Stanford Law School: Law & Policy Lab

Countering Extremist Speech Online

A Report to the Bureau of Conflict & Stabilization Operations
U.S. Department of State

June 2015

Michael Goldfien MA'15
Michael Woolslayer MA'15

Ford Dorsey Program in International Policy Studies Stanford University

Adviser: Beth Van Schaack, Leah Kaplan Visiting Professor in International Human Rights
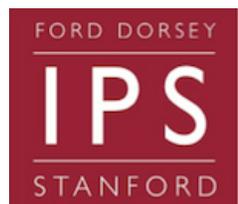
Executive Summary

Dealing with the threat of extremism online has taken on a new urgency in recent years, as prominent extremist organizations like the Islamic State in Iraq and the Levant (ISIL) and Al-Qaeda in the Arabian Peninsula (AQAP) have proven adept at wielding online influence well beyond the territory that they inhabit or control. These groups and their sympathizers use social media sites like Facebook and Twitter—both founded after 9/11—to make their message accessible to people across the world.

While the Internet has undoubtedly become a forum for the promulgation of extremist ideas and the dissemination of know-how for would-be terrorists, its precise role in the radicalization and recruitment process remains open to debate. Though extremists, particularly in the West, are often found to have accessed extremist content online, research into the radicalization process suggests that in-person interactions play a crucial role. People rarely self-radicalize solely through the consumption of extremist media via the Internet.

It is also unclear that "confronting squarely and honestly" the ideas proffered by extremists will be effective, particularly if those instigating the confrontation are unlikely to be seen as credible by the intended audience. Implicit in many official U.S. statements is the notion that challenging the beliefs of those drawn to extremist ideas will lead to the rejection of those ideas and the embrace of moderation. However, recent social science research suggests that seeking to correct beliefs is often ineffective and can even "backfire."

Six recommendations lead from these research findings:

1. ***Don't focus on online extremism at the expense of community-level engagement.*** A strategy to counter extremist narratives online may be part of the solution, but it should be accompanied by a campaign to address extremist discourses and the presence of recruiters in communities.
2. ***Restrictive measures—like website takedowns and content filtering—are just one tool that must be paired with broader counter-messaging and other interventions.*** Many current P/CVE efforts online focus on removing and blocking access to extremist speech where it is found. Such measures must be wielded surgically in support of a wider P/CVE strategy to be effective.
3. ***Tailor the message to the audience.*** Counter-messages that may seem persuasive and reasonable to moderates may lead to a "backfire effect" for those on the fringe. Authors of counter-messages should seek to understand the perspective of those on the fringe, and tailor their messengers and narratives to that audience.
4. ***Provide causal alternatives to false extremist narratives.*** Social science research suggests that people are more likely to be persuaded by corrections that provide a causal alternative to their pre-existing belief. When extremists promote pernicious and false narratives—e.g. that the United States directly targets civilians in Iraq and Syria, for example—it is not enough to simply deny the claim; rather, an accurate causal story must be supplied.
5. ***The messenger matters.*** Research indicates that even moderate populations that reject most extremist ideology will dismiss information emerging from the U.S. Government.

U.S. Government efforts must therefore be focused on empowering voices within vulnerable communities that retain influence and respect.

6. ***Affirm correct information rather than negate incorrect beliefs.*** Some psychological and linguistic research suggests that negations can actually reinforce the idea their purveyors are attempting discredit. Thus, counter-messages, where possible, should seek to affirm accurate claims and facts rather than negating false narratives or lies.

Introduction

Speaking at the White House Summit on Countering Violent Extremism, President Barack Obama told an audience of civil society actors, law enforcement officials, and religious leaders that "[w]e have to confront squarely and honestly the twisted ideologies that…terrorist groups use to incite people to violence."[1] Obama continued, arguing that "[w]e need to find new ways to amplify the voices of peace and tolerance and inclusion, and we especially need to do it online."[2] Since the so-called war on terror began in the wake of 9/11, the battle for the hearts and minds of people around the world has increasingly taken place on the Internet. If extremist groups use the Internet as a vehicle to spread hateful ideas, the argument goes, the West must use the Internet to offer compelling counter-narratives.

As President Obama's comments suggest, dealing with the threat of extremism online has taken on a new urgency in recent years. Today, prominent terrorist organizations like the Islamic State in Iraq and the Levant (ISIL) and Al-Qaeda in the Arabian Peninsula (AQAP) have proven adept at using the online medium to spread propaganda well beyond the territory that they inhabit or control. These groups and their sympathizers use social media sites like Facebook and Twitter—both founded after 9/11—to make their message accessible to people across the world. ISIL uses the internet to disseminate not only clips of beheadings, but "infomercial-like" films tracing its origins to Osama bin Laden and propaganda videos predicting coming clashes between the caliphate and its enemies.[3] Meanwhile, AQAP publishes a slick online English-language magazine called *Inspire*, which includes bomb-making instructions alongside jihadist propaganda.[4] This type of online content is often claimed to have played a role in the radicalization process of Western perpetrators of terrorist attacks.[5]

However, while the Internet has undoubtedly become a forum for the promulgation of extremist ideas and the dissemination of know-how for would-be terrorists, its precise role in the radicalization process remains open to debate. Though terrorists, foreign fighters, and other extremists, particularly in the West, are often found to have accessed extremist content online, research into the recruitment of terrorists, foreign fighters, and jihadists suggests that in-person

---

[1] Barack Obama, "Remarks by the President in Closing of the Summit on Countering Violent Extremism," February 18, 2015.

[2] Ibid.

[3] Graeme Wood, "What ISIS Really Wants," *The Atlantic*, March 2015.

[4] Ghaffar Hussain and Erin Marie Saltman, "Jihad Trending: A Comprehensive Analysis of Online Extremism and How to Counter It," *The Quilliam Foundation*, May 2014, *77*.

[5] For example, in a recent New York Times article about Boston Terror Suspect Usaamah Abdullah Rahim, chairman of the House Homeland Security Committee Representative Michael McCaul was quoted as saying that "[t]hese cases are reminders of the dangers posed by individuals radicalized through social media." Jess Bidgood and Dave Philipps, "Boston Terror Suspect's Shooting Highlights Concerns Over Reach of ISIS," *The New York Times*, June 3, 2015. http://www.nytimes.com/2015/06/04/us/usaama-rahim-boston-terrorism-suspect-planned-beheading-authorities-say.html?smid=nytnow-share&smprod=nytnow

interactions play a crucial role. That is, people rarely self-radicalize solely through the consumption of extremist media via the Internet.

Furthermore, it is unclear that "confronting squarely and honestly" the ideas proffered by extremists will be effective, particularly if those instigating the confrontation are unlikely to be seen as credible in the eyes of those tempted by extremist ideologies and beliefs. Implicit in President Obama's statement is the notion that challenging the beliefs of those drawn to extremist ideas will lead to the rejection of those ideas and the embrace of moderation. However, recent social science research suggests that seeking to correct beliefs is often ineffective and can even "backfire."

This paper engages these issues in five sections. First, it surveys the existing literature on the role of the Internet in the radicalization process and in the recruitment of individuals to extremist causes. Second, the current state of online counter-messaging is analyzed. Third, it assesses the applicability of recent social science research into political beliefs and corrective information to the topic of extremism and counter-messaging. Fourth, it considers the implications of the surveyed research on online extremism and political beliefs for policy. Finally, the paper concludes by summarizing our findings and proposing areas for future research.

The Role of the Internet in the Radicalization and Recruitment Process

The increasingly sophisticated use of the Internet and social media sites by extremists groups like the Islamic State of Iraq and the Levant (ISIL) emphasizes the importance of understanding the role of such online tools in the process of radicalization and recruitment. Videos on hosting sites such as YouTube, virtual extremist discussion forums, and online information hubs and wiki-style how-to guides proliferate. The White House acknowledged that the Internet has "provided violent extremists with access to new audiences and instruments for radicalization" in February 2015 when it created an Interagency Working Group to Counter Online Radicalization to Violence to begin to coordinate U.S. Government efforts in this arena.[6] Yet the effectiveness of these mechanisms in the process of radicalization and recruitment to extremist causes has only recently received significant attention in the academic and policy literature. The preliminary consensus in the literature, still much debated, suggests that while online tools play an important and growing role in the dissemination of extremist ideology and knowledge, the Internet has not yet replaced in-person interactions and networks as a key and necessary step in radicalization and recruitment.

Extensive research has been conducted on the potential for the Internet to facilitate the development of social and political movements of all types. Friedland and Rogerson (2009), for example, indicate that the Internet is enabling groups previously incapable of political action to develop and amplify their voices. Advances in information technology ensure that like-minded individuals are better able to locate and converse than ever before. While such technologies are, in principle, value-neutral, the ease and anonymity of interaction enable negative connections and the dissemination of dangerous ideas to a great extent than in the past. Indeed, recent

---

[6] Quintan Wiktorowicz, "Working to Counter Online Radicalization to Violence in the United States," The White House: February 5, 2015.

research confirms a perception held by many: the Internet extends the reach of extremist messaging in particular.

Edith Cowan University Security Research Centre researcher Simon O'Rourke (2007) writes that the development of virtual communities capable of exploiting rapidly evolving social media technologies provides advantages to amorphous, fluid network organizations like extremist groups through the ease of transnational information flows. Lisa McInerney and Maura Conway (2008), researchers at Dublin City University's Institute for Conflict Resolution and Reconstruction, conducted an exploratory study on extremists' use of YouTube videos and their potential to lead to self-radicalization. Their initial findings suggest that extremist content has spread quickly beyond traditional extremist websites and dedicated forums and into so-called Web 2.0 platforms—including video sharing and social networking—and practices—including user-generated content and the formation of virtual communities. The primary consumers of such Web 2.0 content are outside of core support areas in the Middle East and North Africa, and thus represent a new population of potential Internet-enabled radicals, converts, and political sympathizers within diaspora populations.

The increased extent and availability of extremist content online is well-documented, but researchers still disagree on the degree to which that accessibility translates into related offline activity. Research conducted by Associate Professor of Political Communication at the University of Amsterdam Magdalena Wojcieszak (2009, 2011) indicates that participation in online extremist groups exerts a growing influence on the type and frequency of offline political engagement. Less clear is the extent to which this expanded messaging capacity translates into explicit membership in extremist organizations and violent actions. In one study, Associate Professor of Media, Culture and Creative Arts at Edith Cowan University Dr. Anne Aly (2010) traced the radicalization of two individuals, one Singaporean and one Australian, and attempted to isolate the impact of the Internet on the process. She found that online interactions played a developing role in the progression of radicalization, confirming that the Internet has become an important tactical tool in the extremists' repertoire. Aly also indicates that the exact role of this new tool and its effectiveness in helping convert latent beliefs into violent actions is not fully understood. In contrast, the Director of the German Institute on Radicalization and De-radicalization Studies, Daniel Koehler (2014), has compiled evidence that the Internet is a major driving factor in fueling grievance-based identities. Based on eight German right-wing extremists, he found a link between the Internet and the establishment and fostering of so-called radical "contrast societies"—his term for the space where social movements and their environments interact, with a focus on perceived differences. Koehler suggests that as a place of information exchange, ideological development, and training, the Internet often shaped, or even made possible, the individual radicalization process.

The current academic consensus is that online mechanisms can contribute to radicalization and recruitment to extremist causes, but the tipping point into violent action still requires in-person socialization, despite dissenting voices like Koehler's noted above. Research fellow at the Norwegian Defense Research Establishment Hanna Rogan (2006) shows that while the Internet is an important tool in facilitating ideological cohesion and network-building among extremists, it is not a direct means of recruitment. Likewise, RAND Corporation researchers Von Behr, Reding, Edwards, and Gribbon (2013) suggest that the Internet does create more opportunities to be radicalized and acts as an "echo chamber" that reinforces existing beliefs. Yet they find that

there is little evidence of the Internet accelerating the pace of radicalization, and no indication that the Internet allows radicalization without physical contact; online interaction complements rather than replaces in-person interactions.

Numerous empirical studies strengthen this consensus that in-person interactions remain essential to radicalization and the recruitment to extremist causes. Counterterrorism scholars Gartenstein-Ross and Grossman (2009) empirically investigated radicalization in the U.S. and the U.K., isolating the six major steps in the process.[7] The Internet and online interactions supported progression through the steps, but did not replace in-person facilitation at any point in any of the examined cases. Quilliam Foundation researchers Hussain and Saltman (2014) found in their study of online extremist search terms that the majority of radicalized individuals are exposed to extremist ideology offline before pursuing further online indoctrination. The nature of extremist searches indicates that individuals do not "stumble upon" extremist content, but are already conditioned to seek it when they go online.

The Internet continues to vastly increase the reach of extremist ideology and propaganda, and to create new virtual spaces for its exchange and dissemination. Yet a causal link between extremist presence online and radicalization is, according to recent research, a tenuous one. Internet communities can simplify or accelerate individual movements toward extremism, but there is little evidence in social science research, despite media reports to the contrary like those surrounding the Boston Marathon bomber, to suggest the frequent occurrence of wholly online radicalization. While the Internet is a tool used progressively more by extremists around the world than ever before, the current state of research indicates that in-person interactions remain an essential step in the transition to violence.

Online Counter-Messaging

The growth of the use of the Internet as a tool by extremist organizations has generated calls for governments to leverage online mechanisms to counter extremist messaging. The study of online radicalization and recruitment is quite young, and the field of online counter-messaging is even more nascent. Multilateral efforts like the United Nations' Counter-Terrorism Implementation Task Force (CTITF) Working Group on Use of the Internet for Terrorist Purposes have had limited success in codifying and promoting best practices. The existing literature addressing the proper approaches to countering extremism on the Internet highlight three central facets of a workable counter-messaging program. Director of the Middle East Studies Program in the Institute of Arab and Islamic Studies at the University of Exeter, Dr. Omar Ashour (2011), provides a three-part overview. First, effective counter-messaging must address all aspects of extremist narratives, including theological, political, and historical. Second, credible messengers must be deployed, primarily drawn from former extremists and influential community organizations. Third, counter-messaging needs to leverage all forms of media being used by the extremists.

Reviews of counter-messaging programs show that the emphasis to date has been on restrictive measures, like website takedowns and filtering, but also discuss the limitations to the possible

---

[7] The six steps are: 1) Adopting a legalistic interpretation of Islam; 2) Trusting only select religious authorities; 3) Perceived schism between Islam and the West; 4) Low tolerance for perceived theological deviance; 5) Attempts to impose religious beliefs on others; and 6) Political radicalization.

effectiveness of such responses. Neumann and Stevens (2009) stress that, since online extremism is a reflection of real-world phenomena, it cannot be dealt with through crude, expensive, and counter-productive "pulling-the-plug" measures. Instead, a multi-prong strategy should include: deterring the producers of extremist materials; empowering online communities to self-regulate; reducing the appeal of extremist messages; and promoting positive messages. Former Director of the West Point Combating Terrorism Center James Forest (2010) highlights that U.S. counter-messaging efforts must evolve away from a heavy dependence on restrictive measures to embrace competition with the many sources of information available online, since attempts to simply silence radical sites will be costly and ineffectual. Restrictive measures should be used sparingly as a complement to less expensive and likely more effective approaches.

The most promising counter-messaging strategies have a light or invisible government footprint, deemphasize the reduction of the "supply" of extremist messaging, and address the "demand" through a focus on the leadership in vulnerable communities and social media influencers. One study conducted by Briggs and Feve (2013) of the Institute for Strategic Dialogue described a "counter-messaging spectrum" comprised of three main activities—government strategic communications, alternative narratives, and counter-narratives—with the bulk of resources expended on the direct government strategic communications, which are likely the least effective of the three categories on the spectrum. Government counter-narratives are also likely to be ineffective, and also run the greatest risk of promoting the "backfire effect" and entrenching the very ideologies they seek to counter. RAND Corporation researchers York and Chalk (2013) recommend "de-securitizing," or removing from the control of security forces, efforts to counter extremist messaging, and focusing on engaging and educating civic and social influencers within vulnerable communities. Director of the International Centre for the Study of Radicalization and Kings College London Professor Peter Neumann (2013) argues that governments' pursuit of restrictive measures is undesirable and ineffective, with resources better spent reducing the demand for extremism by promoting education and encouraging civic challenges to extremist narratives.

Political Beliefs and Corrective Information

The widespread use of the Internet and social media by terrorist and jihadist groups has made online counter-messaging and debunking strategies an alluring option for policymakers. The discourse about Islamist extremism among Western politicians and policymakers often portrays Islamists, jihadists, and terrorists as having a fundamentally flawed or incorrect understanding of Islam. President Obama, at the White House CVE Summit, referred to the ideologies used by terrorists to incite violence as "twisted."[8] British Prime Minister David Cameron has similarly argued that problems of extremism in the United Kingdom have "nothing to do with true Islam."[9] In a televised speech following the Charlie Hebdo massacre, French President Francois Hollande made a point to say that the perpetrators had "nothing to do with the Muslim religion."[10] The

---

[8] Barack Obama, "Remarks by the President in Closing of the Summit on Countering Violent Extremism," February 18, 2015.

[9] Christopher Hope and Steven Swinford, "David Cameron: Muslim Council of Britain 'has a problem' over extremism – not Eric Pickles," *The Telegraph*, January 19, 2015.

[10] *Reuters*, "France Pursues Female Suspect After Deadly Sieges," January 10, 2015.

way that these leaders talk about Islamist extremism suggests that Western governments view terrorists and their sympathizers as misinformed.

This view has led Western governments to promote counter-messages that stress a "correct" understanding of Islam and its relationship with the West and that debunk conspiracy theories (e.g. 9/11 was an inside job) widely held among Islamist extremists. As President Obama suggested at the CVE Summit, "voices of peace and tolerance and inclusion" may be important tools for weakening the appeal of extremist ideas. This argument reflects a faith in the ability of civil discourse and the identification of shared facts to narrow the gap between people of differing viewpoints. Yet this assumption should be held to scrutiny if counter-messaging is to become a central component of efforts to prevent or counter violent extremism.

There has been relatively little research into the ways that challenges to people's misperceptions impact their beliefs. That is, under what circumstances do people change their mind? The tendency is to assume that the provision of corrective information will cause people to adjust their views in ways that align more closely with objective truths. However, recent research, much of it done by Brendan Nyhan of Dartmouth and Jason Reifler of the University of Exeter, suggests that corrective interventions are not only often ineffective at changing beliefs, but can actually have the opposite effect. That is, in certain cases, people appear to become more committed to their inaccurate beliefs after being exposed to corrective information.

In one study, Nyhan and Reifler (2010) carried out a series of experiments in which subjects were given a mock news article either containing a factually misleading statement by a politician or a factually misleading statement by a politician followed by a factual refutation or "correction." The authors found that corrections failed "to reduce misperceptions" among the participants most committed to the misperception. In cases in which the subject's ideology dovetailed with the misperception, a correction could "actually *strengthen* [their emphasis] misperceptions." Nyhan and Reifler show that "corrective information in news reports may fail to reduce misperceptions and can sometimes increase them for the ideological group most likely to hold those misperceptions."

The authors' results support the notion that people tend to engage in *motivated reasoning*. There is a significant literature on motivated reasoning, which suggests that people often interpret information subjectively, welcoming facts or ideas that are consonant with their pre-existing beliefs or worldview and downplaying or outright rejecting facts or ideas that are dissonant with their pre-existing beliefs or worldview. The "backfire effect"—which involves an increased commitment to factually incorrect beliefs—observed in Nyhan and Reifler's study may stem from this phenomenon. Those most ideologically committed to a misperception may be provoked into developing counter-arguments that support the previously held belief when confronted with a correction.

Though Nyhan and Reifler's work, as well as the broader literature on motivated reasoning, can be sobering in the context of efforts to disseminate counter-messages of tolerance and moderation, it is nonetheless true that people do change their minds. In a follow-up article, written for the New America Foundation, Nyhan and Reifler (2013) discuss several factors likely to impact the persuasiveness of a correction: its source, its phrasing, and its ability to provide a causal explanation for observed outcomes. According to their research, people may be apt to discount or disregard information coming from sources they do not consider credible or

legitimate. Nyhan and Reifler conduct experiments that test the impact of the source of corrections on subjects' misperceptions by attributing the corrective information to different think tanks and news outlets, each with different ideological leanings. For example, people who identified as conservative were told that President Obama did not raise taxes for families making under $250,000 annually, with the premise that conservatives were more likely to believe, incorrectly, that he had. The authors found misperception to be highest among conservative subjects who had received a correction about President Obama's tax record from sources generally perceived to be liberal-leaning. Furthermore, conservatives exposed to a "liberal" source reported statistically significantly greater negative feelings toward President Obama after receiving a correction than they had before. These results suggest that the persuasiveness of corrective information depends in part on its source; corrections coming from sources seen as credible may be more likely to disabuse people of their misperceptions.

The phrasing of corrections may also make a difference. Nyhan and Reifler cite research (Mayo et al., 2004) suggesting that corrections in the form of a negation may, in some cases, lead people make associations incongruent with the intended meaning of the correction. As Nyhan and Reifler put it, Mayo et al. found that "negating descriptors that lack an opposite concept (e.g., 'criminal') can backfire…saying that 'John is not a criminal' may cause *greater* associations between the concept of John and criminality, reinforcing the association the speaker intends to falsify." Our brains do not do well with the "no or not tags," and can simply fall away in our memories, leading people to recall the idea the correction intended to discredit.[11] By contrast, a statement that "John was exonerated" may enjoy greater accurate retention.

Finally, corrections that provide a logical causal story may make misperceptions more vulnerable to dislodgement. Humans are naturally uncomfortable with uncertainty, and thus seek explanations for what they observe in the world. However, the causal stories that humans construct can be simultaneously incorrect and convincing. Johnson and Seifert (1994) conduct an experiment in which groups of subjects are variously given information about a warehouse fire linked to volatile materials stored inside. They later provided some groups with corrections as to the cause of the fire. The authors show that misinformation—in this case about the cause of the fire—can have a persistent impact on inferential reasoning even after a factual correction is provided; despite receiving subsequent information that volatile materials were in fact not present, volatile materials still factored into subjects' retained explanations of the fire. However, the authors found that the influence of misinformation about the fire could be mitigated when corrections included or were accompanied by an "alternate causal explanation," i.e., that the story about volatile materials was false *and* that evidence indicating arson was discovered. When subjects were provided an "alternate causal explanation," volatile materials were less likely factor into their explanations of the fire.

Nyhan and Reifler observed a similar phenomenon in a study that provided subjects with information regarding the resignation of a (fictional) state senator. Initially provided with speculative information that the resignation was due to corruption, subjects who received a correction accompanied by speculation that the state senator had taken a job as president of a university were statistically significantly less likely to believe the state senator was corrupt.

---

[11] Interview with Brendan Nyhan, May 14, 2015.

Again, corrections that include or are accompanied by an alternative causal explanation appear to be more effective at displacing misperceptions.

Research on political beliefs may be instructive as the United States and other countries embark on counter-messaging campaigns aimed at highlighting flaws in extremists' ideologies, inaccuracies in their propaganda, or hypocrisy in their actions. It is sometimes claimed that facts will speak for themselves in the marketplace of ideas. However, the research cited above seems to undercut this assumption. Rather, facts often fail to influence beliefs even when made explicit. Worse, in some cases, facts are manipulated, minimized, or ignored by people psychologically motivated to protect their deeply held beliefs from being undermined.

It should be noted that the research cited above does not focus on extremists. Further, the corrections involved in the studies conducted by Nyhan, Reifler, and others focus on the correction of *beliefs*, where there is an alternative objective truth, rather than *ideology*, which is inherently subjective. As such, extrapolating their results to efforts to counter violent extremism and strategies for counter-messaging should be done with caution. Nonetheless, the research cited in this section can inform policy discourse on the fight against extremism. Social science indicating the salience of motivated reasoning suggests that in some cases letting the "facts speak for themselves" may have the opposite effect than intended. Further, experimental research demonstrating the importance of sources, audience, causal alternatives, and phrasing in the persuasiveness of corrections may aid in the development of more effective and nuanced counter-messaging strategies.

Implications for Policy

The literature on the role of the Internet in promoting extremism and its efficacy in recruiting individuals to extremist organizations, as well as the effectiveness of corrective interventions in changing beliefs, is young and rapidly evolving. Yet a survey of preliminary research on these issues, focusing on the part played by online interactions, suggests the following six recommendations for policymakers tasked with preventing and countering violent extremism.

1. ***Don't focus on online extremism at the expense of community-level engagement.*** Research suggests that in-person interactions remain a necessary, if not always sufficient, component in the radicalization process. Given the nature of Internet search engine algorithms, it is unlikely that individuals will come across extremist content by chance and "self-radicalize." More likely, they are directed to extremist sites and chat rooms by personal contacts or recruiters from extremist groups. A strategy to counter extremist narratives online may be part of the solution, but it should be accompanied by a campaign to address extremist discourses and the presence of recruiters in communities.

2. ***Restrictive measures—like website takedowns and content filtering—are just one tool that must be paired with broader counter-messaging and other interventions.*** Many current P/CVE efforts online focus on removing and blocking access to extremist speech where it is found. While such attempts can impact the "low-hanging fruit" by limiting access to the most easily discovered content, media openness in Western societies and the complexity and rapidity of changes in technology will leave online outlets available to the most dedicated extremists while increasing the costs of enforcement. Such measures must be wielded surgically in support of a wider P/CVE strategy to be effective.

3. ***Tailor the message to the audience.*** Counter-messages that may seem persuasive and reasonable to moderates may lead to a "backfire effect" for those on the fringe. Those holding moderate beliefs are unlikely to join extremist causes; counter-messages that resonate with them are not likely to have a significant impact on the degree of radicalization. However, those with extreme views, who are at-risk of joining extremist causes, will not be persuaded by messages that might be geared more toward moderates. Thus, authors of counter-messages should seek to understand the perspective of those on the fringe, and tailor their messengers and narratives to that audience.

4. ***Provide causal alternatives to false extremist narratives.*** Social science research suggests that people are more likely to be persuaded by corrections that provide a causal alternative to their pre-existing belief. When extremists promote pernicious and false narratives—e.g. that the United States directly targets civilians in Iraq and Syria, for example—it is not enough to simply deny the claim; rather, an accurate causal story must be supplied.

5. ***The messenger matters.*** Research indicates that even moderate populations that reject most extremist ideology will dismiss information emerging from the U.S. Government. This tendency is greater in communities that feel aggrieved by U.S. actions and is universal among those already holding extremist beliefs. U.S. Government efforts must therefore be focused on empowering voices within vulnerable communities that retain influence and respect. An important aspect of such a strategy would be to remove the control of these efforts from the security services, who are often subjects of deep-seated distrust in the most susceptible communities that harbor narratives of victimization.

6. ***Affirm correct information rather than negate incorrect beliefs.*** Some psychological and linguistic research suggests that negations can actually reinforce the idea they attempt discredit. Thus, counter-messages, where possible, should seek to affirm accurate claims and facts rather than negating false narratives or lies. For example, in response to the widely-held view among Islamist extremists that America is at war with Islam, counter-messages should affirm that the U.S. government supports religious freedom rather than simply saying that "America is not at war with Islam."

Conclusions and Areas for Further Research

Given the salience of extremist ideologies in world politics today, it is easy to succumb to an impulse to "do something" to counter extremist narratives. However, as recent research on political beliefs and corrections has shown, well-intentioned efforts to dispel myths and encourage fact-based discourses can actually worsen the problem of misperception. The goal of counter-messages should not be to provide the sender with the satisfaction of making their point, but to persuade those who find extremist arguments compelling. Thus, the overall message of the recommendations contained in this paper is one of caution. Counter-messages should be carefully tailored to consider their audience, source, phrasing, and potential to provoke backlash or to provide a causal alternative to inaccurate beliefs. Foregoing a counter-message may be preferable to a counter-message that comes from a non-credible source or that can easily be explained away as part of a conspiracy theory. Finally, it should be noted that there are no "magic words" in this battle against extremist ideas. No matter how carefully crafted a counter-message or counter-narrative may be, many people will remain unmoved. First and foremost, a counter-messenger should "do no harm."

Bibliography

Aly, Anne. "The Internet as Ideological Battleground." Edith Cowan University Research Online, 2010.

Ashour, Omar. "Online de-radicalization? Countering violent extremist narratives: message, messenger and media strategy." *Perspectives on Terrorism* 4, no. 6: 2011.

Awan, Akil N., "Radicalization on the Internet?" *The RUSI Journal*, Vol. 152, No. 3, 2007, pp.76–81.

Briggs, Rachel, and Sebastien Feve. "Review of programs to counter narratives of violent extremism." *Institute of Strategic Dialogue* (2013).

Conway, Maura and Lisa McInerney. "Jihadi Video and Auto-Radicalisation: Evidence from an Exploratory YouTube Study." In: Ortiz-Arroyo, Daniel, Henrik Legind Larsen, Daniel Dajun Zeng, David Hicks and Gerhard Wagner, eds, *Intelligence and Security Informatics: First European Conference, EuroISI 2008*: *Esbjerg, Denmark, December 2008 – Proceedings*. Berlin: Springer, 2008: 108–118.

Counter-Terrorism Implementation Task Force (CTITF) Working Group on Use of the Internet for Terrorist Purposes. *Riyadh Conference on "Use of the Internet to Counter the Appeal of Extremist Violence: Conference Summary and Follow-up/Recommendations."* UN CTITF Conference Report: 24-26 January 2011.

Counter-Terrorism Implementation Task Force (CTITF) Working Group on Use of the Internet for Terrorist Purposes. *Countering the Use of the Internet for Terrorist Purposes.* UN CTITF Working Group Report: February 2009.

Forest, James. 2010. "Influence Warfare and Modern Terrorism," from *Influence Warfare: How Terrorists and Governments Fight to Shape Perceptions in a War of Ideas* (Westport, CT: Praeger Security International, 2009). http://www.jamesforest.com/wp-content/uploads/2010/05/Forest_GJIA-article-on-IW.pdf.

Friedland, Jamie and Kenneth Rogerson. "How Political and Social Movements Form on the Internet and How They Change Over Time." Washington, DC: Institute for Homeland Security Solutions, 2009.

Gartenstein-Ross, Daveed and Laura Grossman. "Homegrown Terrorists in the U.S. and U.K.: An Empirical Examination of the Radicalization Process." Washington, DC: FDD Press, 2009.

Hussain, Ghaffar and Erin Marie Saltman. "Jihad Trending: A Comprehensive Analysis of Online Extremism and How to Counter it.*"* Quilliam Foundation, 2014.

Koehler, Daniel. "The Radical Online: Individual Radicalization Processes and the Role of the Internet." *Journal for Deradicalization* 1 (2014): 116-134.

Mayo, Ruth et al, "'I am not guilty' vs 'I am innocent': Successful negation may depend on the schema used for its encoding," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20, no. 6 (2004): 1420-1436.

Neumann, Peter R. "Options and Strategies for Countering Online Radicalization in the United States." *Studies in Conflict & Terrorism*, 2013.

Nyhan, Brendan and Jason Reifler. "Why Corrections Fail: The Persistence of Political Misperception," *Political Behavior* 32 (2010): 303-330.

Nyhan, Brendan and Jason Reifler. "Which Corrections Work? Research Results and Practice Recommendations." *The New America Foundation* (Research Paper), October 2013.

O'Rourke, Simon. "Virtual Radicalisation: Challenges for Police." Edith Cowan University Research Online, 2007.

Rogan, Hanna. "Jihadism Online - A study of how al-Qaida and radical Islamist groups use the Internet for terrorist purposes." Norwegian Defence Research Establishment, 2006.

Seifert, Colleen M., "The continued influence of misinformation in memory: What makes a correction effective?" *Psychology of Learning and Motivation: Advances in Research Theory* 41 (2002): 265-292.

Stevens, Tim and Peter Neumann. "Countering Online Radicalisation: A Strategy for Action." *International Centre for the Study of Radicalisation and Political Violence (ICSR)*, 2009.

Von Behr, Ines, Anaïs Reding, Charlie Edwards, and Luke Gribbon. "Radicalisation in the digital era: The use of the Internet in 15 cases of terrorism and extremism." Brussels, BE: RAND Europe, 2013.

Wojcieszak, Magdalena. "Computer-mediated False Consensus: Radical Online Groups, Social Networks and News Media." *Mass Communication and Society*, Vol. 14, No. 4 (2011): 527–546.

--. "'Carrying Online Participation Offline'—Mobilization by Radical Online Groups and Politically Dissimilar Offline Ties." *Journal of Communication*, 59 (2009): 564–586.

York, Erin, and Peter Chalk. "Promoting Online Voices for Countering Violent Extremism." Santa Monica, CA: RAND Corporation, 2013.