argue that it is difficult to draw any general conclusions about the effectiveness of this pilot project (Reynolds & Scott, 2016).

6.9. SUMMARY

To summarize, while the government actors may not be well-positioned to act as a counter-narrative producer or messenger, they do have an important role to play. First, governments can fruitfully engage in streamlining their own strategic communications. Among other things, it is important that they explain their own actions locally and in an international context. In addition, the government can play a valuable role by facilitating and supporting grassroots and civil society actors best placed to act as alternative and counter-narrative messengers. They can do so by brokering public, private, 3d-sector partnerships, and establishing an infrastructure to streamline and sustain them. Also, they can stimulate grassroots efforts with funding (albeit the government should remain in the shadow), training and expertise. Importantly though, there is only limited evidence regarding the usefulness of online counter-narrative efforts. Hence, if any new initiatives are to be supported, the government should encourage monitoring and evaluation. At the same time, further research to identify meaningful measures or proxies of success in changing attitudes and/or behaviors of specific target audiences is needed, and governments could play a role in supporting those efforts. Likewise, more exhaustive research is necessary on the way audiences engage, interpret and reproduce extremist messaging online and how it relates to violent radicalization, in order to inform meaningful counter-strategies. Finally, government efforts could focus on promoting and supporting the development and implementation of programs in the area of strengthening digital literacy and critical consumption skills. This can be achieved by encouraging efforts and evaluation as well as stimulating knowledge sharing and development in this area.

7. SUMMARY, CONCLUSIONS, LIMITATIONS AND FUTURE DIRECTIONS

7.1. INTRODUCTION

In this report, we have examined the extent to which counter-narrative initiatives via social media can be effective in preventing people from radicalization or can de-radicalize people. We have done so by studying the literature in combination with conducting interviews and focus groups with academics, field workers, young social media experts and some former radical people. In Section 7.2, we present a summary of our answers to the main research questions posed in Chapter 1, the Introduction. In Section 7.3, we discuss more generally the findings from our analysis. In Section 7.4, we describe the most important limitations of the current study. Finally, in Section 7.5, we sketch some potential future directions.

7.2. SUMMARY OF ANSWERS TO RESEARCH QUESTIONS

In the Introduction (Chapter 1), we formulated the following research questions and we now present the following answers to these questions:

1. How can we conceptualize narratives and counter-narratives?

We conceptualize narratives as strategically constructed storylines that are projected and nurtured through (online) strategic communication activities by state and non-state actors in attempts to shape how target audiences feel about or understand events or issues, and ultimately, guide their behavior in a manner that is conducive to their aims and goals. We conceptualize counter-narratives as strategically constructed storylines that are projected and nurtured through strategic communication (or messaging) activities with the intention to undermine the appeal of extremist narratives of violent extremist groups.

2. How are narratives and counter-narratives used via social media?

Narratives have been projected using (amongst other channels) mainly YouTube, Twitter and Facebook. For counter-narratives on social media, we identify 3 domains: (1) counter-messaging (e.g., activities that challenge extremist narratives head on); (2) alternative messaging (e.g., activities that aim provide a positive alternative to extremist narratives); and (3) strategic communication by the government (e.g., activities that provide insight in what the government is doing).

3. To what extent is it possible to use counter-narrative programs via social media to de-radicalize individuals or prevent violent extremism?

We argue that to the extent that people have become more radical, they are less likely to be persuaded by a counter-narrative campaign. They may not pay attention to these messages or even may adopt a stronger attitude in the other direction than intended as a reaction to this persuasive attempt. In particular, when the aim is to directly confront. Thus, we argue that a counter-narrative campaign might best be used for prevention purposes for a well-defined target group. Perhaps for those individuals who can be identified to show some curiosity or sympathy for extremist groups, but are not yet active supporters.

In addition, we have raised the question whether or not such counter narrative efforts may have some effect for the individuals from radical groups who show a glimpse of doubt about their group. For such individuals, a counter-narrative message may fall on fertile grounds and a seed may be planted, although this is highly speculative at the moment and more research is needed to support this notion.

4. What are the pre-requisites for a counter-narrative program for it to be effective?

In developing a campaign and determining the communication strategy, it is important to give consideration to theory, the audience, extremist narratives, goals and objectives, available recourses, the message, the message source, channels, methods of dissemination, and potential risks and limitations of a strategy. First, the application of relevant *theory* is an important strategy to increase the effectiveness of campaigns. Beyond borrowing from existing theory, program planners do well to develop a sound theory of change that explicates how campaign efforts are going to lead to the desired result. With respect to the *audience*, this should be narrowly defined and segmented on the bases of meaningful variables. This can only be done effectively when the target audience is thoroughly researched and understood, hence rigorous audience analysis is necessary.

In addition to studying the audience, it is important to develop an in-depth understanding of the *extremist narratives* one aims to counter. This holds in particular in relation to counter-narratives. When deciding upon which elements of the Salafi-Jihadi narrative one aims to counter, we argue it is also important to consider why members of the target audience in question may be drawn to these narratives. For example, different motives of radical people may request different content of counter-narratives. This is related to the point that, at the outset of a campaign, *goals and objectives* should be delineated that clearly specify which change one aims to achieve in terms of audience knowledge, attitudes or behaviors. Well-defined objectives are specific, measurable, achievable, relevant, and time-bound. Running an effective social media effort also requires adequate *resources* (time, finances, human capital) which should be assessed at the outset of a program, taking into account all relevant stages of a campaign.

Furthermore, *messages* that use the narratives have the advantage of using subtle ways to influence people. Moreover, they may overcome various forms of resistance. In addition, double-sided messages (in which one side is being discredited), coupled with a strong emotional appeal may prove to be persuasive. In terms of *sources*, it is clear that they need to be perceived as credible to function as trustworthy messengers. Most likely candidates are former extremists, victims of violence, peers and family, as well as key members of communities and civil society actors. The potential of the government to serve as a credible source might be limited.

In terms of selected *channels*, obviously they should be frequently utilized by the audience one intends to reach. Using multiple channels is argued to be beneficial. Importantly, social media may not be appropriate in all circumstances, and should be considered in combination with offline activities. In terms of *message dissemination*, we argue to join already on-going 'conversations'. Also, it may be beneficial to partner up with people and organizations that have already established a strong social media presence and are popular among the target audience. Conversely, it may also be useful to ask established online (news) outlets to share the message. Furthermore,

it may be important to create opportunities for audience engagement. Digital advertisement could also help to get the message out, but also has important limitations.

Finally, we argue it is important to assess the potential risks, challenges and limitations when developing a communication strategy. These might include losing control of the message, counter-campaigns, threats to safety and well-being of messengers, and multi-tier approval processes that hamper campaign efforts, as well as the notoriously difficulty of demonstrating the effect of social media efforts.

(5) How can the potential effectiveness of such a counter-narrative program be determined?

Generally speaking, it is not easy to establish an effect of a policy or measure in CVE (e.g., Noordegraaf et al., 2016). For the potential effectiveness of a counter-narrative campaign, it is important to distinguish between a formative evaluation, a process evaluation and a summative evaluation. The realistic evaluation method states it is better to establish specific hypotheses about what element works for whom in which context and why.

Specifically tailored at online interventions, we argue that a combination of awareness (e.g., reach and views) and engagement metrics (e.g., clicks, likes, shares, comments and emoji responses) can provide some insight into the extent to which a counter-narrative campaign achieved its desired effect in terms of reaching and engaging an audience. More traditional research techniques such as interviews and focus groups can give further insight in the reception of online materials (thus useful in the formative phase), but can also provide input in the interpretation of quantitative data in the summative evaluation phase. Experiments (either off line or on line) offer insight in the cause-and-effect-relationship, while “netnography” makes use of data usually obtained from observations of people in their regular social media environments. Sentiment and content analyses are methods to distract the overall evaluation towards a counter-narrative campaign as well as the specific narratives that were put forward in a particular online community. Finally, social network analyses can give insight into the structure of a network, the spread of messages within communities, as well as potential changes in the network.

(6) What can be the role of the government in such a counter-narrative program?

As argued before, government actors are not well-positioned to act as a counter-narrative producer or messenger, as they lack the credibility. However, they do have an important role to play. First, governments can fruitfully engage in streamlining their own strategic communications in terms of explaining their own actions locally and in an international context. Second, the government can play a valuable role by facilitating grassroots and civil society actors best placed to act as counter-narrative messengers. They can do so by sponsoring such efforts (providing help, expertise or financial support) as well as by establishing an infrastructure to support these initiatives. Thirdly, they could stimulate thorough monitoring and evaluation, as there is only limited evidence for the effectiveness of counter-narrative efforts. Finally, government efforts could focus on supporting the development of programs in the area of strengthening digital literacy and critical consumption skills.

7.3. DISCUSSION

Taking into account the insights derived from our literature study, in combination with our interviews and focus groups, we can draw several conclusions. First, we argue that, in general, the development and implementation of online counter-narratives are still in its infancy phases. This seriously undermines the potential to draw any firm conclusions from our analysis at this moment. As such, the arguments presented here need to be treated with caution and deserve further examination at a later date when there is more empirical research available to base our arguments on.

Secondly, we argue that the domain of counter-narratives is severely hampered by a lack of strong empirical research and thorough evaluations. This is perhaps partly due to the difficulties of doing research with radical people, because they are not easy to recruit as participants. In general, it is also extremely difficult to measure changes in attitudes and beliefs (which may take place over a long term), let alone measuring prevention where the outcome is a non-event. More specifically related to online efforts, the anonymity that the internet provides often makes it impossible to establish precisely who engage with a campaign on social media. This leads to the conclusion that it is virtually impossible to establish a direct connection between counter-narrative campaigns and the prevention of extremism.

Thirdly, given the fact that people increasingly make use of social media, it makes it all the more likely that people may be influenced by them, also in terms of radicalization. As such, this makes a crucially important to have a thorough understanding of the techniques and strategies employed by radical groups in how they are capable of reaching out to their audience. However, there are limitations in the use of counter-narratives to influence people. In particular, to the extent that radical groups use ideological narratives about perceived grievances to stir up feelings of resentment, that are rooted in deep experiences (e.g., in terms of everyday experiences of discrimination of one's group), counter-narratives may fail, as they do not directly impact on these actual everyday experiences.

Fourthly, our analysis shows that presenting online counter-narratives may be most fruitful to use in a prevention context, in which people may still be willing to process the message. Alternatively, in terms of de-radicalization, it might be possible to explore the options to expose individuals who have demonstrated some doubts about their group with specific counter-narrative messages. However, ideally, this is not done in an online context, but an offline context, in order to directly monitor the reactions of the individuals.

7.4. LIMITATIONS OF THE CURRENT STUDY

We have tried to answer the research questions using a literature study in combination with interviews and focus groups with experts in terms of radicalization and social media: academics, field workers, social media students and some former radicals. While this has given us a solid and broad scope of the domain, this study (like any study) has its limitations.

First, because the domain of narratives is covered in various disciplines, such as amongst others media studies, history and psychology, we have drawn from a large body of studies and have tried to present an integration of the ideas from these fields. However, we acknowledge we might not have covered all there is to be found in the scientific literature terms of narratives and counter-narratives.

In addition, another limitation from the current study is that it has been done while the field is in full motion. This means that the current state of affairs, as presented in this report, most likely will be outdated in a near future. As such, the analysis presented here has a “limited warranty” in terms of the time that it is relevant and valid.

Another limitation concerns the fact that there is a great deal of “grey literature”: studies or reports by semi-commercial agencies about their own product. We have tried to limit our use of such materials, but in cases where we had other materials, we occasionally had to base our analysis on this grey literature. A further limitation concerns the people we have interviewed and invited for our focus groups. While we have tried to gather a strong line up of people, including skype interviews with international experts, as always, such a selection is limited, given the time and resources available for this study. This means that if we had included more people, this would have broadened our scope. Of course, this could have led to more specific and slightly different, insights.

Moreover, some people we interviewed were commercially involved in the production of counter-narrative campaigns. While such people are experts in the ins-and-outs of counter-narratives, for them it is also a matter of utmost importance. Arguably, this may undermine their objective focus and impartial judgment on this topic.

Finally, much of the literature on counter-narratives is focused on countering Jihadist propaganda and radical groups with a Salafi-Jihadi signature. We should be cautious to generalize these findings derived from this literature to other groups (e.g., extreme right wings groups). Furthermore, given that there is hardly any strong empirical research available on many issues related to the subject matter, we had to base some of our analysis on the general literature on persuasion and communication campaigns, as well as some “grey literature”. This undermines the options to draw strong conclusions from the current analysis, as one might question the extent to which it is possible to generalize from the general field of persuasion to this particular area of interest. Furthermore, “grey literature” included studies or reports by semi-commercial agencies that are involved in counter-narrative work. While we have tried to limit our use of such materials, we occasionally had to base our analysis on this grey literature.

7.5. FUTURE DIRECTIONS

In the domain of counter-narrative initiatives via social media, an underlying assumption is that (a) people become radical via radical propaganda messages and that (b) presenting people with counter-narratives is a fruitful manner to undermine this. Both assumptions can be questioned. There is little hard evidence from which causal inferences can be made that interaction with extremist content on social media leads to participation in violent extremism. Furthermore, *“The theory that the messages, myths, promises, objectives, glamour and other enticements propagated via violent extremism narratives can be replaced with, or dismantled by, an alternative set of communications is an assumption that remains unproven”* (Ferguson, 2016, p. 16). Partly, this is because radicalization is a complex process that may involve a variety of factors and influences and can be triggered by various motives. For example, based on a literature review, Feddes et al. (2015) distinguish between 4 motives, that people can possess:

- (1) Need for *certainty*: people who feel uncertain may search for groups with a clear identity and focus (i.e., radical groups; Hogg, 2014);

- (2) Need for *significance/meaning*: people who experience a loss in significance are more open to radical groups (e.g., Kruglanski et al., 2014);
- (3) Need for *sensation*: people who have a high need for sensation and adventure are more likely to feel attracted to radical groups because this is exciting and might involve weapons etc. (e.g., Bjørge, 2011);
- (4) Need for *justice*: for some people, a clear *ideology* in which the role of justice for there is being spelled out plays an important role in their radicalization process.

Thus, if one wants to use the method of a counter-narrative program, one should decide on the goals in terms of which aspect one primarily wants to focus. This can be any of the 4 above or any combination of them, because the needs in this list can go hand-in-hand. Thus, solely a focus of undermining an ideology by presenting a counter-narrative in terms of this ideology may not be sufficient to de-radicalize a person that is mostly attracted by the radical groups for its need for sensation, for example.

At the same time, a distinction between different motives may inform first-line workers involved in de-radicalization. If one finds out a person has become radical mainly due to, for example, an attraction for sensation (rather than due to strong ideological motives), a program to de-radicalize this person may include elements to fulfill this need in another manner (rather than containing a strong anti-ideological component).

As indicated in the chapter about the role of the government (Chapter 6), another future direction might be a focus on through curriculum development in support of children and young people's resilience to negative online influences, including extremist propaganda. For instance, by examining existing and developing new teaching approaches for media literacy and critical thinking. At the same time, it may be worthwhile to look into methods that aim to foster competences for teaching media literacy and critical thinking.

With regard to this suggestion, in other domains, it has been examined at which age children are "advertisement-wise" in that they understand what advertisement aims to do (i.e., changing their attitudes and behavior) and how it aims to reach these goals (e.g., Roozendaal, Buijzen, & Valkenburg, 2011; Roozendaal, Oprea, & Buijzen, 2016; Valkenburg, Peter & Walther, 2016).

Generally speaking, age is related to the understanding of advertisement by children. Three developmental phases can be distinguished: early childhood (up to 5 years old), middle childhood (6 to 9 years old) and late childhood (10 to 12 years old). In early childhood, advertisement knowledge is limited. In middle childhood, this knowledge is clearly increasing and most of these children understand that the advertisement is trying to persuade them. By the age of 10-12 years, children also understand how this persuasion might work (Roozendaal et al., 2011). While radical propaganda differs from advertisement in a number of ways, we argue that it is possible to make a direct comparison as well, in terms of the aim to influence another (young) person in their attitudes and/or behavior. This might be a fruitful area of future research to examine closely. At what age do children become aware how propaganda might work? Is it possible to increase this awareness and as such strengthen their resilience? Which factors might be involved in this process?

One idea is that inoculation might work: in this view, it should be possible to stimulate resilience in children by exposing them with a mild form of radical narrative (without violence or blood) and help them to refute this propaganda (i.e., provide them with a potential counter-narrative). Subsequently, when they are confronted with “real propaganda”, they should be able to recognize the common elements of propaganda, and be able to come up with counter-narrative arguments themselves. While the strength of the inoculation theory has been tested in other domains among youth (alcohol, cigarette smoking), the potential of this theory for strengthening resilience against extremist messaging has not been studied.

Another potential future direction might be to examine what happens to narratives and counter-narratives in a virtual world running computer simulations, for example in terms of social network analyses. These simulations might inform us how narratives and counter-narratives might evolve in the long run. However, critically, in order to arrive at reliable estimates of such processes over time in computer simulations, we need to have a clear understanding of the most important factors at play in this context. At the moment, unfortunately, this understanding is far from complete. As such, such computer simulations may be fruitfully explored in a (near) future when this understanding is more advanced.

We conclude that, although presenting online counter-narratives appears to be intuitively an appealing strategy to employ, our analysis shows that this may not be an ideal option to de-radicalize people. It might be more fruitful to use counter-narratives in a prevention context, in which people’s minds may still open enough to register and process the information presented. Alternatively, it might be possible to explore the options to expose known individuals (e.g., a convicted prisoner) with specific counter-narratives, but ideally not in an online context, but an offline context, in order to directly monitor the reactions of the individuals and to build the necessary trust.

REFERENCES

- Ackland, R. (2013). *Web social science: Concepts, data and tools for social scientists in the digital age* [Adobe Digital Editions version]. doi: 10.4135/9781446270011
- Adewuyi, E. O., & Adefemi, K. (2016). Behavior change communication using social media: A review. *International Journal of Communication and Health*, 9, 109–116. Retrieved from: <http://communicationandhealth.ro/upload/number9/EMMANUEL-O-ADEWUYI.pdf>
- Against Violent extremism. (n.d.-a). *About Against Violent Extremism*. Retrieved from <http://www.againstviolentextremism.org/about>
- Against Violent Extremism. (n.d.-b). *FAQ*. Retrieved from <http://www.againstviolentextremism.org/about>
- Ahlqvist, T., Bäck, A., Halonen, M., & Heinonen, S. (2008). *Social media roadmaps: Exploring the futures triggered by social media* (VTT research notes 2454). Retrieved from VTT website: <http://www.vtt.fi/inf/pdf/tiedotteet/2008/t2454.pdf>
- Aistrope, T. (2016). Social media and counterterrorism strategy. *Australian Journal of International Affairs*, 70(2), 121-138. doi: 10.1080/10357718.2015.1113230
- Al Raffie, D. (2012). Whose hearts and minds? Narratives and counter-narratives of Salafi jihadism. *Journal of Terrorism Research*, 3(2), 13-31. doi: 10.15664/jtr.304
- Alkhouri, L., & Kassirer, A. (2016, July). *Tech for jihad: Dissecting jihadists' digital toolbox* [Company Report]. Retrieved from Flashpoint website <https://www.flashpoint-intel.com/wp-content/uploads/2016/08/TechForJihad.pdf>
- Allen, M. (1991). Meta-analysis comparing the persuasiveness of one-sided and two-sided messages. *Western Journal of Speech Communication*, 55(4), 390-404. doi: 10.1080/10570319109374395
- Allen, M. (1998). Comparing the persuasive effectiveness of one-and two-sided messages. In M. Allen & R. W. Preiss (Eds.), *Persuasion: Advances through meta-analysis* (pp. 87-98). Cresskill, NJ: Hampton Press.
- Aly, A. (2017). Brothers, Believers, Brave Mujahideen: Focusing attention on the audience of violent jihadist preachers. *Studies in Conflict & Terrorism*, 40(1), 62-76. doi: 10.1080/1057610X.2016.1157407
- Aly, A., Weimann-Saks, D., & Weimann, G. (2014). Making 'noise' online: an analysis of the say no to terror online campaign. *Perspectives on Terrorism*, 8(5), 33-47. Retrieved from: <http://www.terrorismanalysts.com/pt/index.php/pot/article/view/376>
- Amble, J. C. (2012). Combating terrorism in the new media environment. *Studies in Conflict & Terrorism*, 35(5), 339-353. doi: 10.1080/1057610X.2012.666819
- Appiah, O., Knobloch-Westerwick, S., & Alter, S. (2013). Ingroup favoritism and outgroup derogation: Effects of news valence, character race, and recipient race on selective News Reading. *Journal of Communication*, 63(3), 517-534. doi: 10.1111/jcom.12032

Knobloch-Westerwick, S., & Hastall, M. R. (2010). Please your self: Social Identity effects on selective exposure to news about in- and out-Groups. *Journal of Communication*, 60(3), 515-535. doi: 10.1111/j.1460-2466.2010.01495.x

Archetti, C. (2013). *Understanding Terrorism in the age of global media: A communication approach* [Kindle version]. Retrieved from <http://www.amazon.com>

Arquilla, J., & Ronfeldt, D. (2001). *Networks and netwars: The future of terror, crime, and militancy* (RAND Publication No. MR-1382-OSD) [Monograph Report]. Retrieved from the RAND corporation website: https://www.rand.org/pubs/monograph_reports/MR1382.html

Ashour, O. (2011). Online de-radicalization? Countering violent extremist narratives: message, messenger and media strategy. *Perspectives on Terrorism*, 4(6). Retrieved from <http://www.terrorismanalysts.com/pt/index.php/pot/article/view/128/html>

Atkin, C.K. (2004). Media intervention impact: evidence and promising strategies. In R. J. Bonnie & M. E. O'Connell (Eds.), *Reducing underage drinking: A collective responsibility* (pp. 565–596). Washington: The National Academies Press. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK37580>

Atkin, C. K., & Freimuth, V. (2012). Guidelines for formative evaluation research in campaign design. In R. E. Rice & C. K. Atkin (Eds.), *Public communication campaigns* [Adobe Digital Editions version] (4th ed.). Retrieved from <https://play.google.com>

Atkin, C. K., & Rice, R. E. (2012). Theory and principles of public communication campaigns. In R. E. Rice & C. K. Atkin (Eds.), *Public communication campaigns* [Adobe Digital Editions version] (4th ed.). Retrieved from <https://play.google.com>

Atkin, C.K., & Salmon, C.T. (2013). Persuasive Strategies in Health Campaigns. In J. P. Dillard & L. Shen (Eds.), *The Sage handbook of persuasion: Developments in theory and practice* (2d ed.). Retrieved from <https://play.google.com>

Austin, E. W., & Pinkleton, B. E. (2015). *Strategic public relations management: Planning and managing effective communication campaigns* [Adobe Digital Editions version] (3rd ed.). Retrieved from <https://play.google.com>

Bakker, E., & Grol, P. (2015, July). *Motives and considerations of potential foreign fighters from the Netherlands* [Policy Brief]. Retrieved from International Centre for Counter-Terrorism Website <https://www.icct.nl/wp-content/uploads/2015/07/ICCT-Bakker-Grol-Motives-and-Considerations-of-Potential-Foreign-Fighters-from-the-Netherlands-July2015.pdf>

Banas, J. A., & Miller, G. (2013). Inducing resistance to conspiracy theory propaganda: Testing inoculation and metainoculation strategies. *Human Communication Research*, 39(2), 184-207. doi: 10.1111/hcre.12000

Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs*, 77(3), 281-311. doi: 10.1080/03637751003758193

Baray, G., Postmes, T., & Jetten, J. (2009). When I equals we: Exploring the relation between social and personal identity of extreme right-wing political party members. *British Journal of Social Psychology*, 48(4), 625-647. doi: 10.1348/014466608X389582

- Bartlett, J., Birdwell, J., & King, M. (2010). *The edge of violence: A radical approach to extremism* [Interim Research Report]. Retrieved from the Demos website https://www.demos.co.uk/files/Edge_of_Violence_-_full_-_web.pdf?1291806916
- Bartlett, J., & Krasodonski-Jones, A. (2015, October). *Counter-speech: Examining content that challenges extremism online* [Research Report]. Retrieved from the Demos website <https://www.demos.co.uk/wp-content/uploads/2015/10/Counter-speech.pdf>
- Bartlett, J., & Miller, C. (2010, August). The power of unreason: Conspiracy theories, extremism and counter-terrorism. Retrieved from the Demos website https://www.demos.co.uk/files/Conspiracy_theories_paper.pdf
- Bartlett, J., & Reynolds, L. (2015, September). *The State of the Art 2015: A Literature Review of Social Media Intelligence Capabilities for Counter-terrorism* [Research Paper]. Retrieved from the Demos website https://www.demos.co.uk/wp-content/uploads/2015/09/State_of_the_Arts_2015.pdf
- Baxter, L. A., & Babbie, E. R. (2004). *The basics of communication research*. Belmont, CA: Wadsworth/Thomson Learning.
- Bazarova, N. N., Walther, J. B., & McLeod, P. L. (2012). Minority influence in virtual groups. *Communication Research*, 39(3), 295-316. doi: 10.1177/0093650211399752
- Belk, R., & Kozinetz, R. (2016). Videography and netnography. In K. Kubacki & S. Rundle-Thiele (Eds.), *Formative research in social marketing: Innovative methods to gain consumer insights* (pp. 265-279). Singapore: Springer. doi: 10.1007/978-981-10-1829-9_13
- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192-205. doi: 10.1509/jmr.10.0353
- Berger, J.M., & Perez, H. (2016, February). *The Islamic States diminishing returns on Twitter: How suspensions are limiting the social networks of English-speaking ISIS supporters* [Occasional Paper]. Retrieved from George Washington University Program on Extremism website <https://extremism.gwu.edu/sites/extremism.gwu.edu/files/downloads/JMB%20Diminishing%20Returns.pdf>
- Berger, J. M. (2016, May). *Making CVE work: A focused approach based on process disruption* [Research Paper]. Retrieved from International Centre for Counter-terrorism website <https://www.icct.nl/wp-content/uploads/2016/05/J.-M.-Berger-Making-CVE-Work-A-Focused-Approach-Based-on-Process-Disruption-.pdf>
- Berger, J. M., & Morgan, J. (2015, March). *The ISIS twitter census: Defining and describing the population of ISIS supporters on Twitter* (The Brookings Project on U.S. Relations with the Islamic World Analysis Paper No. 20). Retrieved from the Brookings Institution website https://www.brookings.edu/wp-content/uploads/2016/06/ISIS_twitter_census_berger_morgan.pdf
- Birmingham, A., Conway, M., McInerney, L., Hare, N. O., & Smeaton, A. F. (2009, July). *Combining social network analysis and sentiment analysis to explore the potential for online radicalisation*. Paper presented at the 2009 International Conference on Advances in Social Network Analysis and Mining. Retrieved from <http://doras.dcu.ie/4554/>

- Bethlehem, J. (2010). Selection bias in web surveys. *International Statistical Review*, 78(2), 161-188. doi: 10.1111/j.1751-5823.2010.00112.x
- Beutel, A., Weine, S. M., Aliya Saeed, Mihajlovic, A. S., Stone, A., Beahrs, J. O., & Shanfield, S. B. (2016). Guiding principles for countering and displacing extremist narratives. *Journal of Terrorism Research*, 7(3), 35-49. doi: 10.15664/jtr.1220
- Bijvank, M. N., Konijn, E. A., Bushman, B. J., & Roelofsma, P. H. M. P. (2009). Age and violent-content labels make video games forbidden fruits for youth. *Pediatrics*, 123(3), 870-876. doi: 10.1542/peds.2008-0601
- Bilandzic, H., & Busselle, R. (2013). Narrative persuasion. In J. P. Dillard & L. Shen (Eds.), *The Sage handbook of persuasion: Developments in theory and practice* [Adobe Digital Editions version] (2nd ed.). Retrieved from <https://play.google.com>
- Bjørge, T. (2011). Dreams and disillusionment: engagement in and disengagement from militant extremist groups. *Crime, Law and Social Change*, 55(4), 277-285. doi: 10.1007/s10611-011-9282-9
- Bole, R., & Kallmyer, K. (2016). Combatting the Islamic State's digital dominance: Revitalizing U.S. communication strategy. *The Washington Quarterly*, 39(1), 29-48. doi: 10.1080/0163660X.2016.1170478
- Borum, R. (2011). Radicalization into violent extremism I: A review of social science theories. *Journal of Strategic Security*, 4(4), 7-36. doi: 10.5038/1944-0472.4.4.1
- Boslaugh, S. E., Kreuter, M. W., Nicholson, R. A., & Naleid, K. (2005). Comparing demographic, health status and psychosocial strategies of audience segmentation to promote physical activity. *Health Education Research*, 20(4), 430-438. doi: 10.1093/her/cyg138
- Braddock, K., & Horgan, J. (2016). Towards a guide for constructing and disseminating counternarratives to reduce support for terrorism. *Studies in Conflict & Terrorism*, 39(5), 381-404. doi: 10.1080/1057610X.2015.1116277
- Braddock, K., & Dillard, J. P. (2016). Meta-analytic evidence for the persuasive effect of narratives on beliefs, attitudes, intentions, and behaviors. *Communication Monographs*, 83(4), 446-467. doi: 10.1080/03637751.2015.1128555
- Brannon, L. A., Tagler, M. J., & Eagly, A. H. (2007). The moderating role of attitude strength in selective exposure to information. *Journal of Experimental Social Psychology*, 43(4), 611-617. doi: <https://doi.org/10.1016/j.jesp.2006.05.001>
- Brehm, J. W. (1966). *A theory of psychological reactance*. New York, NY: Academic Press.
- Brehm, S. S., & Brehm, J. W. (1981). *Psychological reactance: A theory of freedom and control*. New York, NY: Academic press.
- Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, 17(5), 475-482. doi: 10.1177/0146167291175001
- Briggs, R., & Feve, S. (2013). *Review of programs to counter narratives of violent extremism: What works and what are the implications for government* [Research Report]. Retrieved from the Public Safety Canada website <https://www.publicsafety.gc.ca/lbrr/archives/cn28580-eng.pdf>

- Briggs, R., & Frenett, R. (2014). *Foreign fighters, the challenge of counter-narratives* [Policy Briefing]. Retrieved from the Institute of Strategic Dialogue Counter Extremism website <https://www.counterextremism.org/resources/details/id/683/foreign-fighters-the-challenge-of-counter-narratives>
- Briggs, R., & Silverman, T. (2014). *Western foreign fighters: Innovations in responding to the threat* [Research Paper]. Retrieved from the Institute for Strategic Dialogue website https://www.isdglobal.org/wp-content/uploads/2016/02/ISDJ2784_Western_foreign_fighters_V7_WEB.pdf
- Buijzen, M., & Valkenburg, P. M. (2004). Developing a typology of humor in audiovisual media. *Media Psychology, 6*(2), 147-167. doi: 10.1207/s1532785xmep0602_2
- Bunzel, C. (2016, March). *From paper state to caliphate: The ideology of the Islamic State* (The Brookings Project on U.S. Relations with the Islamic World Analysis Paper No. 19). Retrieved from the Brookings Institution website: <https://www.brookings.edu/wp-content/uploads/2016/06/The-ideology-of-the-Islamic-State.pdf>
- Burgoon, M., Alvaro, E., Grandpre, J., & Voulodakis, M. (2002). Revisiting the theory of psychological reactance: Communicating threats to attitudinal freedom. In J. P. Dillard & M. Pfau (Eds.), *The persuasion handbook: Developments in theory and practice* (pp. 213-232). Thousand Oaks, CA: Sage.
- Bushell, S., Workman, M., & Colley, T. (2016, April). *Towards a unifying narrative for climate change* (Grantham Institute Briefing Paper No. 19). Retrieved from the Imperial College London Website <https://www.imperial.ac.uk/media/imperial-college/grantham-institute/public/publications/briefing-papers/Towards-a-unifying-narrative-for-climate-change-Grantham-BP18.pdf>
- Campo, S., Askelson, N. M., Spies, E. L., Boxer, C., Scharp, K. M., & Losch, M. E. (2013). "Wow, that was funny": The value of exposure and humor in fostering campaign message sharing. *Social Marketing Quarterly, 19*(2), 84-96. doi: 10.1177/1524500413483456
- Carr, C. T., & Hayes, R. A. (2015). Social Media: Defining, Developing, and Divining. *Atlantic Journal of Communication, 23*(1), 46-65. doi: 10.1080/15456870.2015.972282
- Carter, J. A., Maher, S., & Neumann, P. R. (2014, April). *#Greenbirds: Measuring importance and influence in Syrian foreign fighter networks* [Research Report]. Retrieved from International Centre for the Study of Radicalization and Political Violence website <http://icsr.info/wp-content/uploads/2014/04/ICSR-Report-Greenbirds-Measuring-Importance-and-Influence-in-Syrian-Foreign-Fighter-Networks.pdf>
- Choudhary, P., & Singh, U. (2015). A survey on social network analysis for counter-terrorism. *International Journal of Computer Applications, 112*(9). doi: 10.5120/19695-1455
- Civettini, A. J., & Redlawsk, D. P. (2009). Voters, emotions, and memory. *Political Psychology, 30*(1), 125-151. doi: 10.1111/j.1467-9221.2008.00683.x
- Coffman, J. (2002, May). *Public communication campaign evaluation: An environmental scan of challenges, criticisms, practice, and opportunities* [Research Paper]. Retrieved from the Harvard Family Research Project website: <http://www.hfrp.org/evaluation/publications-resources/public->

communication-campaign-evaluation-an-environmental-scan-of-challenges-criticisms-practice-and-opportunities

Cohen, J. (2001). Defining identification: A theoretical look at the identification of audiences with media characters. *Mass Communication and Society*, 4(3), 245-264. doi: 10.1207/S15327825MCS0403_01

Cohen, J., Tal-Or, N., & Mazor-Tregerman, M. (2015). The tempering effect of transportation: Exploring the effects of transportation and identification during exposure to controversial two-sided narratives. *Journal of Communication*, 65(2), 237-258. doi: 10.1111/jcom.12144

Cohen, J., Weimann-Saks, D., & Mazor-Tregerman, M. (2017). Does character similarity increase identification and persuasion? *Media Psychology*, 1-23. doi: 10.1080/15213269.2017.1302344

Cohen, K., Johansson, F., Kaati, L., & Mork, J. C. (2014). Detecting linguistic markers for radical violence in social media. *Terrorism and Political Violence*, 26(1), 246-256. doi: 10.1080/09546553.2014.849948

Conway, M. (2017). Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research. *Studies in Conflict & Terrorism*, 40(1), 77-98. doi: 10.1080/1057610X.2016.1157408

Conway, M., & McInerney, L. (2008). Jihadi video and auto-radicalisation: Evidence from an exploratory YouTube study. In D. O.-A. H. Larsen, D. Zeng & D. Hick (Eds.), *Intelligence and Security Informatics* (pp. 108-118). Berlin, DE: Springer Verlag.

Compton, J. (2013). Inoculation theory. In J. P. Dillard & L. Shen (Eds.), *The Sage handbook of persuasion: Developments in theory and practice* [Adobe Digital Editions version] (2nd ed.). Retrieved from <https://play.google.com>

Corman, S. R. (2011, September). Understanding the role of narrative in extremist strategic communication. In F. Laurie & S. Leventhall (Eds.), *Countering Violent Extremism: Scientific Methods and Strategies* (pp. 36-43). Retrieved from <http://nsiteam.com/social/wp-content/uploads/2016/01/Countering-Violent-Extremism-Scientific-Methods-Strategies.pdf>

Corman, S. R., & Schiefelbein, J. S. (2006). Communication and media strategy in the Jihadi war of ideas (White Paper No. 0601). Retrieved from the Consortium for Strategic Communication website <http://csc.asu.edu/wp-content/uploads/2012/06/119.pdf>

Costello, L., McDermott, M.L., & Wallace, R. (2017). Netnography: Range of practices, misperceptions, and missed opportunities. *International Journal of Qualitative Methods*, 16(1), 1609406917700647.

Coull, A., Yzerbyt, V. Y., Castano, E., Paladino, M.P., & Leemans, V. (2001). Protecting the Ingroup: Motivated allocation of cognitive resources in the presence of threatening ingroup members. *Group Processes & Intergroup Relations*, 4(4), 327-339. doi:10.1177/1368430201004004003

Counter-Narrative Toolkit. (n.d.). *About*. Retrieved from <http://www.counternarratives.org/about-us>

- Dafnos, A. (2014). Narratives as a means of countering the radical Right; Looking into the Trojan T-Shirt project. *Journal EXIT-Deutschland*, 3, 156-188. Retrieved from <http://journals.sfu.ca/jed/index.php/jex/article/viewFile/98/126>
- Dahl, S. (2014). *Social Media Marketing: Theories and Applications* [Adobe Digital Editions version]. Retrieved from <https://play.google.com>
- Dal Cin, S., Zanna, M. P., & Fong, G. T. (2004). Narrative persuasion and overcoming resistance. In E. S. Knowles & J. A. Linn (Eds.), *Resistance and persuasion* (pp. 175-191). Mahwah, NJ: Erlbaum.
- Dalgaard-Nielsen, A. (2010). Violent radicalization in Europe: What we know and what we do not know. *Studies in Conflict & Terrorism*, 33(9), 797-814. doi: 10.1080/1057610X.2010.501423
- Dalgaard-Nielsen, A. (2013). Promoting exit from violent extremism: Themes and approaches. *Studies in Conflict & Terrorism*, 36(2), 99-115. doi: 10.1080/1057610X.2013.747073
- Dare to be Grey. (c2017). *#Dare to be Grey*. Retrieved from <http://www.dtbg.nl/en>
- Davies, G., Neudecker, C., Ouellet, M., Bouchard, M., & Ducol, B. (2016). Toward a framework understanding of online programs for countering violent extremism. *Journal for Deradicalization*(6), 51-86. Retrieved from <http://journals.sfu.ca/jd/index.php/jd/article/view/43/0>
- Dawson, L., Edwards, C., & Jeffray, C. (2014, May). *Learning and adapting: The use of monitoring and evaluation in countering violent extremism* [E-handbook]. Retrieved from the Royal United Services Institute website https://rusi.org/sites/default/files/201406_bk_learning_and_adapting.pdf
- de Graaf, A., Hoeken, H., Sanders, J., & Beentjes, J. W. J. (2012). Identification as a mechanism of narrative persuasion. *Communication Research*, 39(6), 802-823. doi: 10.1177/0093650211408594
- de Graaf, B., Dimitriu, G., & Ringsmose, J. (2015). Introduction: Shaping societies for war: Strategic narratives and public opinion. In B. de Graaf, G. Dimitriu & J. Ringsmose (Eds.), *Strategic narratives, public opinion and war: Winning domestic support for the Afghan War* [Adobe Digital Editions version]. Retrieved from <https://play.google.com>
- de Koning, M. (2013). Between the Prophet and Paradise: The Salafi struggle in the Netherlands. *Canadian Journal of Netherlandic Studies*, 33(2)-34(1), 17-34.
- de Wolf, A., & Doosje, B. (2010). *Aanpak van radicalisme: een psychologische analyse* [Dealing with radicalization: A psychological analysis]. Amsterdam, NL: SWP
- Dean, G. (2016). Framing the challenges of online violent extremism: “Policing-public-policies-politics” framework. In M. Khader, L. S. Neo, G. Ong, E. Tan & J. Chin (Eds.), *Countering violent extremism and radicalisation in the digital era* [Adobe Digital Editions version] (pp. 226-269). doi: 10.4018/978-1-5225-0156-5.ch012
- Demant, F., Sloomman, M., Buijs, F., & Tillie, J. (2008). *Decline and disengagement: An analysis of processes of deradicalisation*. Amsterdam, NL: Institute for Migration and Ethnic Studies.
- Demos. (2016, 13 December). *Pioneering school pilot a breakthrough in building students’ resilience to extremism online* [Press release]. Retrieved 10 September, 2017, from

<https://www.demos.co.uk/press-release/pioneering-school-pilot-a-breakthrough-in-building-students-resilience-to-extremism-online/>

Dillard, J. P., Plotnick, C. A., Godbold, L. C., Freimuth, V. S., & Edgar, T. (1996). The multiple affective outcomes of AIDS PSAs. *Communication Research*, 23(1), 44-72. doi: 10.1177/009365096023001002

Dillard, J. P., & Shen, L. (2005). On the nature of reactance and its role in persuasive health communication. *Communication Monographs*, 72(2), 144-168. doi: 10.1080/03637750500111815

Dimitriu, G., & de Graaf, B. (2016). Fighting the war at home: strategic narratives, elite responsiveness, and the Dutch mission in Afghanistan, 2006–2010. *Foreign Policy Analysis*, 12, 2-23. doi: 10.1111/fpa.12070

Doosje, B., Loseman, A., & Bos, K. (2013). Determinants of radicalization of Islamic youth in the Netherlands: Personal uncertainty, perceived injustice, and perceived group threat. *Journal of Social Issues*, 69(3), 586-604. doi: 10.1111/josi.12030

Doosje, B., Moghaddam, F. M., Kruglanski, A. W., de Wolf, A., Mann, L., & Feddes, A. R. (2016). Terrorism, radicalization and de-radicalization. *Current Opinion in Psychology*, 11, 79-84. doi: 10.1016/j.copsyc.2016.06.008

Eagly, A. H., & Chaiken, S. (1984). Cognitive Theories of Persuasion. *Advances in Experimental Social Psychology*, 17, 267-359. doi: 10.1016/S0065-2601(08)60122-7

Edelman, A. (2014, 16 September). State department's 'embarrassing' 'Think Again Turn Away' Twitter campaign could actually legitimize terrorists: Expert. *New York Daily News*. Retrieved from <http://www.nydailynews.com/news/politics/state-department-embarrassing-turn-twitter-campaign-legitimizes-terrorists-expert-article-1.1941990>.

EdVenture Partners. (c2017). *Peer to peer*. Retrieved from <https://edventurepartners.com/peer2peer>

Edwards, C., & Gribbon, L. (2013). Pathways to violent extremism in the digital Era. *The RUSI Journal*, 158(5), 40-47. doi: 10.1080/03071847.2013.847714

Egner, M. (2009). Social-science foundations for strategic communications in the global war on Terrorism. In P. K. Davis, K. Cragin & D. Noricks (Eds.), *Social science for counterterrorism: Putting the pieces together* [e-book] (pp. 323-355). Retrieved from the RAND corporation website https://www.rand.org/content/dam/rand/pubs/monographs/2009/RAND_MG849.pdf

Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual review of psychology*, 53(1), 161-186. doi: 10.1146/annurev.psych.53.100901.135228

Ernst, J., Schmitt, J. B., Rieger, D., Beier, A. K., Vorderer, P., Bente, G., & Roth, H.J. (2017). Hate beneath the counter Speech? A qualitative content analysis of user comments on YouTube related to counter speech videos. *Journal for Deradicalization*, 10, 1-49. Retrieved from <http://journals.sfu.ca/jd/index.php/jd/article/view/91>

Escalas, J. E. (2004). Imagine yourself in the product: Mental simulation, narrative transportation, and persuasion. *Journal of advertising*, 33(2), 37-48. doi: 10.1080/00913367.2004.10639163

- Esposito, S. R., Hornsey, M. J., & Spoor, J. R. (2013). Shooting the messenger: Outsiders critical of your group are rejected regardless of argument quality. *British Journal of Social Psychology*, 52(2), 386-395. doi: 10.1111/bjso.12024
- Evers, C. W., Albury, K., Byron, P., & Crawford, K. (2013). Young people, social media, social network sites and sexual health communication in Australia: "This is funny, you should watch it". *International Journal of Communication*, 7. Retrieved from <http://ijoc.org/index.php/ijoc/article/download/1106/853>
- Eysenbach, G., & Wyatt, J. (2002). Using the Internet for surveys and health research. *Journal of Medical Internet Research*, 4(2), e13. doi: 10.2196/jmir.4.2.e13
- Fairlie, B. M. (2017). The spread of information via social media. In J. V. Cohn, S. Schatz, H. Freeman & D. J. Y. Combs (Eds.), *Modeling Sociocultural Influences on Decision Making: Understanding Conflict, Enabling Stability* (pp. 43-61). Boca Raton, FL: CRC Press.
- Feddes, A. R., Nickolson, L., & Doosje, B. (2015). *Triggerfactoren in het radicaliseringsproces* [Trigger factors in the radicalisation process]. Retrieved from the Social Stability Expertise Unit, Dutch Ministry of Social Affairs & Employment website <https://www.socialestabiliteit.nl/documenten/publicaties/2015/10/13/triggerfactoren-in-het-radicaliseringsproces>
- Felten, H., Taouanza, I., & Keuzenkamp, S. (2016, June). *Klaar met discriminatie? Onderzoek naar effectiviteit van sociale media-campagnes tegen discriminatie* [Done with discrimination? Research into the effectiveness of social media campaigns against discrimination]. Retrieved from Platform Integration & Society <https://www.kis.nl/publicatie/klaar-met-discriminatie>
- Ferguson, K. (2016, March). *Countering violent extremism through media and communication strategies: A review of the evidence* [Research Report]. Retrieved from the Partnership for Conflict, Crime & Security Research website <http://www.paccsresearch.org.uk/wp-content/uploads/2016/03/Countering-Violent-Extremism-Through-Media-and-Communication-Strategies-.pdf>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row & Peterson.
- Fink, N. C., & Barclay, J. (2013, February). *Mastering the narrative: Counterterrorism strategic communications and the United Nations* [Research Report]. Retrieved from the Center on Global Counterterrorism Cooperation website <http://www.globalcenter.org/publications/mastering-the-narrative-counterterrorism-strategic-communication-and-the-united-nations/>
- Fisher, A. (2015). How jihadist networks maintain a persistent online presence. *Perspectives on Terrorism*, 9(3), 3-20. Retrieved from <http://www.terrorismanalysts.com/pt/index.php/pot/article/view/426/html>
- Fransen, M. L., Smit, E. G., & Verlegh, P. W. J. (2015). Strategies and motives for resistance to persuasion: an integrative framework. *Frontiers in Psychology*, 6, 1201. doi: 10.3389/fpsyg.2015.01201
- Freedman, L. (2006). *The transformation of strategic affairs*. London: Routledge.

- Freedman, L. (2015). The possibilities and limits of strategic narratives. In B. De Graaf, G. Dimitriu & J. Ringsmose (Eds.), *Strategic narratives, public opinion and war: Winning domestic support for the Afghan War* [Adobe Digital Editions]. Retrieved from <https://play.google.com>
- Freeman, B., Potente, S., Rock, V., & McIver, J. (2015). Social media campaigns that make a difference: what can public health learn from the corporate sector and other social change marketers. *Public Health Research Practice*, 25(2), 2-8. doi: 10.17061/phrp2521517
- Freimuth, V., Cole, G., & Kirby, S. (2011). Issues in evaluating mass media-based health communication campaigns. In J. S. Detrani (Ed.), *Mass communication: Issues, perspectives and techniques* (pp. 77–98). Oakville, ON, CAN: Apple Academic Press.
- Frennet, R. & Dow, M. (2015, September). *One to One online interventions: A pilot CVE methodology* [Research Report]. Retrieved from the Institute for Strategic Dialogue website https://www.isdglobal.org/wp-content/uploads/2016/04/One2One_Web_v9.pdf
- Frenett, R., & Silverman, T. (2016). Foreign fighters: Motivations for travel to foreign conflicts. In A. de Guttery, F. Capone & C. Paulussen (Eds.), *Foreign fighters under International Law and beyond* (pp. 63-76). The Hague, NL: T.M.C. Asser Press.
- Geeraerts, S. B. (2012). Digital radicalization of youth. *Social cosmos*, 3(1), 25-32. Retrieved from <https://dspace.library.uu.nl/bitstream/handle/1874/237584/41-150-1-PB.pdf%3Bsequence=2>
- German Institute on Radicalization and De-radicalization Studies (GIRDS). (n.d.). *Open Letter to our Sons and Daughters in Syria and Iraq*. Retrieved from <http://girds.org/mothersforlife/open-letter-to-our-sons-and-daughters>
- German Institute on Radicalization and De-radicalization Studies (GIRDS). (n.d.). *A Second Letter to Abu Bakr al-Baghdadi from the Mothers for Life*. Retrieved from <http://girds.org/mothersforlife/a-second-letter-to-abu-bakr-al-baghdadi-from-the-mothers-for-life>
- Gielen, A. J. (2017). Countering Violent Extremism: A realist review for assessing what works, for whom, in what circumstances, and how? *Terrorism and Political Violence*, 1-19. doi: 10.1080/09546553.2017.1313736
- Dutch General Intelligence and Security Service (GISS). (2014). *The transformation of jihadism in the Netherlands Swarm dynamics and new strength*. The Hague: General Intelligence and Security Service, Ministry of the Interior Kingdom Relations of the Netherlands. Retrieved from <https://english.aivd.nl/publications/publications/2014/10/01/the-transformation-of-jihadism-in-the-netherlands>
- Godbold, L. C., & Pfau, M. (2000). Conferring resistance to peer pressure among adolescents. *Communication Research*, 27(4), 411-437. doi: 10.1177/009365000027004001
- Gold, J., Pedrana, A. E., Stooze, M. A., Chang, S., Howard, S., Asselin, J., & Hellard, M. E. (2012). Developing health promotion interventions on social networking sites: Recommendations from The FaceSpace project. *Journal of Medical Internet Research*, 14(1), e30. doi: 10.2196/jmir.1875
- Goodall, H., Cheong, P. H., Fleischer, K., & Corman, S. R. (2012). Rhetorical charms: The promise and pitfalls of humor and ridicule as strategies to counter extremist narratives. *Perspectives on*

- Terrorism*, 6(1). Retrieved from <http://www.terrorismanalysts.com/pt/index.php/pot/article/view/goodall-et-al-rhetorical/html>
- Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, 26(3), 213-224. doi: 10.1002/bdm.1753
- Gouvernement.fr. (2015, 28 January). #StopJihadism: everyone on alert and taking action against Jihadism. Retrieved from <http://www.gouvernement.fr/en/stopjihadism-everyone-on-alert-and-taking-action-against-jihadism>
- Grandpre, J., Alvaro, E. M., Burgoon, M., Miller, C. H., & Hall, J. R. (2003). Adolescent reactance and anti-smoking campaigns: A theoretical approach. *Health communication*, 15(3), 349-366. doi: 10.1207/S15327027HC1503_6
- Green, J. (2000). The role of theory in evidence-based health promotion practice. *Health Education Research*, 15(2), 125-129. doi: 10.1093/her/15.2.125
- Green, M. C. (2004). Transportation into narrative worlds: The role of prior knowledge and perceived realism. *Discourse Processes*, 38(2), 247-266. doi: 10.1207/s15326950dp3802_5
- Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of personality and social psychology*, 79(5), 701-721. doi: 10.1037//0022-3514.79.5.701
- Greene, K. J. (2015). ISIS: Trends in terrorist media and propaganda. *International Studies Capstone Research Papers*, 3. 1-58 Retrieved from http://digitalcommons.cedarville.edu/international_studies_capstones/3
- Grossman, M. (2015). Disenchantments: counter-terror narratives and conviviality. In F. Mansouri (Ed.), *Cultural, Religious and Political Contestations: The Multicultural Challenge* (pp. 71-89). Cham: Springer.
- Global Counter-Terrorism Forum (GCTF). (2013). *Good practices on community engagement and community-oriented policing as tools to counter violent extremism* [Best Practice Document]. Retrieved from the GCTF website https://www.thegctf.org/documents/10162/159885/13Aug09_EN_Good+Practices+on+Community+Engagement+and+Community-Oriented+Policing.pdf
- Guadagno, R. E., Rempala, D. M., Murphy, S., & Okdie, B. M. (2013). What makes a video go viral? An analysis of emotional contagion and Internet memes. *Computers in Human Behavior*, 29(6), 2312-2319. doi: <http://dx.doi.org/10.1016/j.chb.2013.04.016>
- Hafez, M., & Mullins, C. (2015). The Radicalization puzzle: A theoretical synthesis of empirical approaches to homegrown extremism. *Studies in Conflict & Terrorism*, 38(11), 958-975. doi: 10.1080/1057610X.2015.1051375
- Halverson, J. R., Corman, S. R., & Goodall, H. L., Jr. (2013). *Master narratives of Islamist extremism* (Paperback ed.). New York, NY: Palgrave Macmillan.
- Hanna, R., Rohm, A., & Crittenden, V. L. (2011). We're all connected: The power of the social media ecosystem. *Business horizons*, 54(3), 265-273. doi: <http://dx.doi.org/10.1016/j.bushor.2011.01.007>

- Harmon-Jones, E. (2012). Cognitive dissonance theory. In V. S. Ramachandran (Ed.), *The Encyclopedia of Human Behavior* (2nd ed., Vol. 1, pp. 543-549). Cambridge, MA: Academic Press.
- Harmon-Jones, E., & Harmon-Jones, C. (2007). Cognitive dissonance theory after 50 years of development. *Zeitschrift für Sozialpsychologie*, *38*(1), 7-16. doi: 10.1024/0044-3514.38.1.7.
- Harmon-Jones, E., Harmon-Jones, C., & Levy, N. (2015). An Action-Based Model of Cognitive-Dissonance Processes. *Current Directions in Psychological Science*, *24*(3), 184-189. doi: 10.1177/0963721414566449
- Harmon-Jones, E., & Mills, J. (1999). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (pp. 3-21). Washington, DC: American Psychological Association.
- Harris-Hogan, S., Barrelle, K., & Zammit, A. (2016). What is countering violent extremism? Exploring CVE policy and practice in Australia. *Behavioral Sciences of Terrorism and Political Aggression*, *8*(1), 6-24. doi: 10.1080/19434472.2015.1104710
- Harris, K., Gringart, E., & Drake, D. (2014, December). *Understanding the role of social groups in radicalisation*. Paper presented at the 7th Australian Security and Intelligence Conference, Edith Cowan University, Perth, AUS. doi: 10.4225/75/57a83235c833d
- Harris, K. J. (2011, December) *Entitativity and ideology: a grounded theory of disengagement*. Paper presented at the 4th Australian Security and Intelligence Conference, Edith Cowan University, Perth, AUS. doi: 10.4225/75/57a01cd3ac5c7
- Hedayah & International Center for Counter-Terrorism (ICCT). (2014). *Developing effective counter narrative frameworks for countering violent extremism* [Meeting Note]. Retrieved from the International Center for Counter-Terrorism website <https://icct.nl/publication/developing-effective-counter-narrative-frameworks-for-countering-violent-extremism>
- Hegghammer, T. (2014). *Interpersonal trust on Jihadi internet forums*. Norwegian Defence Research Establishment. Retrieved from http://hegghammer.com/_files/interpersonal_trust.pdf
- Heldman, A. B., Schindelar, J., & Weaver, J. B. (2013). Social media engagement and public health communication: Implications for public health organizations being truly “Social”. *Public Health Reviews*, *35*(1), 13. doi: 10.1007/bf03391698
- Helmus, T. C., York, E., & Chalk, P. (2013). *Promoting online voices for countering violent extremism* (RAND Publication No. RR-130-OSD) [Research Report]. Retrieved from the RAND corporation website https://www.rand.org/content/dam/rand/pubs/research_reports/RR100/RR130/RAND_RR130.pdf
- Hemmingsen, A.S., & Castro, K. I. (2017). *The trouble with counter-narratives* (DIIS Report no. 1). Retrieved from the Danish Institute for International Studies website: http://pure.diis.dk/ws/files/784884/DIIS_RP_2017_1.pdf
- Hoeken, H., & Flikkers, K. M. (2014). Issue-relevant thinking and identification as mechanisms of narrative persuasion. *Poetics*, *44*, 84-99. doi: 10.1016/j.poetic.2014.05.001

- Hogg, M. A. (2000). Subjective uncertainty reduction through self-categorization: A motivational theory of social identity processes. *European Review of Social Psychology*, 11(1), 223-255. doi: 10.1080/14792772043000040
- Hogg, M. A. (2014). From uncertainty to extremism: Social categorization and identity processes. *Current Directions in Psychological Science*, 23, 338-342. doi: 10.1177/0963721414540168
- Holtmann, P. (2013). Countering al-Qaeda's single narrative. *Perspectives on Terrorism*, 7(2), 142-146. Retrieved from <http://www.terrorismanalysts.com/pt/index.php/pot/article/view/262/html>
- Hong, S.-M., Giannakopoulos, E., Laing, D., & Williams, N. A. (1994). Psychological reactance: Effects of age and gender. *The Journal of Social Psychology*, 134(2), 223-228. doi: 10.1080/00224545.1994.9711385
- Horgan, J. (2008). Deradicalization or disengagement? *Perspectives on Terrorism*, 2(4), 3-8. Retrieved from <http://www.terrorismanalysts.com/pt/index.php/pot/article/view/32/html>
- Hornsey, M. J. (2008). Social identity theory and self-categorization theory: A historical review. *Social and Personality Psychology Compass*, 2(1), 204-222. doi: 10.1111/j.1751-9004.2007.00066.x
- Howe, L. C., & Krosnick, J. A. (2017). Attitude Strength. *Annual review of psychology*, 68(1), 327-351. doi: 10.1146/annurev-psych-122414-033600
- Hu, Y., & Sundar, S. S. (2009). Effects of online health sources on credibility and behavioral intentions. *Communication Research*, 37(1), 105-132. doi: 10.1177/0093650209351512
- Hui, J. Y. (2016). Social media analytics for intelligence and counter-extremism. In M. Khader, Loo Seng Neo, G. Ong, E. T. Mingyi & J. Chin (Eds.), *Combating violent extremism and radicalization in the digital era* [Adobe Digital Editions version] (pp. 328-348). doi: 10.4018/978-1-5225-0156-5.ch001
- Hussain, G., & Saltman, E. M. (2014). *Jihad trending: A comprehensive analysis of online extremism and how to counter it* [Research report]. Retrieved from the Quilliam Foundation website <http://www.quilliamfoundation.org/wp/wp-content/uploads/publications/free/jihad-trending-quilliam-report.pdf>
- Igartua, J. J., & Vega, J. (2015, May). *Identification with characters and cognitive processes in entertainment-education interventions through audiovisual fiction*. Paper presented at the Mass Communication Division of the International Communication Association 65th Annual Conference, San Juan, Puerto Rico. Retrieved from https://www.researchgate.net/profile/Juan_Jose_Igartua/publication/280314331_Identification_With_Characters_and_Cognitive_Processes_in_Entertainment-Education_Interventions_Through_Audiovisual_Fiction/links/55b205b008aed621ddfd7a9b.pdf
- Iles, I. A., & Nan, X. (2017). It's no laughing matter: An exploratory study of the use of ironic versus sarcastic humor in health-related advertising messages. *Health marketing quarterly*, 34(3), 187-201. doi: 10.1080/07359683.2017.1346432
- Ingram, H. J., & Reed, A. G. (2016). *Lessons from history for counter-terrorism strategic communications* [Policy Brief]. Retrieved from the International Centre for Counter-Terrorism website <https://icct.nl/publication/a-brief-history-of-propaganda-during-conflict-a-lesson-for-counter-terrorism-strategic-communications>

- Institute for Strategic Dialogue (ISD). (2014). *Government engagement and communication strategies with communities* [Policy Paper]. Retrieved from the Institute for Strategic Dialogue Counter Extremism website https://www.counterextremism.org/download_file/257/134/540/
- Jacobson, M. (2010). Terrorist drop-outs: One way of promoting a counter-narrative. *Perspectives on Terrorism*, 3(2). Retrieved from <http://www.terrorismanalysts.com/pt/index.php/pot/article/view/66/html>
- Jervis, R. (2017). *Perception and misperception in international politics*. Princeton, NJ: Princeton University Press.
- Jigsaw. (n.d.). *The Redirect Method: A blueprint for bypassing extremism* [Project Information Document]. Retrieved from <https://redirectmethod.org/downloads/RedirectMethod-FullMethod-PDF.pdf>
- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192-205. doi: 10.1509/jmr.10.0353
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business horizons*, 53(1), 59-68. doi: :10.1016/j.bushor.2009.09.003
- Katz, R. (2014, September 26). The state department's Twitter war with ISIS is embarrassing. *Time*. Retrieved from <http://time.com/3387065/ISIS-twitter-war-state-department>
- Khalil, J. (2014). Radical beliefs and violent actions are not synonymous: How to place the key disjuncture between attitudes and behaviors at the heart of our research into political violence. *Studies in Conflict & Terrorism*, 37(2), 198-211. doi: 10.1080/1057610X.2014.862902
- Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business horizons*, 54(3), 241-251. doi: 10.1016/j.bushor.2011.01.005
- Kim, A., Lee, S., Marguleas, O., & Beyer, J. L. (2016, October 3). *Do counter-narrative programs slow terrorist recruiting?* (JSIS Cybersecurity Report). Retrieved from The Henry M. Jackson School of International Studies website <https://jsis.washington.edu/news/jsis-cybersecurity-report-counter-narrative-programs-stop-terrorist-recruiting-efforts>
- Kitzinger, J. (1995). Qualitative research. Introducing focus groups. *British Medical Journal*, 311(7000), 299-302. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2550365/pdf/bmj00603-0031.pdf>
- Klausen, J. (2015). Tweeting the jihad: Social media networks of Western foreign fighters in Syria and Iraq. *Studies in Conflict & Terrorism*, 38(1), 1-22. doi: 10.1080/1057610X.2014.974948
- Knobloch-Westerwick, S., & Hastall, M. R. (2010). Please yourself: Social Identity effects on selective exposure to news about in- and out-Groups. *Journal of Communication*, 60(3), 515-535. doi: 10.1111/j.1460-2466.2010.01495.x
- Konijn, E., Oegema, D., Schneider, I., Vos, B. d., Krijt, M., & Prins, J. (2010). *Jong en multimediaal: Mediagebruik en meningsvorming onder jongeren, in het bijzonder moslimjongeren* [Young and multi-media: Media use and information seeking among adolescents and young adults, especially among Muslim youth]. Amsterdam / Den Haag: Breckfield Hall Publishers / WODC, Ministry of Justice.

- Konijn, E. A. (2008). Affects and media exposure. In W. Donsbach (Ed.), *The International Encyclopedia of Communication* (Vol. 1, pp. 123-129). Oxford: Wiley-Blackwell.
- Konijn, E. A., & Hoorn, J. F. (2005). Some like it bad: Testing a model for perceiving and experiencing fictional characters. *Media Psychology*, 7(2), 107-144. doi: 10.1207/S1532785XMEP0702_1
- Konijn, E. A., van der Molen, J. H. W., & van Nes, S. (2009). Emotions bias perceptions of realism in audiovisual media: Why we may take fiction for real. *Discourse Processes*, 46(4), 309-340. doi: 10.1080/01638530902728546
- Korda, H., & Itani, Z. (2013). Harnessing social media for health promotion and behavior change. *Health Promotion Practice*, 14(1), 15-23. doi: doi:10.1177/1524839911405850
- Kozinets, R. V. (2002). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research*, 39(1), 61-72. doi: 10.1509/jmkr.39.1.61.18935
- Kozinets, R. V. (2010). *Netnography: Doing ethnographic research online*: Sage publications.
- Kozinets, R. V. (2015). *Netnography: Redefined*. Los Angeles, CA: Sage.
- Kozinets, R. V., Dolbec, P.-Y., & Earley, A. (2014). Netnographic analysis: Understanding culture through social media data. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 262-276). London, GB: SAGE Publications.
- Kreuter, M. W., Green, M. C., Cappella, J. N., Slater, M. D., Wise, M. E., Storey, D., . . . Woolley, S. (2007). Narrative communication in cancer prevention and control: A framework to guide research and application. *Annals of Behavioral Medicine*, 33(3), 221-235. doi: 10.1007/bf02879904
- Kreuter, M. W., & Wray, R. J. (2003). Tailored and targeted health communication: strategies for enhancing information relevance. *American journal of health behavior*, 27(1), S227-S232. doi: 10.5993/AJHB.27.1.s3.6
- Krueger, R. A., & Casey, M. A. (2015a). Focus group interviewing in *Handbook of Practical Program Evaluation*. In K. E. Newcomer, H. P. Hatry & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation* (4th ed., pp. 506-534): San Francisco, CA: John Wiley & Sons, Inc.
- Krueger, R. A., & Casey, M. A. (2015b). *Focus groups: A practical guide for applied research* [Adobe Digital Editions version] (5th ed). Retrieved from <https://play.google.com>
- Kruglanski, A. W., Gelfand, M. J., Bélanger, J. J., Sheveland, A., Hetiarachchi, M., & Gunaratna, R. (2014). The psychology of radicalization and deradicalization: How Significance quest impacts violent extremism. *Political Psychology*, 35, 69-93. doi: 10.1111/pops.12163
- Lee, E.J. (2006). When and how does depersonalization increase conformity to group norms in computer-mediated communication? *Communication Research*, 33(6), 423-447. doi: 10.1177/0093650206293248
- Lee, E.J. (2007). Deindividuation effects on group polarization in computer-mediated communication: The role of group identification, public-self-awareness, and perceived argument quality. *Journal of Communication*, 57(2), 385-403. doi: 10.1111/j.1460-2466.2007.00348.x
- Lentini, P. (2013). *Neojihadism: Towards a new understanding of terrorism and extremism?* Cheltenham, GB: Edward Elgar Publishing.

- Leuprecht, C., Hataley, T., Moskalenko, S., & McCauley, C. (2010). Containing the narrative: Strategy and tactics in countering the storyline of global Jihad. *Journal of Policing, Intelligence and Counter Terrorism*, 5(1), 42-57. doi: 10.1080/18335300.2010.9686940
- Levac, J. J., & O'Sullivan, T. (2010). Social media and its use in health promotion. *Revue interdisciplinaire des sciences de la santé-Interdisciplinary Journal of Health Sciences*, 1(1), 47-53. doi: 10.18192/riss-ijhs.v1i1.1534
- Lindekilde, L. (2012). Introduction: Assessing the effectiveness of counter-radicalisation policies in northwestern Europe. *Critical Studies on Terrorism*, 5(3), 335-344. doi: 10.1080/17539153.2012.723522
- Lovejoy, K., & Saxton, G. D. (2012). Information, community, and action: how nonprofit organizations use social media. *Journal of Computer-Mediated Communication*, 17(3), 337-353. doi: 10.1111/j.1083-6101.2012.01576.x
- MacKie, D. M., Gastardo-Conaco, M. C., & Skelly, J. J. (1992). Knowledge of the advocated position and the processing of in-group and out-group persuasive messages. *Personality and Social Psychology Bulletin*, 18(2), 145-151. doi: 10.1177/0146167292182005
- Mackie, D. M., Worth, L. T., & Asuncion, A. G. (1990). Processing of persuasive in-group messages. *Journal of personality and social psychology*, 58(5), 812. doi: 10.1037/0022-3514.58.5.812
- Macnair, L., & Frank, R. (2017). Voices Against Extremism: A case study of a community-based CVE counter-narrative campaign. *Journal for Deradicalization*, 10, 147-174. Retrieved from <http://journals.sfu.ca/jd/index.php/jd/article/view/86>
- Macnamara, J. (2017). *Evaluating public communication: Exploring new models, standards, and best practice*. London, GB / New York, NY: Routledge.
- Mahood, S., & Rane, H. (2017). Islamist narratives in ISIS recruitment propaganda. *The Journal of International Communication*, 23(1), 15-35. doi: 10.1080/13216597.2016.1263231
- Maley, W. (2015). The war in Afghanistan: Australia's strategic narratives. In B. De Graaf, G. Dimitriu & J. Ringsmose (Eds.), *Strategic narratives, public opinion and war: Winning domestic support for the Afghan War* [Adobe Digital Editions version]. Retrieved from <https://play.google.com>
- Mann, L. Doosje, B., Konijn, E. A., Nickolson, L., Moore, U., & Ruigrok, N. (2015). *Indicatoren en manifestaties van weerbaarheid van de Nederlandse bevolking tegen extremistische boodschappen: Een theoretische en methodologische verkenning* [Indicators and manifestations of resilience of the Dutch population against extremist messages. A theoretical and methodological exploration]. Retrieved from the Research and Documentation Centre, Dutch Ministry of Security & Justice website <https://www.wodc.nl/onderzoeksdatabase/2488-de-meetbaarheid-van-de-weerbaarheid.aspx>
- Marcellino, W., Smith, M., Paul, C., & Skrabala, L. (2017). *Monitoring social media: Lessons for future Department of Defense social media analysis in support of information operations* (RAND Publication no. RR-1742-OSD) [Research Report]. Retrieved from the RAND corporation website https://www.rand.org/pubs/research_reports/RR1742.html

- Marques, J., Abrams, D., Paez, D., & Martinez-Taboada, C. (1998). The role of categorization and in-group norms in judgments of groups and their members. *Journal of personality and social psychology*, 75(4), 976. doi: 10.1037/0022-3514.75.4.976
- Marques, J. M., Abrams, D., Páez, D., & Hogg, M. A. (2001). Social categorization, social identification, and rejection of deviant group members. In M. A. Hogg & R. S. Tindale (Eds.), *Blackwell handbook of social psychology: Group processes* (pp. 400–424). Oxford, GB: Blackwell Publishers Ltd. doi: 10.1002/9780470998458.ch17
- Marques, J. M., Yzerbyt, V. Y., & Leyens, J.-P. (1988). The “Black Sheep Effect”: Extremity of judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology*, 18(1), 1-16. doi: 10.1002/ejsp.2420180102
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior Research Methods*, 44(1), 1-23. doi: 10.3758/s13428-011-0124-6
- McGarty, C., Haslam, S. A., Hutchinson, K. J., & Turner, J. C. (1994). The effects of salient group memberships on persuasion. *Small Group Research*, 25(2), 267-293. doi: 10.1177/1046496494252007
- McGuire, W. J. (1964). Inducing resistance to persuasion: Some contemporary approaches. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 1, pp. 191-229). New York: Academic Press.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113. doi: <http://dx.doi.org/10.1016/j.asej.2014.04.011>
- Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, 59, 210-220. doi: <http://dx.doi.org/10.1016/j.pragma.2013.07.012>
- Metzger, M. J., & Flanagin, A. J. (2017). Digital media and perceptions of source credibility in political communication. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford Handbook of Political Communication*. New York, NY: Oxford University Press.
- Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic Approaches to credibility evaluation online. *Journal of Communication*, 60(3), 413-439. doi: 10.1111/j.1460-2466.2010.01488.x
- Miller, C., & Bartlett, J. (2012). 'Digital fluency': towards young people's critical use of the internet. *Journal of Information Literacy*, 6(2), 35-55. doi: 10.11645/6.2.1714
- Miller, C. H., Burgoon, M., Grandpre, J. R., & Alvaro, E. M. (2006). Identifying principal risk factors for the initiation of adolescent smoking behaviors: The significance of psychological reactance. *Health communication*, 19(3), 241-252. doi: 10.1207/s15327027hc1903_6
- Miller, C. H., Lane, L. T., Deatrck, L. M., Young, A. M., & Potts, K. A. (2007). Psychological reactance and promotional health messages: The effects of controlling language, lexical concreteness, and the restoration of freedom. *Human Communication Research*, 33(2), 219-240. doi: 10.1111/j.1468-2958.2007.00297.x

- Miron, A. M., & Brehm, J. W. (2006). Reactance Theory - 40 Years later. *Zeitschrift für Sozialpsychologie*, 37(1), 9-18. doi: 10.1024/0044-3514.37.1.9
- Miskimmon, A., O'Loughlin, B., & Roselle, L. (2012). Forging the world: Strategic narratives and international relations. London: Centre for European Politics/New Political Communications Unit.
- Miskimmon, A., O'Loughlin, B., & Roselle, L. (2015). Great power politics and strategic narratives of war. In B. De Graaf, G. Dimitriu & J. Ringsmose (Eds.), *Strategic narratives, public opinion and war: Winning domestic support for the Afghan War* [Adobe Digital Editions version]. Retrieved from <https://play.google.com>
- Moghaddam, F. M. (2005). The staircase to terrorism: A psychological exploration. *American Psychologist*, 60(2), 161. doi: 10.1037/0003-066X.60.2.161
- Moran, M. B., & Sussman, S. (2014). Translating the link between social identity and health behavior into effective health communication strategies: An experimental application using antismoking advertisements. *Health communication*, 29(10), 1057-1066. doi: 10.1080/10410236.2013.832830
- Moreno, M. A., Goniu, N., Moreno, P. S., & Diekema, D. (2013). Ethics of Social Media Research: Common Concerns and Practical Considerations. *Cyberpsychology, Behavior and Social Networking*, 16(9), 708-713. doi: 10.1089/cyber.2012.0334
- Morgan, S. E., King, A. J., Ivic, R., Thompson, T. L., Parrott, R., & Nussbaum, J. F. (2011). Using new technologies to enhance health communication research methodology. In T. L. Thompson, J. F. Nussbaum & J. F. Nussbaum (Eds.), *Handbook of health communication* (2nd ed.). Retrieved from <https://play.google.com>
- Moyer-Gusé, E. (2008). Toward a theory of entertainment persuasion: Explaining the persuasive effects of entertainment-education messages. *Communication Theory*, 18(3), 407-425. doi: 10.1111/j.1468-2885.2008.00328.x
- Moyer-Gusé, E., & Dale, K. (2017). Narrative persuasion theories. In R. Patrick (Ed.), *The International Encyclopedia of Media Effects* (Vol. 3, pp. 1345-1354). Chichester, GB, John Wiley & Sons, Inc. doi: 10.1002/9781118783764.wbieme0082
- Moyer-Gusé, E., Mahood, C., & Brookes, S. (2011). Entertainment-education in the context of humor: Effects on safer sex intentions and risk perceptions. *Health communication*, 26(8), 765-774. doi: <http://dx.doi.org/10.1080/10410236.2011.566832>
- Moyer-Gusé, E., & Nabi, R. L. (2010). Explaining the effects of narrative in an entertainment television Program: Overcoming resistance to persuasion. *Human Communication Research*, 36(1), 26-52. doi: 10.1111/j.1468-2958.2009.01367.x
- Nabi, R. L. (2015). Emotional flow in persuasive health messages. *Health communication*, 30(2), 114-124. doi: 10.1080/10410236.2014.974129
- Nabi, R. L. (2016). Laughing in the face of fear (of disease detection): Using humor to promote cancer self-examination behavior. *Health communication*, 31(7), 873-883. doi: 10.1080/10410236.2014.1000479

- Nabi, R. L., Moyer-Gusé, E., & Byrne, S. (2007). All joking aside: A serious investigation into the persuasive effect of funny social issue messages. *Communication Monographs*, 74(1), 29-54. doi: <http://dx.doi.org/10.1080/03637750701196896>
- Neiger, B. L., Thackeray, R., Burton, S. H., Giraud-Carrier, C. G., & Fagen, M. C. (2012). Evaluating Social Media's Capacity to Develop Engaged Audiences in Health Promotion Settings. *Health Promotion Practice*, 14(2), 157-162. doi: 10.1177/1524839912469378
- Nelson-Field, K., Riebe, E., & Newstead, K. (2013). The emotions that drive viral video. *Australasian Marketing Journal*, 21(4), 205-211. doi: <http://dx.doi.org/10.1016/j.ausmj.2013.07.003>
- Neo, L. S., Dillon, L., Shi, P., Tan, J., Wang, Y., & Gomes, D. (2016). Understanding the psychology of persuasive violent extremist online platforms. In M. Khader, L. S. Neo, G. Ong, E. Tan & J. Chin (Eds.), *Countering violent extremism and radicalisation in the digital era* [Adobe Digital Editions version] (pp. 1-15). doi: 10.4018/978-1-5225-0156-5.ch001
- Neumann, P. R. (2013a). Options and strategies for countering online radicalization in the United States. *Studies in Conflict & Terrorism*, 36(6), 431-459. doi: 10.1080/1057610X.2013.784568
- Neumann, P. R. (2013b). The trouble with radicalization. *International Affairs*, 89(4), 873-893. doi: 10.1111/1468-2346.12049
- Neumann, P. R. (2015, September). *Victims, perpetrators, assets: the narratives of Islamic State defectors* [ICSR Report]. Retrieved from the International Centre for the Study of Radicalization and Political Violence website: <http://www.ceskatelevize.cz/ct24/sites/default/files/1627584-icsr-report-victims-perpertrators-assets-the-narratives-of-islamic-state-defectors.pdf>
- Nissen, T. E. (2015). *#The weaponization of social media: @Characteristics of contemporary conflicts* [Monograph Report]. Retrieved from the Royal Danish Defense College website <http://www.fak.dk/publikationer/Pages/The-Weaponization-Of-Social-Media.aspx>
- Noar, S. M. (2006). A 10-year retrospective of research in health mass media campaigns: Where do we go from here? *Journal of health communication*, 11(1), 21-42. doi: 10.1080/10810730500461059
- Noar, S. M. (2011). An audience-channel-message-evaluation (ACME) framework for health communication campaigns. *Health Promotion Practice*, 13(4), 481-488. doi: 10.1177/1524839910386901
- Noar, S. M., Harrington, N. G., & Aldrich, R. S. (2009). The role of message tailoring in the development of persuasive health communication messages. *Annals of the International Communication Association*, 33(1), 73-133. doi: 10.1080/23808985.2009.11679085
- Noar, S. M., & Head, K. J. (2011). Trends in the research and practice of health communication campaigns. *Sociology Compass*, 5(6), 426-438. doi: 10.1111/j.1751-9020.2011.00379.x
- Noar, S. M., Palmgreen, P., Chabot, M., Dobransky, N., & Zimmerman, R. S. (2009). A 10-year systematic review of HIV/AIDS mass communication campaigns: Have we made progress? *Journal of health communication*, 14(1), 15-42. doi: 10.1080/10810730802592239.

- Noordegraaf, M., Douglas, S., Bos, A., & Klem, W. (2016). *Evaluation Counterterrorism Strategy 2011-2015*. Retrieved from the Research and Documentation Centre, Dutch Ministry of Security & Justice website <https://www.wodc.nl/onderzoeksdatabase/2608-evaluatie-nationale-contraterrorisme-strategie-2011-2015.aspx>
- O'Keefe, D. J. (1999). How to Handle Opposing Arguments in Persuasive Messages: A Meta-Analytic Review of the Effects of One-Sided and Two-Sided Messages. *Annals of the International Communication Association*, 22(1), 209-249. doi: 10.1080/23808985.1999.11678963
- O'Keefe, D. J. (2015). *Persuasion: theory and research* (3rd ed.). Los Angeles, CA: SAGE Publications.
- Oumnia works. (n.d.). *Wat is Oumnia Works?* [What is Oumnia Works? Retrieved from <http://www.oumniaworks.nl>
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135. doi: 10.1561/1500000011
- Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23(3), 184-188. doi: 10.1177/0963721414531598
- Paul, C., Yeats, J., Clarke, C. P., & Matthews, M. (2015). Assessing and evaluating Department of Defense efforts to inform, influence, and persuade: desk reference (RAND Publication No. RR-809/1-OSD) [Research Report]. Retrieved from the RAND corporation website: https://www.rand.org/content/dam/rand/pubs/research_reports/RR800/RR809z1/RAND_RR809z1.pdf
- 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*. Retrieved from: https://www.researchgate.net/publication/271511172_Explaining_and_understanding_the_role_of_exposure_to_new_social_media_on_violent_extremism_an_integrative_quantitative_and_qualitative_approach
- 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. doi: 10.1080/09546553.2013.876414
- Pawson, R. & Tilly, N. (2004). *Realist Evaluation*. Retrieved from http://www.communitymatters.com.au/RE_chapter.pdf
- Perloff, R. M. (2010). *The dynamics of persuasion communication and attitudes in the 21st century* (4th ed.). New York, NY: Routledge.
- Peters, G.J. Y., Ruiters, R. A. C., & Kok, G. (2013). Threatening communication: a critical re-analysis and a revised meta-analytic test of fear appeal theory. *Health Psychology Review*, 7(sup1), S8-S31. doi: 10.1080/17437199.2012.703527
- Petrosino, A., Turpin-Petrosino, C., & Buehler, J. (2003). Scared Straight and other juvenile awareness programs for preventing juvenile delinquency: A systematic review of the randomized

- experimental evidence. *The ANNALS of the American Academy of Political and Social Science*, 589(1), 41-62. doi: 10.1177/0002716203254693
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19, 123-205. doi: 10.1016/S0065-2601(08)60214-2
- Pfau, M., & Bockern, S. V. (1994). The persistence of inoculation in conferring resistance to smoking initiation among adolescents. *Human Communication Research*, 20(3), 413-430. doi: 10.1111/j.1468-2958.1994.tb00329.x
- Pfau, M., Bockern, S. V., & Kang, J. G. (1992). Use of inoculation to promote resistance to smoking initiation among adolescents. *Communication Monographs*, 59(3), 213-230. doi: 10.1080/03637759209376266
- Pfau, M., & Burgoon, M. (1988). Inoculation in political campaign communication. *Human Communication Research*, 15(1), 91-111. doi: 10.1111/j.1468-2958.1988.tb00172.x
- Pfau, M., Kenski, H. C., Nitz, M., & Sorenson, J. (1990). Efficacy of inoculation strategies in promoting resistance to political attack messages: Application to direct mail. *Communication Monographs*, 57(1), 25-43. doi: 10.1080/03637759009376183
- Postmes, T., Spears, R., Lee, A. T., & Novak, R. J. (2005). Individuality and social influence in groups: inductive and deductive routes to group identity. *Journal of personality and social psychology*, 89(5), 747. doi: 10.1037/0022-3514.89.5.747
- Postmes, T., Spears, R., Sakhel, K., & de Groot, D. (2001). Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin*, 27(10), 1243-1254. doi: doi:10.1177/01461672012710001
- Prevent Violent Extremism. (c2017). *Violent extremism and social media: A research portal*. Retrieved 15 rom <https://preventviolentextremism.info>.
- Price, M. E.; *Free expression, globalism, and the new strategic communication* [Adobe Digital Editions version], retrieved from <https://play.google.com>
- Keller, & Lehmann, D. R. (2008). Designing Effective Health Communications: A Meta-Analysis. *Journal of Public Policy & Marketing*, 27(2), 117-130. doi: 10.1509/jppm.27.2.117
- Pyszczynski, T., Abdollahi, A., Solomon, S., Greenberg, J., Cohen, F., & Weise, D. (2006). Mortality salience, martyrdom, and military might: The great Satan versus the axis of evil. *Personality and Social Psychology Bulletin*, 32(4), 525-537. doi: doi:10.1177/0146167205282157
- Quick, B. L., & Considine, J. R. (2008). Examining the use of forceful language when designing exercise persuasive messages for adults: A test of conceptualizing reactance arousal as a two-step process. *Health communication*, 23(5), 483-491. doi: 10.1080/10410230802342150
- Quick, B. L., & Stephenson, M. T. (2008). Examining the role of trait reactance and sensation seeking on perceived threat, state reactance, and reactance restoration. *Human Communication Research*, 34(3), 448-476. doi: 10.1111/j.1468-2958.2008.00328.x
- Rabasa, A., & Benard, C. (2015). *Eurojihad: Patterns of Islamist radicalization and terrorism in Europe*. New York: Cambridge University Press.

- Radicalisation Awareness Network (RAN). (2015, October). Counter narratives and alternative narratives [Ran Issue Paper]. Retrieved from https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/networks/radicalisation_awareness_network/ran-papers/docs/issue_paper_cn_oct2015_en.pdf
- Radicalisation Awareness Network (RAN). (2016, May). Voices of victims of terrorism [E-handbook]. Retrieved from https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/networks/radicalisation_awareness_network/about-ran/ran-rvt/docs/ran_vvt_handbook_may_2016_en.pdf
- Radicalisation Awareness Network (RAN). (2017). *Preventing radicalisation to terrorism and violent Extremism Approaches and Practices* [Collection of Best Practices]. Retrieved from https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/networks/radicalisation_awareness_network/ran-best-practices/docs/ran_collection_approaches_and_practices_en.pdf
- Rane, H. (2016). Narratives and counter-narratives of Islamist extremism. In A. Aly, S. MacDonald, L. Jarvis & T. Chen (Eds.), *Violent extremism online: new perspectives on terrorism and the internet* [Adobe Digital Editions version]. DOI: 10.4324/9781315692029
- Ranstorp, M. (2010). *Understanding violent radicalisation: terrorist and jihadist movements in Europe*: London, GB / New York, NY: Routledge.
- Reicher, S., Spears, R., & Haslam, A. (2010). The social identity approach in social psychology. In M. Wetherell & C. T. Mohanty (Eds.), *The SAGE Handbook of Identities* (pp. 45-62). London, GB: Sage
- Reyes, R. J., Van den Putte, B., Van Breukelen, R., Loef, J., Otte, M., & Wennekers, C. (2011, May). *Gedragverandering via campagnes* [Behaviour change through campaigns]. Retrieved from the website of the Public Information and Communications Service, Dutch Ministry of General Affairs <https://www.communicatierijk.nl/binaries/communicatierijk/documenten/publicaties/2011/05/18/gedragverandering-via-campagnes/literatuurstudie-gedragverandering-via-campagnes.pdf>
- Reynolds, L., & Scott, R. (2016). *Digital citizens: Countering extremism online* [Demos Report]. Retrieved from the Demos website <https://www.demos.co.uk/wp-content/uploads/2016/12/Digital-Citizenship-web-1.pdf>
- Reynolds, L., & Tuck, H. (2016, November). *The counter-narrative monitoring & evaluation handbook* [E-handbook]. Retrieved from the Institute for Strategic Dialogue website: <http://www.isdglobal.org/wp-content/uploads/2017/06/CN-Monitoring-and-Evaluation-Handbook.pdf>
- Richards, I. (2016). "Flexible" capital accumulation in Islamic State social media. *Critical Studies on Terrorism*, 9(2), 205-225. doi: 10.1080/17539153.2015.1125642
- Richardson, R. (2013, December). *Fighting fire with fire: target audience responses to online anti-violence campaigns* [ASPI Report]. Retrieved from the Australian Strategic Policy Institute website <https://www.aspi.org.au/report/fighting-fire-fire-target-audience-responses-online-anti-violence-campaigns>

- Riff, D., Lacy, S., & Fico, F. (2014). *Analyzing media messages: Using quantitative content analysis in research* (3rd ed.). New York, NY: Routledge.
- Rimal, R. N., & Adkins, A. D. (2003). Using computers to narrowcast health messages: The role of audience segmentation, targeting, and tailoring in health promotion. In T. L. Thompson, A. M., Dorsey, K. I. Miller & R. Parrott (Eds.), *Handbook of health communication* (pp. 497-513). Mahwah, NJ: Lawrence Erlbaum Associates.
- Roex, I. (2014). Should we be scared of all Salafists in Europe? *Perspectives on Terrorism*, 8(3). Retrieved from <http://www.terrorismanalysts.com/pt/index.php/pot/article/view/346/html>
- Rogers, R. W., & Mewborn, C. R. (1976). Fear appeals and attitude change: Effects of a threat's noxiousness, probability of occurrence, and the efficacy of coping responses. *Journal of personality and social psychology*, 34(1), 54–61. doi: 10.1037/0022-3514.34.1.54
- Romaniuk, P. (2015, September). *Does CVE work? Lessons learned from the global effort to counter violent extremism* [Research Report]. Retrieved from the Global Center on Cooperative Security website http://www.globalcenter.org/wp-content/uploads/2015/09/Does-CVE-Work_2015.pdf
- Romaniuk, P., & Fink, N. C. (2012). *From input to impact: Evaluating terrorism prevention programs* [Policy Report]. Retrieved from the Center on Cooperative Security website http://globalcenter.org/wp-content/uploads/2012/10/CGCC_EvaluatingTerrorismPrevention.pdf
- Rossmann, C. (2015). Strategic health communication: Theory- and evidence-based campaign development. In A. Z. Derina Holtzhausen (Ed.), *The Routledge Handbook of Strategic Communication, P, Editors: Derina Holtzhausen, Ansgar Zerfass, pp.409-423* (pp. 409-423). New York: Routledge.
- Rozendaal, E., Buijzen, M., & Valkenburg, P. (2011). Children's understanding of advertisers' persuasive tactics. *International Journal of Advertising*, 30(2), 329-350. doi: 10.2501/IJA-30-2-329-350. Doi: 10.2501/IJA-30-2-329-350
- Rozendaal, E., Oprea, S. J., & Buijzen, M. (2016). Development and Validation of a Survey Instrument to Measure Children's Advertising Literacy. *Media Psychology*, 19(1), 72-100. doi: 10.1080/15213269.2014.885843
- Ruiter, R. A. C., Kessels, L. T. E., Peters, G.-J. Y., & Kok, G. (2014). Sixty years of fear appeal research: Current state of the evidence. *International Journal of Psychology*, 49(2), 63-70. doi: 10.1002/ijop.12042
- Russell, J., & Rafiq, H. (2016). Countering Islamist extremist narratives: A strategic briefing. Retrieved from the Quilliam Foundation website <https://www.quilliaminternational.com/countering-islamist-extremist-narratives-a-strategic-briefing>
- Sageman, M. (2008). *Leaderless jihad: Terror networks in the twenty-first century*. Philadelphia, PA: University of Pennsylvania Press.
- Saltman, E. M., & Frenett, R. (2016). Female radicalization to ISIS and the role of women in CVE. In N. C. Fink, S. Zeiger & R. Bhulai (Eds.), *A man's world? Exploring the roles of women in countering terrorism and violent extremism* [E-book] (pp. 142-162). Retrieved from the the Global Center on Cooperative Security website: <http://www.globalcenter.org/wp->

content/uploads/2016/07/AMansWorld_FULL.pdf and the Hedayah Center website:
<http://www.hedayahcenter.org/Admin/Content/File-44201684919.pdf>

Sassenberg, K., Boos, M., & Rabung, S. (2005). Attitude change in face-to-face and computer-mediated communication: private self-awareness as mediator and moderator. *European Journal of Social Psychology, 35*(3), 361-374. doi: 10.1002/ejsp.254

Sassenberg, K., & Jonas, K. J. (2009). Attitude change and social influence on the net. In A. N. Joinson, K. Y. A. McKenna, T. Postmes & U.-D. Reips (Eds.), *Oxford Handbook of Internet Psychology*. New York, NY: Oxford University Press. doi: 10.1093/oxfordhb/9780199561803.013.0018

Schmid, A. P. (2012, August). *Strengthening the role of victims and incorporating victims in efforts to counter violent extremism and terrorism* [Research Paper]. Retrieved from the International Centre for Counter-Terrorism website <https://www.icct.nl/download/file/ICCT-Schmid-Strengthening-the-Role-of-Victims-August-2012.pdf>

Schmid, A. P. (2013, March). *Radicalisation, de-radicalisation, counter-radicalisation: A conceptual discussion and literature review* [Research Paper]. Retrieved from the International Centre for Counter-Terrorism website <https://www.icct.nl/download/file/ICCT-Schmid-Radicalisation-De-Radicalisation-Counter-Radicalisation-March-2013.pdf>

Schmid, A. P. (2014, January). *Al-Qaeda's "single narrative" and attempts to develop counter-narratives: The state of knowledge* [Research Paper]. Retrieved from the International Centre for Counter-Terrorism website <https://www.icct.nl/download/file/Schmid-Al-Qaeda's-Single-Narrative-and-Attempts-to-Develop-Counter-Narratives-January-2014.pdf>

Schmid, A. P. (2014, May). *Violent and non-violent extremism: Two sides of the same coin* [Research Paper]. Retrieved from the International Centre for Counter-Terrorism website <https://www.icct.nl/download/file/ICCT-Schmid-Violent-Non-Violent-Extremism-May-2014.pdf>

Schmid, A. P. (2015, June). *Challenging the narrative of the "Islamic State"* [Research Paper]. Retrieved from the International Centre for Counter-Terrorism website <https://www.icct.nl/wp-content/uploads/2015/06/ICCT-Schmid-Challenging-the-Narrative-of-the-Islamic-State-June2015.pdf>

Schmid, A. P., & Tinnes, J. (2015, December). *Foreign (terrorist) fighters with IS: A European perspective* [Research Paper]. Retrieved from the International Centre for Counter-Terrorism website <https://www.icct.nl/wp-content/uploads/2015/12/ICCT-Schmid-Foreign-Terrorist-Fighters-with-IS-A-European-Perspective-December2015.pdf>

Scrivens, R., Davies, G., & Frank, R. (2017). Searching for signs of extremism on the web: an introduction to Sentiment-based Identification of Radical Authors. *Behavioral Sciences of Terrorism and Political Aggression, 1*-21. doi: 10.1080/19434472.2016.1276612

Sedgwick, M. (2012). Jihadist ideology, Western counter-ideology, and the ABC model. *Critical Studies on Terrorism, 5*(3), 359-372. doi: 10.1080/17539153.2012.723520

Sherif, M., & Hovland, C. I. (1961). *Social judgment: assimilation and contrast effects in communication and attitude change*. New Haven, CT: Yale University Press.

- Sherif, C. W., Sherif, M., & Nebergall, R. E. (1965). *Attitude and attitude change: The social judgment-involvement approach*. Philadelphia: W. B. Saunders
- Shi, J., Poorisat, T., & Salmon, C. T. (2016). The use of social networking sites (SNSs) in health communication campaigns: review and recommendations. *Health communication*, 1-8. doi: 10.1080/10410236.2016.1242035
- Sieckelinck, S., & de Winter, M. (2015). Introduction. In S. Sieckelinck & M. de Winter (Eds.), *Formers and families: Transitional journeys in and out of extremism in the United Kingdom, Denmark and The Netherlands* (pp. 15-24), Retrieved from the National Coordinator Security and Counterterrorism (NCTV), Dutch Ministry of Security and Justice website https://english.nctv.nl/binaries/end-report-formers-and-families_tcm32-84248.pdf
- Silke, A. (1998). Cheshire-cat logic: The recurring theme of terrorist abnormality in psychological research. *Psychology, Crime & Law*, 4(1), 51-69. doi: 10.1080/10683169808401747
- Silverman, T., Stewart, C. J., Birdwell, J., & Amanullah, Z. (2016, August). *The Impact of counter-narratives. Insights from a year-long cross-platform pilot study of counter-narrative curation, targeting, evaluation and impact* [ISD Report]. Retrieved from the Institute for Strategic Dialogue website https://www.isdglobal.org/wp-content/uploads/2016/08/Impact-of-Counter-Narratives_ONLINE_1.pdf
- Slater, M. D. (1996). Theory and method in health audience segmentation. *Journal of health communication*, 1(3), 267-284. doi: 10.1080/108107396128059
- Slater, M. D., Kelly, K. J., & Thackeray, R. (2006). Segmentation on a shoestring: Health audience segmentation in limited-budget and local social marketing interventions. *Health Promotion Practice*, 7(2), 170-173. doi: 10.1177/1524839906286616
- Smith, A. N., Fischer, E., & Yongjian, C. (2012). How Does Brand-related User-generated Content Differ across YouTube, Facebook, and Twitter? *Journal of Interactive Marketing*, 26(2), 102-113. doi: 10.1016/j.intmar.2012.01.002
- Smith, J. R., & Hogg, M. A. (2008). Social identity and attitudes. In W. Crano & R. Prislin (Eds.), *Attitudes and attitude change* (pp. 337-360). New York: Psychology Press.
- Snyder, L. B. (2007). Health communication campaigns and their impact on behavior. *Journal of nutrition education and behavior*, 39(2), S32-S40. doi: 10.1016/j.jneb.2006.09.004
- Spears, R., Lea, M., & Lee, S. (1990). De-individuation and group polarization in computer-mediated communication. *British Journal of Social Psychology*, 29(2), 121-134. doi: 10.1111/j.2044-8309.1990.tb00893.x
- Spears, R., Lea, M., & Postmes, T. (2001). *Social psychological theories of computer-mediated communication: Social pain or social gain?* In W. P. Robinson and H. Giles (eds.) *The New Handbook of Language and Social Psychology* (2nd ed. pp. 601-623). Chichester: Wiley.
- Spears, R., Lea, M., Postmes, T., & Wolbert, A. (2011). A SIDE look at computer-mediated interaction. In Z. Birchmeier, B. Dietz-Uhler & G. Stasser (Eds.), *Strategic uses of social technology: An interactive perspective of social psychology* (pp. 16-39). Cambridge: Cambridge University Press.

- Statista (2017). *Twitter: Number of monthly active users 2010-2017*. Retrieved from <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users>
- Statista (2017). *Number of Facebook users worldwide 2008-2017*. Retrieved from <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide>
- Steindl, C., Jonas, E., Sittenthaler, S., Traut-Mattausch, E., & Greenberg, J. (2015). Understanding Psychological Reactance. *Zeitschrift für Psychologie*, 223(4), 205-214. doi: 10.1027/2151-2604/a000222
- Stevens, T. (2010). New Media and Counter-Narrative Strategies In E. J. A. M. Kessels (Ed.), *Countering Violent Extremist Narratives* (pp. 112-123). The Hague, NL: National Coordinator for Counterterrorism (NCTV).
- Stewart, D. W., & Shamdasani, P. N. (2014). *Focus Groups: Theory and Practice* (3rd ed.): SAGE Publications.
- Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems*, 29(4), 217-248. doi: 10.2753/MIS0742-1222290408
- Stryker, J. E., Wray, R. J., Hornik, R. C., & Yanovitzky, I. (2006). Validation of Database Search Terms for Content Analysis: The Case of Cancer News Coverage. *Journalism & Mass Communication Quarterly*, 83(2), 413-430. doi: 10.1177/107769900608300212
- Qatar International Academy for Security Studies (QIASS). (2013, September). *Countering violent extremism: the counter narrative study* [Research Report]. Retrieved from <http://qiass.org/wp-content/uploads/2016/05/CVE-Counter-Narrative-Study.pdf>
- Sue, V. M., & Ritter, L. A. (2012). *Conducting Online Surveys* (2nd ed.) Thousand Oaks, CA: Sage <http://dx.doi.org/10.4135/9781506335186>
- Sundar, S. S., & Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*, 51(1), 52-72. doi: 10.1111/j.1460-2466.2001.tb02872.x
- Szmania, S., & Fincher, P. (2017). Countering Violent Extremism Online and Offline. *Criminology & Public Policy*, 16(1), 119-125. doi: 10.1111/1745-9133.12267
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Monterey, CA: Brooks/Cole.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of inter group behaviour. In S. Worchel & L. W. Austin (Eds.), *Psychology of Intergroup Relations* (2nd ed., pp. 7-24). Chicago, IL: Nelson-Hall.
- Tanis, M., & Postmes, T. (2003). Social cues and impression formation in CMC. *Journal of Communication*, 53(4), 676-693. doi: 10.1111/j.1460-2466.2003.tb02917.x
- Taylor, H. (2012). *Social media for social change. Using the Internet to tackle Intolerance* (ISD Action Research Report). London: Institute for Strategic Dialogue.

- Thackeray, R., Neiger, B. L., Hanson, C. L., & McKenzie, J. F. (2008). Enhancing promotional strategies within social marketing programs: use of Web 2.0 social media. *Health Promotion Practice, 9*(4), 338-343. doi: 10.1177/1524839908325335
- Thackeray, R., Neiger, B. L., & Keller, H. (2012). Integrating social media and social marketing. *Health Promotion Practice, 13*(2), 165-168. doi: doi:10.1177/1524839911432009
- Tormala, Z. L. (2008). A new framework for resistance to persuasion: The resistance appraisals hypothesis. In R. Prislin & W. D. Crano (Eds.), *Attitudes and attitude change* (pp. 213-234). New York, NY: Psychology Press. .
- Torok, R. (2016). Social media and the use of discursive markers of online extremism and recruitment. In M. Khader, L. S. Neo, G. Ong, E. Tan & J. Chin (Eds.), *Countering Violent Extremism and Radicalisation in the Digital Era* [Adobe Digital Editions version] (pp. 39-69). doi: 10.4018/978-1-5225-0156-5.ch003
- Tuck, H., & Silverman, T. (2016, June). *The counter-narrative handbook* [E-handbook]. Retrieved from the Institute for Strategic Dialogue website <http://www.strategicdialogue.org/wp-content/uploads/2016/12/CN-Monitoring-and-Evaluation-Handbook.pdf>
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Cambridge, MA: Basil Blackwell.
- Valente, T. W. (2002). *Evaluating Health Promotion Programs*: Oxford University Press, USA.
- Valente, T. W., & Kwan, P. P. (2012). Evaluating communication campaigns. In R. E. Rice & C. K. Atkin (Eds.) *Public communication campaigns* [Adobe Digital Editions version] (4th ed.). Retrieved from <https://play.google.com>
- Valente, T. W., Palinkas, L. A., Czaja, S., Chu, K.-H., & Brown, C. H. (2015). Social network analysis for program implementation. *PloS one, 10*(6), 1-18. doi: 10.1371/journal.pone.0131712
- Valente, T. W., & Pitts, S. R. (2017). An appraisal of social network theory and analysis as applied to public health: Challenges and opportunities. *Annual Review of Public Health, 38*(1), 103-118. doi: 10.1146/annurev-publhealth-031816-044528
- Valkenburg, P. M., Peter, J., & Walther, J. B. (2016). Media effects: Theory and research. *Annual review of psychology, 67*, 315-338. doi: 10.1146/annurev-psych-122414-033608
- Van der Hulst, (2008). Sociale netwerkanalyse en de bestrijding van criminaliteit en terrorisme [Social network analysis for the combat of crime and terrorism]. *Justitiële verkenningen, 38*(8), 10-32. Retrieved from https://www.wodc.nl/binaries/jv0805-volledige-tekst_tcm28-77018.pdf
- van Atteveldt, W. H. (2008). *Semantic network analysis: Techniques for extracting, representing, and querying media content*. Charleston, SC: BookSurge Publishers. Retrieved from http://vanatteveldt.com/p/vanatteveldt_semanticnetworkanalysis.pdf
- van Ginkel, B. T. (2015, March). *Responding to cyber jihad: Towards an effective counter narrative* [Research Paper]. Retrieved from the International Centre for Counter-Terrorism website <https://www.icct.nl/download/file/ICCT-van-Ginkel-Responding-To-Cyber-Jihad-Towards-An-Effective-Counter-Narrative-March2015.pdf>

- Veilleux-Lepage, Y. (2014, December). *Retweeting the caliphate: The role of soft-sympathizers in the Islamic State's social media strategy*. Paper presented at the 6th International Symposium on Terrorism and Transnational Crime, Antalya, TUR. Retrieved from https://www.researchgate.net/publication/273896091_Retweeting_the_Caliphate_The_Role_of_Soft-Sympathizers_in_the_Islamic_State's_Social_Media_Strategy
- Veldhuis, J., Konijn, E. A., & Seidell, J. C. (2014). Counteracting Media's Thin-Body Ideal for Adolescent Girls: Informing Is More Effective Than Warning. *Media Psychology*, 17(2), 154-184. doi: 10.1080/15213269.2013.788327
- Veldhuis, T. (2012, March). *Designing rehabilitation and reintegration programmes for violent extremist offenders: A realist approach* [Research Paper]. Retrieved from the International Centre for Counter-Terrorism website <https://www.icct.nl/download/file/ICCT-Veldhuis-Designing-Rehabilitation-Reintegration-Programmes-March-2012.pdf>
- Veldhuis, T., & Staun, J. (2009). *Islamist radicalisation: a root cause model*. The Hague, NL: Netherlands Institute of International Relations Clingendael.
- Venhaus, J. (2010, May). *Why youth join al-Qaeda* (USIP Special Report No. 236). Retrieved from the United States Institute of Peace website <https://www.usip.org/sites/default/files/SR236Venhaus.pdf>.
- Vidino, L. (2010, November). *Countering radicalization in America* (USIP Special Report No. 262). Retrieved from the United States Institute of Peace website https://www.usip.org/sites/default/files/SR262%20-%20Countering_Radicalization_in_America.pdf
- Vitale, H. M., & Keagle, J. M. (2014). A Time to Tweet, as Well as a Time to Kill: ISIS's Projection of Power in Iraq and Syria. *Defense Horizons*, 77, 1-12. Retrieved from <https://www.files.ethz.ch/isn/185255/DH-77.pdf>
- Von Behr, I., Reding, A., Edwards, C., & Gribbon, L. (2013). *Radicalisation in the digital era: The use of the internet in 15 cases of terrorism and extremism* (RAND Publication No. RR-453-RE) [Research Report]. Retrieved from the RAND corporation website https://www.rand.org/content/dam/rand/pubs/research_reports/RR400/RR453/RAND_RR453.pdf
- VOX-Pol. (2017). *Objectives*. Retrieved from <http://www.voxpol.eu/about-us/>
- Wagemakers, J. (2009). A purist Jihadi-Salafi: The ideology of Abu Muhammad al-Maqdisi. *British Journal of Middle Eastern Studies*, 36(2), 281-297. doi: 10.1080/13530190903007327
- Waldman, S., & Verga, S. (2016, November). *Countering violent extremism on social media: An overview of recent literature and Government of Canada projects with guidance for practitioners, policy-makers, and researchers* (DRDC Scientific Report No. DRDC-RDDC-2016-R229). Retrieved from the website of Defense Research and Development Canada http://cradpdf.drdc-rddc.gc.ca/PDFS/unc262/p805091_A1b.pdf
- Wasserman, S., & Faust, K. 1994. *Social network analysis: Methods and applications*. Cambridge, GB: Cambridge University Press.

- Weimann, G. (2014). *New terrorism and new media* (Research series vol. 2) [Research Paper]. Retrieved from the Wilson Center's Common Labs website
<https://www.wilsoncenter.org/publication/new-terrorism-and-new-media>
- Weimann, G. (2015). *Terrorism in Cyberspace: The Next Generation* [Adobe Digital Editions Version], retrieved from <https://play.google.com>
- Weinreich, N. K. (2011). *Hands-on social marketing: a step-by-step guide to designing change for good* (2nd ed.). Thousand Oaks CA: Sage.
- Wiktorowicz, Q. (2006). Anatomy of the Salafi movement. *Studies in Conflict & Terrorism*, 29(3), 207-239. doi: 10.1080/10576100500497004
- Wright, K. B. (2005). Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services. *Journal of Computer-Mediated Communication*, 10(3), 00-00. doi: 10.1111/j.1083-6101.2005.tb00259.x
- Wyer, N. A. (2010). Selective self-categorization: Meaningful categorization and the in-group persuasion Effect. *The Journal of Social Psychology*, 150(5), 452-470. doi: 10.1080/00224540903365521
- YouTube (c2017). *YouTube for Press*. Retrieved from, <https://www.youtube.com/intl/en-GB/yt/about/press>
- Zeiger, S. (2016, August). *Undermining violent extremist narratives in South-East Asia: a how-to guide* [Compendium]: Retrieved from the Hedayah Center website
<http://www.hedayahcenter.org/Admin/Content/File-3182016115528.pdf>
- Zimmer, M. (2010). "But the data is already public": on the ethics of research in Facebook. *Ethics and Information Technology*, 12(4), 313-325. doi: 10.1007/s10676-010-9227-5

APPENDIX A: METHOD

This research relied predominantly on a literature review. Data from the literature was supplemented by subject-matter expert interviews and several focus groups interviews.

The literature study has been done using Google Scholar for scientific literature and Google for other “grey literature”. Due to the multi-faceted nature of the research questions, we have searched for scientific literature in various disciplines, including psychology, political science, communication science, health science, and computer science. In addition, we have made use of the “grey literature”, such as research reports, policy papers, working papers, evaluations and recommendations. Sometimes this involved documents of private companies involved in counter-narrative work, which we have treated with some caution, for obvious reasons. We have used search terms related to persuasion, narratives, counter narratives, online, social media, radicalization, terrorism, as well as effectiveness, test, approach and method. In addition, we have used the search backward method (i.e., checking the citations in an article) as well as the search forward method (i.e., checking online who has cited this article).

Based on engagement with our area of interest, we have conducted interviews (n=8) and 3 focus group interviews (n=6, n=7 and n=8) on the 12th and 14th of April 2016 with experts with roles in academia, industry, and government. We have used the following inclusion criteria when selecting the mix of people in our study:

- People who have a direct expertise in or an academic understanding of the use of counter-narratives.
- People with a thorough academic understanding of radicalization processes.
- People who work on practical matters related to radicalization, such as trainers or advisors.
- People with academic or practical understanding of strategic communication, persuasion and online campaigns.
- Policy makers and civil servant from municipalities and ministries.

We were also interested in the views of young people. Therefore one focus group (n=8) was conducted with young people (on the 12th of April 2016). Participants included young people studying social media, people involved in counter-narrative campaigning, so called-key-figures (i.e., young role models trained to develop activities to prevent radicalization), and a person with a personal history in radicalization. Focus groups were conducted by a moderator and an assistant moderator using a semi-structured interview scheme. Focus group interviews lasted for about 1,5 hour. Both the interviews and focus groups were audio taped and transcribed in full. After analysis key findings were integrated in the report.

APPENDIX B: NEDERLANDSE SAMENVATTING [DUTCH SUMMARY]

Radicale groepen komen tot bloei door hun extremistische boodschap te verspreiden. Ze dragen hun narratief dan ook actief uit met propaganda op een breed scala aan sociale media, onder meer op YouTube, Twitter en Facebook. In deze studie onderzoeken we in hoeverre sociale mediacampagnes en programma's gericht op het uitdragen en voeden van een contra-narratief kunnen worden gebruikt om gewelddadig extremisme te voorkomen dan wel personen te deradicaliseren. We richten ons daarbij op een specifieke vorm van gewelddadige radicalisering, namelijk het salafi-jihadisme of jihadistisch salafisme. Vanwege de politieke en maatschappelijke context waarin dit project werd gestart, waren NCTV en WODC geïnteresseerd in deze specifieke vorm van radicalisering. De onderzoeksmethodes die we hebben gebruikt zijn een uitgebreide literatuurstudie, interviews (n = 8) en drie focusgroepengesprekken (n = 6, n = 7 en n = 8) met mensen met verschillende expertise en achtergronden (waaronder academici, veldwerkers, media en design studenten, 'sleutelfiguren' en een voormalig radicaal individu).

In deze studie worden narratieven geconceptualiseerd als strategisch geconstrueerde verhaallijnen. Staats- en niet-staatsactoren projecteren en voeden deze verhaallijnen via (online) strategische communicatie om vorm te geven aan hoe het publiek bepaalde gebeurtenissen of problemen interpreteert en hun gedrag te beïnvloeden op een manier die bevorderlijk is voor hun doelen. In de context van het bestrijden van radicalisering, conceptualiseren wij counter-narratives of contra-narratieven als strategisch geconstrueerde verhaallijnen die door middel van strategische communicatie worden uitgedragen en gevoed om de aantrekkingskracht van de narratieven van gewelddadige extremistische groepen te ondermijnen. Hierbij kunnen drie vormen van strategische communicatie worden onderscheiden: (1) counter-messaging (bijvoorbeeld communicatieactiviteiten die extremistische boodschappen direct weerspreken) (2) alternative messaging (bijvoorbeeld communicatieactiviteiten die een positief alternatief voor de extremistische boodschap aanreiken); en (3) strategische communicatie door de overheid (bijvoorbeeld communicatieactiviteiten die inzicht geven in wat de overheid doet).

Een centrale uitkomst van onze analyse is dat naar mate mensen radicaler worden en zich sterk gaan identificeren met en committeren aan een extremistische ideologie of groep, het minder waarschijnlijk is dat zij vatbaar zullen zijn voor counter-narratieve beïnvloedingspogingen. In algemene zin zijn mensen met een sterke overtuiging lastig te overreden middels een (online) campagne. Zij zijn namelijk vaak weinig ontvankelijk voor boodschappen die hen op andere gedachten willen brengen. Maar het is echter heel moeilijk om mensen met sterk extremistische attitudes op andere gedachten te brengen. Het is maar de vraag of de boodschappen wel 'aankomen' en niet worden genegeerd. Wanneer deze hen wel bereiken, is het tevens mogelijk dat zij een ongewenst effect sorteren. Het kan er namelijk toe leiden dat zij juist in hun initiële overtuigingen worden gesterkt. Als zodanig achten wij het waarschijnlijk dat counter-narrative initiatieven meer kans hebben effect te sorteren wanneer zij worden ingezet als preventiemiddel, bijvoorbeeld bij mensen die 'gevoelig' blijken te zijn voor extremistische opvattingen en informatie – in zoverre zij kunnen worden geïdentificeerd – in een poging een verder radicaliseringsproces te voorkomen.

Daarnaast hebben we ons de vraag gesteld of dergelijke inspanningen effect kunnen hebben voor personen die een glimp van twijfel hebben waar het gaat om hun betrokkenheid in

een radicale groep. Potentieel kunnen contra-narratieve boodschappen op sociale media een zaad van twijfel planten. Maar dit is hoogst speculatief, en er is meer onderzoek nodig om deze notie te ondersteunen. Het startpunt voor het ontwerpen van een counter-narrative campagne is onderzoek. Er dient een grondig begrip ontwikkeld te worden van (de context van) 'het probleem', onderliggende determinanten, en de wijze waarop communicatie kan (en niet kan) bijdragen aan een oplossing. Hierbij is het verstandig om relevante theorieën en inzichten omtrent radicalisering en attitude- en gedragsverandering in overweging te nemen. Naast het meewegen van bestaande theorie, wordt aangeraden een degelijke programmatheorie te ontwikkelen die verklaart hoe campagneinspanningen tot het gewenste resultaat dienen te leiden.

Het is cruciaal dat ontwikkelaars van contra-narratieve campagnes hun doelpubliek zeer zorgvuldig afbakenen. Deze dient te worden gesegmenteerd op basis van relevante variabelen. Om dit te kunnen doen, is het essentieel de doelgroepen grondig te bestuderen en begrijpen, hetgeen een zorgvuldige en grondige publieksanalyse vereist. Contra-narratieve campagnes die zich richten op een te breed en algemeen doelpubliek lopen het risico ineffectief te zijn en kunnen mogelijk averechts uitpakken. Het valt dan ook aan te raden inspanningen te concentreren op kleinere, specifieke segmenten.

Naast het bestuderen van het publiek is het belangrijk om een diepgaand begrip van de verschillende elementen van extremistische narratieven te ontwikkelen, zodat alternatieve of tegengeluiden hierop kunnen worden toegespitst. Ook dient in ogenschouw te worden genomen op welke wijze deze narratieven appelleren aan en weerklank vinden bij leden van doelgroep. Individuen die radicaliseren en op zoek gaan naar extremistische informatie kunnen verschillende motieven of drijfveren hebben en die kunnen toegespitste alternatieve of tegengeluiden vereisen. Sommigen voelen zich aangetrokken tot een ideologie. Maar er zijn ook andere factoren die mensen kunnen motiveren, zoals de behoefte aan groepslidmaatschap en kameraadschap, de behoefte aan zingeving en de behoefte aan spanning en sensatie.

Zoals eerder aangeven, is het van belang zicht te hebben op het gewenste resultaat en de stappen die hier naartoe zouden moeten leiden. Hieraan gerelateerd dienen doelen en doelstellingen te worden afgebakend. Welke verandering men wil men eigenlijk bereiken? Goed gedefinieerde doelstellingen zijn specifiek, meetbaar, haalbaar, relevant en tijdsgebonden. Realistische doelen stellen vereist ook dat er inzicht is in de beschikbare middelen zoals tijd, financiën, mensen. Het is ook belangrijk om potentiële partners in ogenschouw te nemen. Het betrekken van relevante actoren binnen lokale gemeenschappen bij campagneontwikkeling en implementatie kan campagne-inspanning ten goede komen.

Hoewel controle over de boodschap geenszins verzekerd is wanneer deze wordt 'losgelaten' in een sociale media omgeving, is het noodzakelijk om goed na te denken over de inhoud de boodschap. Zoals gesteld, in zoverre dit mogelijk is, loont het hierbij wellicht in te spelen op verschillende motieven van radicalen. Afgezien van de inhoud is noodzakelijk om na te denken over andere eigenschappen van boodschappen. Boodschappen met een narratieve structuur (bijvoorbeeld een verhaalvorm) kunnen mogelijk krachtig zijn, onder meer omdat zij de persuasieve strekking subtiel brengen en kunnen bijdragen aan een reductie van verschillende vormen van weerstand. Daarnaast zouden tweezijdige boodschappen (waarbij één zijde wordt

weerlegd), en boodschappen met een sterke emotionele aantrekkingskracht mogelijkwijs overtuigend kunnen zijn.

Voor een doeltreffende boodschap is het tevens van belang dat het doelpubliek de boodschapper als geloofwaardig waarneemt. Kandidaten die vaak worden genoemd zijn voormalige extremisten, slachtoffers van geweld, peers en familie, en belangrijke leden uit de gemeenschappen en spelers uit het middenveld. Het potentieel van de overheid om op te treden als geloofwaardige boodschapper is mogelijk beperkt vanwege de kloof met het publiek en een gepercipieerde 'say-do gap'. Echter, het antwoord op de vraag welke van deze spelers de boodschap het meest doeltreffend kan brengen, hangt onder meer af van de context, de doelgroep die men wil bereiken en van de boodschap die wordt doorgegeven. Daarbij dient gemeld te worden dat in een sociale media omgeving de bron van een boodschap of post ambigu of onduidelijk kan zijn. Processen van informatieoverdracht kunnen de originele bron bijvoorbeeld verwateren en maken het soms moeilijk om te interpreteren wie de bron is. Ook geven sommige onderzoekers aan dat traditionele aannames waar het gaat om de indicatoren die een rol spelen bij het beoordelen van geloofwaardigheid binnen een sociale mediaomgeving worden gecompliceerd.

Er dient tevens nagedacht te worden over het kanaal dat wordt gebruikt om de counternarratieve boodschap te verspreiden. Logischerwijs dienen deze gebruikt te worden door de doelgroep. Daarbij zijn sociale media campagnes die meerdere kanalen inzetten mogelijk efficiënter dan campagnes met slechts één kanaal. Het is belangrijk om erop te wijzen dat sociale media niet altijd de meest efficiënte kanalen zijn om een doelpubliek te bereiken en dat we aanraden om online en offline kanalen te combineren. Waar het gaat om het verspreiden van het bericht, zou men mee kunnen doen in de 'conversaties' die reeds gaande zijn. In sommige gevallen kan het voordelig zijn om samen te werken met mensen en organisaties, die al een sterke aanwezigheid van sociale media hebben gevestigd en populair zijn onder de doelgroep. Soms kan het ook nuttig zijn om online (nieuws) outlets te benaderen om de boodschap te delen. Bij het verspreiden van een boodschap op sociale media kanalen dient daarbij ook rekening te worden gehouden met timing en volume. Ten slotte, kan het creëren van engagement met het publiek bijdragen aan het verspreiden van de boodschap.

Het is belangrijk om de potentiële risico's, uitdagingen en beperkingen mee te wegen bij het ontwikkelen en implementeren van een communicatiestrategie. Dit omvat onder meer het verliezen van controle over de boodschap, tegencampagnes, eventuele bedreigingen van de veiligheid en het welzijn van de boodschappers, en moeilijke beslissingsprocessen. Een andere grote uitdaging is het meten van het effect van een campagne.

Het meten van het effect van beleid en maatregelen gericht op het tegengaan van radicalisering is in algemene zin lastig. Dit is niet anders bij counter-narratieve campagnes. Om de effectiviteit van een campagne vast te stellen, is het van belang om een onderscheid te maken tussen een formatieve evaluatie, een proces evaluatie en een summatieve evaluatie. We beargumenteren dat bij het meten van effect van online counter-narrative inspanningen een combinatie van relevante awareness metrics (zoals bereik en aantal views) en engagement metrics (zoals likes, shares, comments en emoji responses) enig inzicht kan bieden in de vraag of de campagne de gewenste online impact heeft gehad in relatie tot gestelde doelen. Vaststellen of de juiste doelgroep ook wordt bereikt en of een campagne ook een blijvend, offline effect heeft op

diegene die bereikt zijn is echter uiterst ingewikkeld. Aanvullende traditionele onderzoeksmethoden, bijvoorbeeld (online) interviews en focusgroepen kunnen inzicht bieden in de ontvangst van online materialen (en dus van nut in de formatieve evaluatiefase), maar ze kunnen ook input bieden bij de interpretatie van kwantitatieve data verkregen in de summatieve evaluatiefase. Experimenten (offline of online) bieden inzicht in de oorzaak-gevolg-relatie, terwijl “netnography” gebruik maakt van data verkregen door observaties van mensen in hun reguliere sociale media omgeving. Sentiment- en inhoudsanalyses zijn methoden om te achterhalen hoe een counter-narratieve campagne overall wordt waargenomen, maar ook om uit te zoeken hoe de doelgroep de specifieke narratieven uit een campagne waarneemt. Ten slotte, sociale netwerk analyses kunnen inzicht bieden in de structuur van een groep en hoe deze structuur mogelijk kan veranderen over tijd, hoewel het gebruik van deze techniek in dit veld nog niet volledig tot groei is gekomen. Een combinatie van dergelijke kwalitatieve en kwantitatieve methoden zou bij kunnen dragen aan een dieper inzicht in het mogelijke effect van een campagne.

Zoals reeds naar voren kwam, wordt veelal verondersteld dat overheid niet zo geschikt is om als counter-narrative boodschapper op te treden. Zij mist namelijk de nodige geloofwaardigheid bij de relevante doelgroepen. Maar wat kan de overheid dan wel doen? Ten eerste kan de overheid haar eigen strategische communicatie stroomlijnen en helder en consistent uitleg en toelichting geven op haar beleid en acties, lokaal en binnen de internationale context. Daarbij is het zaak dat woorden en daden in overeenstemming zijn. Ten tweede kan de overheid een waardevolle rol vervullen door actoren uit gemeenschappen en middenveldorganisaties te faciliteren. Dit zouden zij kunnen doen door partnerschappen te stimuleren tussen het middenveld en de privésector (bijvoorbeeld in de technologiesector of de reclamewereld), door expertise te delen, en financiële ondersteuning te bieden. Overheden moeten zich er wel van bewust zijn dat een publieke associatie met initiatieven kan fungeren als een doodsteek voor deze inspanningen. Ten derde, overheden zouden systematische evaluaties van initiatieven kunnen stimuleren. Het bewijs van de doeltreffendheid van deze initiatieven is tot op heden spaarzaam. Tot slot, de overheid zou zich ook kunnen focussen op de ontwikkeling en ondersteuning van programma's op het gebied van digitale geletterdheid en kritische consumptievaardigheden.

Wat de beperkingen van de huidige studie betreft, we hebben onze analyse gericht op counter-narratives voor het ontkrachten van de jihadistische boodschap en ideologie. Als zodanig is het niet mogelijk deze bevindingen te generaliseren naar andere groepen. Bovendien, doordat er nauwelijks gedegen empirisch onderzoek beschikbaar is, moesten we een deel van onze analyse baseren op de algemene literatuur over beïnvloeding en communicatiecampagnes. Ook hebben wij soms gebruik moeten maken van de grijze literatuur. Dit beperkt de mogelijkheden om harde conclusies te trekken op basis van de huidige analyse.

We concluderen dat, hoewel een online counter-narrative campagne intuïtief een aantrekkelijke strategie lijkt, onze analyse laat zien dat dit naar alle waarschijnlijkheid niet heel effectief zal zijn als de-radicaliseringsinstrument. Counter-narrative activiteiten op sociale media hebben de meeste kans om doeltreffend te zijn wanneer mensen zich in een vroege fase van het radicaliseringsproces bevinden, wanneer de geesten van mensen nog open genoeg staan om de gepresenteerde informatie te registreren en verwerken. Het is wellicht ook mogelijk om opties te verkennen om bekende individuen (bijvoorbeeld een veroordeelde gevangene) aan specifieke

counter-narratives bloot te stellen, maar idealiter niet in een online context, maar in een offline context, om zo de reacties van de individuen te kunnen vaststellen.

