

Combating extremism in a public sphere at risk: platforms' affordances, dilemmas and opportunities of social media campaigns

Original

Combating extremism in a public sphere at risk: platforms' affordances, dilemmas and opportunities of social media campaigns / Monaci, S.. - In: H-ERMES. - ISSN 2284-0753. - ELETTRONICO. - 15:(2019), pp. 229-248.

Availability:

This version is available at: 11583/2774416 since: 2019-12-18T12:46:29Z

Publisher:

ESE Editoria Scientifica Elettronica

Published

DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

Combating extremism in a public sphere at risk: platforms' affordances, dilemmas and opportunities of social media campaigns

Sara Monaci

The spread of social media highlights controversial changes in the public sphere: new opportunities of access and expression in fact go alongside aberrant phenomena of extremist propaganda. In recent years, civil society institutions and bodies have responded to the wave of hatred and violence on social media through online awareness-raising campaigns aimed at combating ideological propaganda and offering alternative narratives to vulnerable individuals. The essay develops a critical reflection on some of these initiatives in view of social media affordances and the communication strategies adopted by promoters.

What is the role of social media campaigns in consideration of the public sphere at risk? What are the limits and criticalities for issuers? What are the potential effects on audiences?

Starting with a number of case histories, the contribution highlights that issuers often find themselves facing the conservative dilemma between the possibility of using social media strategically and the risk of legitimising extremist organisations that use the same media channels. The study also reveals that awareness-raising messages, developed in the form of a counter-narrative or alternative narrative, can trigger counterproductive actions such as the backfire effect which reinforce, rather than mitigate, the polarisation between individuals, particularly online, and which thereby invalidate the arguments and critical intentions of the campaigns.

Keywords: public sphere, social media, extremist propaganda.

1. The public sphere in the era of social media

According to Jurgen Habermas, the dissemination of mass media marks a further weakening of the bourgeois public sphere which transforms informed citizens into consumers of cultural products, passive subjects with respect to the manipulative strategies of the mass media. The ideal of a bourgeois public sphere in which private citizens - able to understand and grasp the problems of the community - negotiated power with the State is replaced by a more mainstream and conflict-free public sphere unified by mass consumption. Consensus becomes the political objective pursued through rampant consumerist culture; media propaganda is its main tool.

Habermas also encompasses in his reflections, albeit marginally, the nascent reality of electronic media, despite not focusing specifically on investigating their qualities (Habermas 1962).

The issue of the relationship between social media and the public sphere lends itself to different interpretations, partly relating to the broader relationship between public sphere and network society, partly focused on social media as

original elements of a new public sphere. The reflections of Manuel Castells and of Yochai Benkler refer to the first set. Castells emphasises that: - “The construction of the new public sphere in the network society proceeds by building protocols of communication between different communication processes”- (Castells, 2013, p. 125). Yochai Benkler writes: - “The easy possibility of communicating effectively into the public sphere allows individuals to reorient themselves from passive readers and listeners to potential speakers and participants in a conversation” (Benkler 2006, p. 213) (..) “The network allows all citizens to change their relationship to the public sphere. They no longer need to be consumers and passive spectators. They can become creators and primary subjects. It is in this sense that the Internet democratizes (ivi, p. 272)”. Unlike the mass media, the Internet encourages citizens to become actively involved in conversations and in producing content rather than being passive spectators; according to the most optimistic interpretations, the Internet thus represents an opportunity to democratise access to the public sphere, giving citizens the chance to form an active part of it. The enormous dissemination of social media has prompted further reflection on their potential to be new propulsive elements of expansion of public debate, participation and widespread creativity: Zizi Papacharissi describes the emergence of a: - “virtual sphere 2.0 in which citizen-consumers participate and express dissent with a public agenda [...] by expressing political opinion on blogs, viewing or posting content on YouTube, or posting a comment in an online discussion group” (Papacharissi 2009, p. 244). Jean Burgess and Joshua Green – authors of the first essay on YouTube as a cultural system - state that the world's most important video-blogging platform represents a: “cultural public sphere because it is an enabler of encounters with cultural differences and the development of political ‘listening’ across belief systems and identities” (2009, p. 77). The mature phase of Facebook, YouTube, Twitter - now widespread globally – is aligned, however, with a more critical interpretation of social media, considered as profit-oriented technical-economic apparatuses and not just as platforms for public conversation. Christian Fuchs emphasises (2014) the contradictions inherent in social media considered as public spheres; the most important is constituted by the relationship of power between those who own the

platforms – private profit-oriented corporations – and the end users. That tension between free use of social spaces and the profit demands of technological players imposes enormous distortions of the concept of *public*: the *publicity* of individuals on social networks goes hand in hand with the systematic monitoring of user profiles, use of personal data for publicity purposes and manipulative propaganda aimed at influencing voting decisions. The *publicity* of the data of users registered on social networks thus immediately reveals their contradictory aspect: the possibility of being on social media and participating in a public sphere - virtual and potentially global – takes for granted the provision of data to private entities which transform them into *commodities* that can be sold for marketing or targeted propaganda purposes. Recently, the case of Cambridge Analytica¹ has revealed how the economic use of data can represent a danger to democratic life by manipulating the consensus of millions of unsuspecting voters. A further critical argument is linked to the affordances (Boyd 2010) of social media platforms: what these allow or do not allow to be done, in view of how these technologies have been designed and implemented. According to authors such as Eli Parisier and Evgeny Morozov (2011) social media technologies offer spaces of discussion and freedom, but at the same time they can be used by authoritarian regimes to strengthen control and monitoring of dissident voices and to filter – opaquely – information for end users. This is the filter bubble phenomenon which is produced as a result of extreme personalisation performed by social media platforms, by search engines such as Google, to offer information so adherent to the tastes and opinions of the persons as to force the latter into an informative bubble which excludes any content or opinion that differs from the algorithmically defined user profile (Parisier 2012). The problem is that this process does not result from a deliberate choice of the individual but is the result of selection processes implemented by the platforms based upon micro-marketing and targeted propaganda objectives. By selecting information based upon the propensities and affinities of individuals, the filters strengthen a phenomenon already studied in social psychology, namely *confirmation bias*: individuals seek

¹ For an in-depth report on the case see: <https://www.theguardian.com/news/series/cambridge-analytica-files>, 20 November 2019.

online information and opinions that tend to strength rather than deny their visions of the world (Sunstein 2017, pp. 93-97).

If the filters support the inclinations of individuals who are having difficulty orienting themselves within a growing information overload, what could be the repercussions on the public sphere? Can we envisage the constitution of information silos that are increasingly impermeable to each other, isolating individuals within their reassuring bubbles?

Another crucial factor is the role of propaganda: *spreadability* (Jenkins, Ford, Green 2013), namely that the diffusive nature of social media enhances both the free expression of millions of citizens - finally emancipated from the constraints and restrictions of broadcast media - but they are also an opportunity for extremist organisations to spread ideologically misleading messages, connoted by verbal violence or aimed at ideological co-opting.

2. Extremist propaganda and the public sphere at risk

In a complex scenario where social media technologies are moving further and further away from the ideal image of “technologies of freedom” (de Sola Pool 1983), some recent phenomena have strengthened a vision of the public sphere 2.0 as a *public sphere at risk*.

Recent years have seen the dissemination of different forms of extremist propaganda: from the xenophobia of the extreme right which raged on platforms such as Stormfront, to the jihadist propaganda of ISIS (Islamic State of Iraq and Syria) which strategically used social media to recruit new followers in the West (Meleagrou-Hitchens, Kaderbhai 2017). Consider also the fact that propaganda was in many cases the reverberation of numerous terrorist attacks undertaken by ISIS in Europe between 2014 and 2016 and that still today - the Easter 2019 massacre in Sri Lanka is an example of this - represent a creeping threat. Even the phenomenon of *foreign fighters* – thousands of young fighters who left Europe between 2014 and 2016 to join the Islamic State army in Syria or in Iraq - can be attributed to some extent to the effects of an increasingly pervasive and seductive propaganda (Boutin et al. 2016; Reed, Pohl 2017).

The videos of ISIS executions posted on the social networks and often republished on the broad scale by mainstream news represented a macabre viral phenomenon that significantly contributed to creating a climate of widespread terror. The messages of hatred and violence amplified on social media made clear the limits, risks and enormous influence of the media in relation to the fears and hopes of voters. Online propaganda, on one side, and, on the other, the perception – fuelled by attacks in Europe - of an incumbent threat by Islam - have contributed to fuelling the vision of a *public sphere at risk* (Atwan 2015) in addition to the perception of the objective risk of terrorist attacks - I also refer to the climate of increasing suspicion towards Islamic organisations, foreign citizens in Europe and the enormous reverberation that the phenomenon of terrorism has had on the issue of migrants, seen as possible “infiltrators” or “silent cells” likely to attack Europe at the first opportunity. It is not the task of this essay to investigate the relationship between the phenomenon of terrorism and that of the perception of migrants in Europe but that reference is, however, fundamental for outlining the climate of the public sphere in the years between 2014 - 2017 as a result of different and concomitant phenomena whose relationships are still to be analysed. The concept of *public sphere at risk* thus refers to the risks linked to social media propaganda as an instrument used by extremist organisations for recruitment and ideological co-opting, but, on the other hand, also as a political lever aimed at exacerbating attitudes of fear and closure towards groups and social minorities.

Extremist propaganda – or tout court propaganda – thus acts in the sense of polarising the positions between those, for example, who wish to combat terrorism through integration and social inclusion measures referring to the disadvantaged groups of the Parisian banlieu – hot beds of jihadist extremism (Kepel 2015; Khosrokhavar 2017) – and those, on the other hand, who advocate the tightening of security measures and the development of a police state. The risk of not reaching an agreement between these divergent positions is higher the more aggressive and pervasive the propaganda from both sides becomes, and obviously the more the objective acts of violence - on one side and the other - reinforce that divide.

3. Awareness-raising initiatives: online campaigns against violent extremism

That scenario has led different institutions to undertake measures of online containment of extremist propaganda and to define counter-measures aimed at strengthening social awareness on the issue of violent radicalisation. Those measures are expressed at different levels: at international level through forms of collaboration with the main technological companies - Google, Facebook, Twitter etc. - to adopt measures and strategies of censure and containment of online extremist propaganda (Gillespie 2018), at European level through the creation of collaboration networks focused on monitoring, analysis and development of communication and awareness-raising initiatives of citizens on the risks linked to extremist propaganda (Briggs, Silverman T. 2016), and at national level - such as in France and in the United States - with the definition of targeted initiatives (e.g. the creation of free-phone numbers, or online listening points) to offer citizens an initial contact with the institutions in the case of emergency. The communication initiatives also include the promotion of social media campaigns aimed at developing counter-narratives able to combat those of extremist organisations and to enhance positive values such as social and cultural integration, freedom of expression and above all dialogue between citizens². The initiatives reviewed in the literature include online campaigns such as *Abdullah X, Average Mohamed, Extreme Dialogue* (Colliver, Davey 2017, pp. 175-182; RAN 2015) and *Breaking the ISIS Brand Counter Narratives Project* aimed at Albanian-speaking Facebook users (Speckard et al. 2018).

The *Average Mohamed*³ campaign is split into short animated cartoons aimed above all at young people, dedicated to the issue of identity, free expression, and the role of women in Islamic culture. The campaign was created

² See, in particular, the campaigns created by the RAN (Radicalisation Awareness Network) and by the Institute of Strategic Dialogue (ISD). The RAN involves scholars and operators working at European level to prevent radicalisation. The network is promoted and funded by the European Commission - Directorate-General "Migration and Home Affairs", in https://ec.europa.eu/home-affairs/what-we-do/networks/radicalisation_awareness_network_en, 1 July 2019. ISD is a non-governmental organisation dedicated to raising the awareness of the new generations against hatred and extremism. <https://www.isdglobal.org/about-isd/>, 1 July 2019.

³ <https://www.averagemohamed.com/>, 20 November 2019.

by a lay association of Muslims of Somali origin currently residing in Minnesota. Reflecting on these general issues, the campaign develops a message of moderation and conciliation in an attempt to combat the propaganda of ISIS violence and to draw attention to the pacifist culture of Islam.

*Abdullah X*⁴ is the protagonist of an animated graphic novel: the star is a young Muslim in search of a role in the Western society where he was born and grew up. Through the YouTube channel dedicated to the campaign, the young man expresses his opinions on important issues that affect the lives of many young Muslims in Europe in an historical phase lacerated by terrorism and by conflicts involving the Middle East. Abdullah X is a fictitious character who can give a voice to all young people wanting to reflect critically on issues of identity, faith, sense of belonging, duty, and also on the awareness of injustice and social inequality.

The *Extreme dialogue*⁵ campaign has an educational objective and offers many informative resources and documentation for teachers, operators and scholars on the subject of preventing violent extremism. The communicative message is split into a series of video testimonies of former extremist militants belonging to different organisations of jihadist nature or of the extreme right. The video contributions include, in particular, the testimony of Christianne Boudreau, mother of a young Canadian foreign fighter who joined the ISIS army and later died in Syria. The story of Christianne and her son Demian became a major case and led to the opportunity to establish an association – Mothers for Life – which brings together different women who have lost sons in Syria. Christianne Boudreau herself met the former President of the Chamber of Deputies in Italy, Laura Boldrini, in the context of a Parliament initiative on the issue of preventing jihadist radicalisation. Alongside the described initiatives, other campaigns – between 2015 and 2016 – have emerged as part of social media as interesting counter-narrative cases. The viral video *Not in my name*⁶, for example, is the initiative of an association of young British Muslims wanting to denounce their distance and opposition to terrorist violence, reaffirming the principle that Islam is

⁴ <https://www.youtube.com/user/abdullahx>, 20 November 2019.

⁵ <http://extremedialogue.org/>, 20 November 2019.

⁶ <https://www.youtube.com/watch?v=wfYanI-zJes>, 20 November 2019.

not ISIS. In response to this and almost reinforcing the message of the young Brits, the Islamic community of Bra (Cuneo, Italy), created in 2016 a short video teaser - *In my name*⁷ - which emphasises the sense of belonging and friendship that binds the community to the territory and its social context, thereby countering ISIS's message of hatred and violence with a real example of integration and hope for the future. In this case, the young Muslims want to say “who they are” and not just “what they are not”, challenging the prejudices and stereotypes that contribute to isolating Islamic communities in Europe.

The government-led campaigns include the French *Stop djihadisme*⁸: an interesting example that integrates information and documentation resources on the phenomenon of terrorism with interactive storytelling contents aimed at raising awareness of young people and families on the consequences of violent radicalisation: these tell, for example, of the destiny of French fighters who join ISIS, the ordeals of young women attracted to Syria by the promise of a fighter husband and then abandoned together with their children in a war-torn territory, or they emphasise the criminal responsibilities that these decisions imply.⁹

Another very interesting initiative - which could be considered a generator of online campaigns - is the Redirect¹⁰ project launched by the Jigsaw incubator of the Alphabet/Google group. This is a system able to re-direct the paths of online searches by aspiring jihadists towards video contents that offer counter-narratives to the messages of pro-ISIS propaganda¹¹. The project is based on the awareness that in many cases the phenomenon of indoctrination and recruitment is a process that starts from the bottom up and that sees young people very active in seeking information, texts and videos on terrorist organisations such as ISIS. The aim of the initiative is thus to divert attention from the outset from certain types of contents in order to emphasise, on the other hand, messages able to question and discredit extremist propaganda.

⁷ <https://www.youtube.com/watch?v=r24MKjdPMqg>, 20 November 2019.

⁸ <http://www.stop-djihadisme.gouv.fr/>, 20 November 2019.

⁹ See, in particular, the interactive storytelling created by the project <http://www.toujourslechoix.fr/> as part of the government campaign *Stop Djihadisme*.

¹⁰ <https://redirectmethod.org/>, 20 November 2019.

¹¹ https://www.wired.com/2016/09/googles-clever-plan-stop-aspiring-isis-recruits/?utm_content=buffer04d12&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer, 20 November 2019.

3.1 The social media campaign theheartofdarkness.eu (#hod)

As part of the European SAFFRON (Semantic Analysis against Foreign Fighters Recruitment Online Networks) project¹², the writer is the author and coordinator of a social media campaign to combat the narratives used by ISIS to recruit young European fighters. The aim of the campaign #hod¹³ is to offer alternative narratives that attempt a critical understanding of the motivations that lead young people to join the ISIS army, seeking to dismantle ISIS rhetoric while strengthening positive values, such as cultural integration, friendship and opportunities of self-expression protected by democratic societies. The initiative is an upstream campaign: it is not designed to target a specific audience in a defined geographic area, but is aimed at a vast online audience (Tuck, Silverman 2016, pp. 7-12). The campaign has a main website, a Facebook page and a Twitter account. The campaign's alternative narrative is split into seven videos dedicated to a subject or to a motivation linked to the issue of violent radicalisation, for example, the theme of *discrimination*, the *youth role in western societies*, the *influence of the media*, etc. Each video is split into two micro-narrations linked between them through a hypertext link: the first video section narrates the motivation, while the second presents the alternative narrative from the perspective of an expert or with the contribution of a young practitioner on the theme. As regards the issue of *discrimination*, for example, the first video describes the fictitious experience of Omar, a Barcelona taxi driver who is regularly insulted due to his traditional appearance and clothing (djellaba, long beard, etc.). The video highlights Omar's anger and concerns linked to his job: lots of people avoid him as they think he is a terrorist and he is also forced to leave the gym he attends due to the prejudices of the other clients. The first video emphasises, also thanks to images having strong emotional impact - Omar's growing frustrations and prefigures his turning point to violence as a possible

¹² SAFFRON is a European Union funded project (2016-2018) involving different institutions (private ICT companies, a military academia and one university) in France, Italy and Romania; its main objectives are:
– deliver and test a tool to be used by all relevant players (which are also part of the consortium) to identify in a timely fashion both all internet activities of direct and indirect recruitment of foreign fighters and all signals (weak or strong) pointing at radicalization of single individuals;
– analyse the recent trends about recruitment of young European people by terrorist groups;
– analyse the online communication strategy of terrorist groups;
– develop a social media campaign to contrast their propaganda. Cfr.: www.saffron-project.eu, 20 November 2019.

¹³ The campaign website is www.theheartofdarkness.eu, 20 November 2019.

form of social redemption. Omar's anger is followed by a linked video which suggests, on the other hand, a positive alternative as a possible response to these insults: here, a young representative of the Balon Mondial association¹⁴ tells of the opportunities and potential of football as a means for integration between different nationalities, cultures and religions. The Balon Mondial testimony aims to offer an alternative view on the problems of discrimination experienced by many young refugees and immigrants in Europe: it is not necessarily meant to be a solution, but, rather, a narrative which highlights the tangible opportunities available. The alternative narrative of the campaign is developed, therefore, around the dialectic between the expression of distress that could even lead to violence and the alternative narrative that enhances the options for overcoming prejudices and encouraging social integration. The first video is aimed at telling a story: an individual, fictitious or real experience, with which many young people can easily identify; the second is aimed at calling into question society as a whole - through the contribution of operators, scholars and professionals - to offer a different perspective on the motivations that lead to violent radicalisation. The campaign thus integrates the subjective experiences linked to some kinds of experienced or perceived insult, within a social system exemplified by the different testimonies, in which it is possible to identify an option alternative to violence. The alternative narrative presents the motivations linked to violence as problematic aspects of Western society and not as an external phenomenon. The storytelling articulated in the two videos is also aimed at involving online users as much as possible. The social media videos are often intended for quick consumption: the attention span of young people is low and often limited to a few seconds. In view of this, each video has been split into two interweaving parts: the first, more oriented to the emotional impact, is aimed at attracting attention using audiovisual effects and captivating dialogues, while the second, dedicated to presenting alternative narratives, is developed through a more reflective register

¹⁴ Balon Mondial is a non-profit association from Turin that works on issues of cultural integration through sport. Every year, it organises the “world cup of migrants”, a national football tournament with mixed teams made up mostly of migrants and refugees. <https://www.balonmundial.it/>, 20 November 2019.

through interviews; the hypertext link between the first and second part invites users to continue viewing and not to lose the attention of younger audiences.

As highlighted in a previous analysis (Monaci et al. 2017, p. 226), the campaign communication strategies are expressed both through counter-narratives - arguments that contrast and oppose those of ISIS - and through alternative narratives that develop different perspectives, for example, by emphasising the opposition of the Muslim world to ISIS, or by documenting the cruelty of the organisation through real testimonies or by developing a reflection on problems linked to youth identity of second and third generations of Muslims in the West or by involving the parental figures of young foreign fighters to send a message with emotional impact.

All of the analysed campaigns use videos as the preferred form of communication and use one or more social media channels such as YouTube, Facebook or Twitter thanks to which the videos are disseminated. In some cases, the storytelling of the campaigns involve real testimonies (Extreme Dialogue, theheartofdarkness.eu) or fictitious testimonies (Abdullah X, Average Mohammed, [Toujourslechoix](http://Toujourslechoix.com)) or a group of persons - e.g. the Muslim community ([#inmyname](https://twitter.com/inmyname) , [#notinmyname](https://twitter.com/notinmyname)); all campaigns have a young target as their preferred audience. The campaigns also have elements in common: they are created and promoted by national institutions (governments) or international institutions (European Commission) often in collaboration with important technological partners such as the Alphabet/Google group (Extreme Dialogue, Abdullah X) so as to best develop the opportunities of dissemination and personalisation of the messages.

Based upon personal experiences of research and in view of the cases analysed¹⁵, the essay intends to develop a reflection starting with some critical questions: What is the role of social media campaigns in consideration of the public sphere at risk? What are the limits and criticalities for their issuers? What are the potential effects on audiences?

¹⁵ In particular, the most homogeneous campaigns by objectives and communication strategies are: theheartofdarkness.eu, Average Mohamed, Abdullah X, Extreme Dialogue, [#inmyname](https://twitter.com/inmyname), [#notinmyname](https://twitter.com/notinmyname), [tousjourslechoix](http://tousjourslechoix.com).

4. *Which medium for which message? Social media campaigns and the “Conservative Dilemma”*

The issuers of the analysed campaigns are, for the most part, Institutions - European Commissions, National Governments, Research Institutes, etc. - prompted by the need to develop a response to extremist propaganda addressed to the online public sphere. The analysed campaigns, on the other hand, are initiatives focused on prevention, aimed at informing citizens of the risks linked to undertaking a path of violent radicalisation, which, however, have little or no effect on persons who are already radicalised or co-opted by extremist organisations. Increasing the awareness of individuals to the phenomenon - above all, young Muslims in the West - and combating extremist propaganda are among the main aims of the social media campaigns.

Those campaigns intervene at informative level of the public sphere and use social media to achieve the maximum *spreadability* and to attract the attention of a young audience. The codes of social media storytelling - short videos that use languages of comics and graphic novels or subjective videos so congenial to YouTubers – are used to counter, using the same narrative expedients, the attractive contents produced by jihadist propaganda. That strategy may, however, be found to be a double-edged sword: it may lead the Institutions into what Asa Briggs (2002) described as the *conservative dilemma* in relation to the stance of the Catholic Church during the Lutheran Reform in the 16th century. With the aim of expanding as far as possible, above all in Germany, the base of believers critical of the waste and privileges of the Roman Papacy; Luther and the reformist movement made extensive use of pamphlets - thanks to the recent invention of the printing press - and of forms of popular theatre as instruments of persuasion aimed at the general public. The translation of the Bible into the vernacular and its dissemination on the broad scale by German printers contributed to “privatising” the relationship between believers and the divine, freeing the act of faith from the Catholic liturgy which was strictly in Latin. Against the protestant challenge,

Catholics (..) did not respond (..) with the same media, or at least not on a comparable scale or for an equally broad audience; they did not produce as many pamphlets in defence of the church as the Protestants did to attack it, and they did not publish their own translations of the Bible, as the church considered them dangerous. When they used theatre, they generally targeted an elite rather than a popular audience, for example, the parents of the noble pupils of the Jesuit colleges in France, Italy and central Europe (ivi, p. 103).

In view of the conservative dilemma, the Catholics decided that the correct response on the level of the message could be the incorrect response on the level of the medium. The same dilemma - which message for which medium? - was recently discussed by Clay Shirky in relation to the political power of social media:

The dilemma is created by new media that increase public access to speech or assembly; with the spread of such media, whether photocopiers or Web browsers, a state accustomed to having a monopoly on public speech finds itself called to account for anomalies between its view of events and the public's. The two responses to the conservative dilemma are censorship and propaganda (2011, pp. 6,7).

Today, institutions and governments, above all if attacked in some of their fundamental values - as often occurs in jihadist propaganda which denounces the West, economic imperialism and moral relativity of liberal societies - are faced with a double dilemma: how to respond? By censoring or by launching counter-propaganda initiatives? And if the second option is chosen, which media should be used? Which messages should be conveyed?

While the deliberate censorship of extremist contents - above all if done on the most widespread social networks - can lead to polarising even further the positions (Zuckerman 2008) – on the other side, the production of social media counter-narratives may dangerously legitimate extremist organisations as issuers in the public sphere. Using the same media strategies as the opponents can, therefore, on one side, lead to an excessive simplification of very complex issues such as the cultural identity of Muslims in the West, economic inequality between the North and South of the world, etc., and, on the other side, also offer a

sounding board to messages and points of view that strengthen polarisation, rather than reduce it.

The analysed campaigns arise from the intention of the Institutions to use social media strategically, seeking to position against the extremist voices the liberal perspectives of Western democracies: some experiences demonstrate, however, that the task is very complex and that the public sphere of social media is a problematic space. This is the case of the campaign *Think again, Turn Away*: a 2015 initiative of the US State Department on the main social media platforms (Facebook, Twitter, ask.fm) aimed at dissuading *foreign fighters* from joining the ISIS army in Syria above all within the US Islamic community. The campaign invoked different criticisms: US Islamic communities denounced it as a pro-Israeli action aimed at subtly associating the entire Islamic community with the atrocities of ISIS¹⁶; some commentators and experts in Diplomacy criticised it for the superficiality of the arguments made in response to the provocations of users on Twitter, and for the carelessness of having published without verification the fake news relating to alleged violence perpetrated by ISIS in the city of Mosul. As a further counterproductive effect, the Twitter channel of the campaign triggered repeated violent verbal exchanges between the reckless moderators of the channel - who, moreover, only wrote in English - and individuals writing in Arabic, dismantling and criticising the main arguments and using the Twitter channel precisely as a sounding board for a counter-narrative against the initial counter-narrative (Rogers 2015).

The very campaign www.theheartofdarkness.eu, created by the writer was subject to many different revisions related to its video contents and the possibility that they may be manipulated and reused for purposes opposed to the declared aims of the campaign: several times between 2017 and 2018 the videos were correctly, contextualised and redeveloped with phrases of context aimed at clarifying - also in an excessively pedagogic manner - the video contents of the campaign so as to make politically unequivocal the key message of the social media campaign.

¹⁶[https://www.aaiusa.org/think again turn away we re looking at you state department](https://www.aaiusa.org/think_again_turn_away_we_re_looking_at_you_state_department), 2 July 2019

Institutions and governments have enjoyed for quite some time a condition of information asymmetry in which they held a certain level of control over the dissemination of information, over the possibilities of censure or of propaganda: it was then a public sphere in which there were few issuers and the many recipients had little control over the information. Social media may significantly alter that “privilege”: as tools of *spreadability*, they are means of *mass self-communication* (Castells, Ib.), in which there are many issuers and anyone can - the Governments as well as the extremist organisations - grab the attention of millions of people; the culture of social media, in addition, as highlighted by Burgess and Green (Ib.), values, rather than sanctions, the redevelopment of other people's contents – see the phenomenon of *memes* – and the spreading of altered contents as a practice of remixing and not as appropriation of other people's materials.

Social media are also ductile tools, oriented at the same time towards practices of participatory communication which replace the mediation of institutional issuers but at the same time also moulded by pervasive technologies that determine their current affordances and future uses.

5. The role of communication campaigns: instruments of awareness-raising or polarisation?

As illustrated in the above paragraph, the issuers of campaigns - governments, international institutions, organisations of civil society – are facing the conservative dilemma since social media can trigger strategies of counter-narration, of counterproductive reprocessing of messages, thereby legitimating precisely the issuers that the campaigns are seeking to discredit.

From the audience perspective, what role can social media campaigns play? Are social media really the place to increase social awareness or do they contribute to the polarisation of individuals on irreconcilable positions?

Recent studies state that in most cases the process of ideological radicalisation occurs in a self-directed way: it is the single individuals - also thanks to online anonymity - that actively seek on social media news, references or profiles referable to forms of violent extremism (Vidino 2014). Social media, on the other hand, are the ideal platform for strengthening the personalisation of

individual choices by encouraging the phenomenon of the filter bubble introduced in par. 1. The reasoning of Parisier (ib.) develops as follows: to maximise the user base - which also corresponds to a useful value for advertisers - social platforms such as Facebook suggest contents and contacts that are as close as possible to the tastes, political and religious opinions of the person active on the platform. That strategy derives from consolidated knowledge in the field of social psychology, according to which it is likely that an individual will read news or share contents with similar persons (same political stance, same cultural consumptions, musical tastes, etc.), and not with completely different persons. This is the phenomenon of *homophilia* which social networks tend to amplify thanks to the introduction of filters: algorithms able to select information in place of users in view of their marketing aims. The distortions made by the filters are, however, opaque for the final user: we do not realise that we are inside a bubble as it is the algorithm that filters what could be of interest to us. That phenomenon obviously becomes more serious the more the opinions of the individual are polarised on extreme ideas, positions and tastes: in these cases, the filter does nothing but amplify and strengthen already consolidated positions and beliefs. From this point of view, the only campaigns that can invert or at least mitigate the positions of the persons are those that re-direct the decisions of the user from his very first interactions: for example, the re-direct campaign promoted by Jigsaw. The campaign, as described in section 3, intends to intercept from the outset the search for extremist contents, re-directing the choices of the person towards critical sources and audiovisual contents able to counter violent messages. The project develops starting with the analysis of propaganda contents in English and in Arabic (videos, magazine texts, interviews, etc.); it identifies some prominent narratives and develops audiovisual or textual contents in response to the propaganda material. Those contents are directed to users whenever they search with sensitive keywords from the Google search engine. This thereby avoids the person, still in the cognitive exploration phase, falling into an “extreme” bubble further amplified by the filters which reinforce the person's initial choices.

Faced with persons who are already polarised or have decidedly extremist positions, those actions have, however, a weak hold: the counter-narratives that

propose visions and perspectives substantially in contrast with extremist visions can even be harmful to the extent that they produce what Sunstein defines as the *backfire effect* (ivi, pp. 93-94). As demonstrated by various experiments in the political field, in fact, a denial or a correction does not produce a change of opinion in persons already polarised, but tends, on the other hand, to corroborate their initial positions, and are therefore unsustainable. Ultimately, rational and argumentative criticism - precisely the counter-narrative - does not necessarily lead to a change of stance but rather can produce a more obstinate polarisation. Other studies also demonstrate that that process is all the more effective, the more interactions between persons occur online and not face to face: in the case of virtual discussions, even a-synchronous, there is a stronger polarisation of the opinions of persons around an idea or a principle compared to what happens in control groups which also interact on the physical level.

An example of the *backfire effect* is the reaction to the campaign #toujourslechoix: the YouTube channel videos, promoted by the French government to highlight the sense of individual responsibility in choosing the life path, were redeveloped, remixed and then spread online in satirical form using the hashtag #paslechoix precisely by a group of French supporters of ISIS. The parody triggered a counterproductive spread which drastically de-strengthened the initial intention, raising even more criticisms of the French Government's attempt to diffuse the propaganda with a crude action of ineffective counter-propaganda (Trautenberg 2016). Here, the counter-narrative was manipulated in an attempt to call into question its fundamental premises: in the face of the narrative based upon the possibility of individual choice, the social media parodies responded by completely overturning the argument, emphasising the hypocrisy of the campaign message and highlighting the radical lack of alternatives to the path of violence. What they parodied was the false presupposition of choice and the inevitability of violence as the only viable option.

6. Conclusions

The turn of the Millennium hailed social media as new technologies of freedom - freedom of expression and freedom of access - and as fundamental carriers for enhancing and even in some cases exporting the values typical of liberal democracies (Morozov, Ib.). A public sphere of the few - as in the argument of Habermas – was replaced by a public sphere of the many, potentially extendible to all world citizens. The expansion of the public sphere, however, incorporated not only the multiplication of voices, but also the reverberation of the broader economic, political and social upheavals that have characterised the last twenty years. The growing polarisation of civil society on fundamental issues such as the migrant emergency, the threat of terrorism, etc., and at political level, the spread of political and religious extremism as a bulwark of *resistance identity* (Castells 1996) have found in the public sphere of social media an enormous sounding board, uncontainable, seductive in terms of the media and easily accessible.

As I have attempted to describe in the above paragraphs, in some cases the very affordances of social media - *pervasiveness, scalability, searchability* - make them so easy to manipulate from the technical perspective, re-directing users often without them being aware of those distortions. On the other hand, however, as stated by Kratzenberg, the technology is not, in itself, positive, or negative, not even neutral. All technology - particularly when it is a communication medium - can be considered a socio-technical system that is developed, also, based upon principles, values and orientations defined by rational subjects. So what is the role of political and social institutions in the public sphere of social media? It cannot be just to censor or to respond to attacks on Democracy by the same means as propaganda.

This essay has sought to reflect on the role of communication campaigns in combating jihadist extremism and on the criticalities of those initiatives in the scenario of a public sphere at risk. The Institutions and civil society cannot and must not leave to the technological players the possibility of regulating - censoring or, conversely, multiplying – the voices and points of view that ideologically justify violence. The online public sphere is basically also a public

asset with which it is possible to raise awareness and educate individuals: it is not just a case of freedom of expression and access but of understanding of the rules and values that guarantee those rights, even online.

Bibliography

- Atwan, A. B. (2015). *Islamic state: The digital caliphate*. Univ. of California Press.
- Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.
- Boutin, B., Chauzal, G., Dorsey, J., Jegerings, M., Paulussen, C., Pohl, J., ... & Zavagli, S. (2016). The foreign fighters phenomenon in the European Union. *Profiles, threats & policies. The International Centre for Counter-Terrorism–The Hague*. doi, 10(2016.1), 02.
- Boyd, D. (2010). Social network sites as networked publics: Affordances, dynamics, and implications. In *A networked self* (pp. 47-66). Routledge.
- Briggs A., Burke P. (2000). *A Social History of the Media. From Gutenberg to the Internet*, Polity Press, Cambridge; trad. It., *Storia sociale dei media. Da Gutenberg a Internet*, 2002, Il Mulino, Bologna.
- Briggs, R. & Silverman, T. (2016). Western Foreign Fighters. Innovations in Responding to the Threat. Institute for Strategic Dialogues. http://www.strategicdialogue.org/wp-content/uploads/2016/02/ISDJ2784_Western_foreign_fighters_V7_WEB.pdf 1 July 2019
- Burgess, J., Green J. (2009). *YouTube*. Cambridge: Polity Press.
- Castells, M. (1996). *The power of identity*, Blackwell, Oxford.
- Castells, M. (2013). *Communication power*. OUP Oxford.
- Colliver, C., & Davey, J. (2017) Cross-Spectrum Counter Violent Extremism: Prevention and Intervention Models. *Digitale Medien und politisch-weltanschaulicher Extremismus im Jugendalter*, 164.
- Fuchs, C. (2014). Social media and the public sphere. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 12(1), 57-101.
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
- Habermas, J. (1962). *Strukturwandel der Öffentlichkeit*, Suhrkamp Verlag, Frankfurt am Main; tr. it., 2015, *Storia e critica dell'opinione pubblica*, Laterza, Bari – Rome.
- de Sola Pool I. (1983). *Technologies of freedom*. Harvard University Press.
- Jenkins, H. Ford. S., & Green, J.(2013). *Spreadable media: Creating value and meaning in a networked culture*; trad. it., 2013, *Spreadable media. I media tra condivisione, circolazione, partecipazione*, Maggioli Editore, Santarcangelo di Romagna.
- Kepel G., *Terreur dans l'Hexagone. Genèse du djihad français*, Paris: Gallimard, 2015.

- Khosrokhavar, F. (2017). *Radicalisation*. Éditions de la Maison des sciences de l'homme.
- Meleagrou-Hitchens, A., & Kaderbhai, N. (2017). Research perspectives on online radicalisation. *International Centre for the Study of Radicalisation*.
- Monaci S., Mazza C., Taddeo G. (2017). Designing a Social Media Strategy against Violent Extremism Propaganda: The# heartofdarkness Campaign. *Digitale Medien und politisch-weltanschaulicher Extremismus im Jugendalter*, 213. https://www.dji.de/fileadmin/user_upload/bibs2017/Digitale_Medien.AFS.Band.13.pdf#page=215 1 Luglio 2019.
- Morozov, E. (2011). *The net delusion: How not to liberate the world*. Penguin UK; trad. it., 2011, L'ingenuità della rete, Codice, Turin.
- Papacharissi, Zizi. (2009). The Virtual Sphere 2.0. The Internet, the Public Sphere, and Beyond. In *Routledge Handbook of Internet Politics*, eds. Andrew Chadwick and Philip N. Howard, 230-245. New York: Routledge.
- Parisi, E. (2012). Il filtro. *Milano: Il Saggiatore*.
- RAN (Radicalization Awareness Network); 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, 1 July 2019
- Reed, A. G., & Pohl, J. (2017). The four dimensions of the Foreign Fighter threat: making sense of an evolving phenomenon. *Journal Title: International Centre for Counter-Terrorism Policy Brief*, 8(1).
- Rogers, A. "'Think Again, Turn Away'...from Lousy Public Diplomacy," *Middle East Report Online*, October 21, 2015. <https://merip.org/2015/10/think-again-turn-away-from-lousy-public-diplomacy/> 1 July 2019
- Shirky, C. (2011). The political power of social media: Technology, the public sphere, and political change. *Foreign affairs*, 28-41.
- Speckhard, A., Shajkovci, A., Bodo, L., & Fazliu, H. (2018). Bringing Down the Digital Caliphate: A Breaking the ISIS Brand Counter-Narratives Intervention with Albanian Speaking Facebook Accounts. International Center for the Study of Violent Extremism (ICSVE), 14th February. <https://www.icsve.org/bringing-down-the-digital-caliphate-a-breaking-the-isis-brand-counter-narratives-intervention-with-albanian-speaking-facebook-accounts/>, 1 July 2019
- Sunstein C. (2017). *#republic*. Princeton University Press, Princeton.
- Zuckerman, E. (2008, March). The cute cat theory of digital activism. In *E-Tech Conference, San Diego, CA, USA*. Abstract retrieved from: <http://en.oreilly.com/et2008/public/schedule/detail/1597>. 1 Luglio 2019
- Vidino, L. (2014). Home-Grown Jihadism in Italy: Birth, Development and Radicalization Dynamics. *Foreword by Stefano Dambruoso*. Milan: ISPI and European Foundation for Democracy.