The Routledge Companion to Media Disinformation and Populism

Howard Tumber, Silvio Waisbord

Countering hate speech

Publication details

Babak Bahador
Published online on: 24 Mar 2021

How to cite: Babak Bahador. 24 Mar 2021, Countering hate speech from: The Routledge Companion to Media Disinformation and Populism Routledge
Accessed on: 12 Jul 2023

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Counterspeech refers to communication that responds to hate speech in order to reduce it and negate its harmful potential effects. This chapter begins by defining hate speech and examining some of its negative impacts. It then defines and disaggregates the concept of counterspeech, differentiating five of its dimensions – audiences, goals, tactics, messages, and effects. This is presented in two sections. The first examines audiences, goals, and tactics. Audiences refers to the groups exposed to counterspeech, including hate groups, violent extremists, the vulnerable, and the public. Goals are the aims of those engaging in counterspeech efforts, which often vary by audience. Tactics assesses the different means and mediums used to reach these audiences. The second section examines messaging and effects. Messaging refers to the content typologies used to try and influence audiences. Effects analyses how the audiences exposed to counterspeech are influenced based on a review of recent studies in which the approach has been tested. In this section, five key findings from the counterspeech research literature are presented.

Defining hate speech

In its narrow definition, hate speech is based on the assumption that the emotion of hate can be triggered or increased towards certain targets through exposure to particular types of information. The emotion of hate involves an enduring dislike, loss of empathy, and possible desire for harm against those targets (Waltman and Mattheis 2017). Hate speech, however, is usually defined more broadly to include any speech that insults, discriminates, or incites violence against groups that hold immutable commonalities such as a particular ethnicity, nationality, religion, gender, age bracket, or sexual orientation. While hate speech refers to the expression of thoughts in spoken words, it can be over any form of communication, including text, images, videos, and even gestures. The term hate speech is widely used today in legal, political, and popular discourse. However, it has been criticised for its connection to the human emotion of hate and other conceptual ambiguities (Gagliardone et al. 2015; Howard, 2019). This has led to proposals for more precise terms such as dangerous (Benesch 2014; Brown 2016), fear (Buyse 2014) and ignorant speech (Lepoutre 2019).

While jurisdictions differ in the way they define and attempt to remedy the negative consequences of hate speech (Howard 2019), there is some international consensus through institutions such as the United Nations about what constitutes hate speech. According to Article 20 of
the International Covenant on Civil and Political Rights (United Nations n.d.), ‘[a]ny advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law’. Other international legal instruments that address hate speech include the Genocide Convention on the Prevention and Punishment of the Crime of Genocide (1951), the International Convention on the Elimination of All Forms of Racial Discrimination or ICERD (1969) and the Convention on the Elimination of All Forms of Discrimination Against Women or CEDAW (1981).

While these international agreements are important representations of broadly recognised principles on the topic, in practice, much of hate speech today occurs on private social media platforms such as Facebook, Twitter, and Google. It can be argued that, as a result, a new type of jurisdiction has emerged in which the policies or ‘community standards’ of these organisations ultimately set the boundaries between free speech and hate speech (Gagliardone et al. 2015). These policies are ever evolving and are largely modelled on the same core principles enshrined in more traditional national laws and international treaties, with sanctions ranging from flagging hateful posts to removing such posts and closing associated accounts. However, many criticise the notion of relying on profit-making corporations to set such policies for a myriad of reasons including their slow response, inconsistent policy application, inappropriate enforcements (e.g. banning journalists or activists), and favouritism towards the powerful (Laub 2019).

Research on media effects and persuasion demonstrate that hateful messages are likely to have different effects on different members of in-group audiences, often determined by their predispositions. Hate speech, therefore, should be understood as speech that only has the potential to increase hate. Even when hate increases, however, other moral, cultural, political, and legal inhibitions can prevent hateful views from manifesting into behavioural responses such as violence and crime. Factors that can increase the risk of speech leading to violence include the speaker’s influence, audience susceptibility, the medium, and the social/historical context (Benesch 2013; Brown 2016).

While hate speech can target individuals, it is much more concerning when groups are targeted. This is because, in such scenarios, all members of the group can become ‘guilty by association’ and the focus on collective blame and vicarious retribution, even if few were responsible for the purported negative actions (Lickel et al. 2006; Bahador 2012; Bruneau 2018). Hate speech targeting groups is almost always a form of disinformation because rarely is an entire group guilty of the negative actions and characteristics allocated to them. In the vast majority of cases, such allegations are either outright false or exaggerated or conflate the actions of a minority associated with the group to the entire group.

Hate speech is often used by populist leaders and politicians to shift blame for real or perceived societal problems or threats on domestic minorities within societies who have historically been the victims of prejudice and past hate. While one aspect of hate speech involves calls to dehumanise and demonise such groups, another involves incitement to exclude, discriminate, and commit violence against such groups as a solution to overcoming the social problems and threats. This type of speech is also used between nation-states and is a well-known precursor to international conflict to prepare and socially mobilise societies for war and the normalisation of mass violence (Dower 1986; Keen 1991; Carruthers 2011).

To offset the potential negative impacts of hate speech, governments, technology companies, non- and intergovernmental organisations, and experts have often proposed content removal (or take-downs) and other forms of punishment. While there is much criticism of the interpretation and implementation of such sanctions, as mentioned, there is a larger ethical issue at stake as such actions violate the human right to free speech, which is widely considered to be fundamental to a properly functioning democracy. Free speech is protected in the
Countering hate speech

Universal Declaration of Human Rights and within the constitutions of a number of countries, including the First Amendment to the United States Constitution. Within a US context, almost all speech is protected, with limitations only in very rare cases when ‘imminent lawless action’ is advocated (Tucker 2015). Furthermore, there is also the unintended risk of the Streisand effect, in which attempts to hide or censor something inadvertently increase its publicity (Lepoutre 2019).

Counterspeech

To reduce the harmful effects of hate speech without infringing on the freedom of speech, counterspeech is often evoked as a non-regulatory and morally permissible solution (Engida 2015; Strossen 2018; Howard 2019; Lepoutre 2019). Counterspeech is, by definition, communication that directly responds to the creation and dissemination of hate speech with the goal of reducing its harmful effects. Bartlett and Krasodomski-Jones define counterspeech as speech that argues, disagrees with, and presents an opposing view to offensive and extreme online content (2015). However, in practice, efforts under the counterspeech umbrella are sometimes much broader and include communication that is not directly responsive but rather aims to prevent or create the conditions to reduce hate speech and its negative effects (Benesch 2014; Lepoutre 2019). For the purposes of this chapter, analysis will primarily focus on attempts to counter hate speech directly. Furthermore, while counterspeech can occur across a range of online and offline mediums, especially when it is more broadly understood, the focus in this chapter will largely be on counterspeech in an online digital setting.

Before examining the concept of counterspeech more directly, it is important to understand that counterspeech is based on a set of assumptions. The first of these involves an underlying conviction in persuasion and media effects, especially at the individual psychological level. By exposing audiences to particular messaging, counterspeakers hope to achieve attitudinal and behavioural outcomes that will either reduce hate speech or negate its harmful effect. This approach, focused on change at the individual level, is in contrast to structuralists who view concepts like hate and its negative manifestations, such as violence, as functions of social and political structures that enable them to exist (Waltz 1959). For structuralists, hate speech only becomes more widely prevalent and dangerous under particular societal circumstances, so to solve the problem, the structures must be changed, not the speech itself. Structural solutions often focus on the importance of systemic change and often see media literacy and education as enabling ‘digital citizens’ to both identify hate speech and react to it through critical engagement that respects human rights and incorporates a broad set of ethical principles (Gagliardone et al. 2015; Porten–Cheé 2020).

Counterspeech, at its core, is rooted in a set of classical liberal values that assume a public that is both moral and rational in its decision-making, capable of deciphering truth from falsehood with sufficient information in the so-called marketplace of ideas (Schauer 1982; Ingber 1984; Carr 2001; Lombardi 2019). Hatred and prejudice, from this perspective, derive from ignorance and a lack of accurate information about others. This assumption is built into UNESCO’s charter, which blames ‘ignorance of each other’s ways and lives’ as the reason ‘wars begin in the minds of men’ (UNESCO, n.d.). However, through various means including education and communication, this same approach assumes that ‘it is in the minds of men that the defences of peace must be constructed’. Inherent in this remedy is belief in the power of speech and deliberation to have almost therapeutic powers to correct errors and find solutions. They can also be seen as a way to enhance democracy and enable a new form of political participation and even civic duty (Porten–Cheé 2020).
From this perspective, counterspeech represents rational, accurate, and corrective information while hate speech represents irrationality based on falsehoods. These assumptions, of course, are problematic for a range of reasons, including the research findings that show humans to be driven less by rationality and more by emotion and bias in their decision-making. In addition, the so-called marketplace of ideas assumes that the public, and especially those holding opposing views, can meet in open public forums in which ideas can be readily exchanged. But communication in the online world is often marked by echo chambers of like-minded views that reinforce and radicalise polarised positions (Langton 2018).

There is a relatively limited body of research and literature on counterspeech specifically and, compared to the large body of work on hate speech, it can be considered under-researched, undertheorised and underdeveloped, and what is written is often by practitioners who have worked in various ways to operationalise the concept (Howard 2019). However, it is important to note that a number of parallel literatures overlap with it. Three that are particularly significant are the countering violent extremism (CVE)/deradicalisation literature, which aims to examine and/or counter the online and offline efforts of terrorist and extremist groups (Berger and Strathearn 2013; Koehler 2017; Bjola and Pamment 2018); the countering online incivility literature, which aims to understand and/or improve the tone or ‘health’ of communication online (Munger 2017; Tromble n.d.; Rossini 2019; Porten-Cheé 2020); and the countering dis/misinformation literature, which aims to examine and/or reduce and correct false or fake information online (Chan et al. 2017; Porter and Wood 2019).

To deconstruct counterspeech, it is important to examine it from at least five dimensions. While at its core, counterspeech is about messaging and content, there are at least four other dimensions that need examination to fully appreciate the contours of the concept and the effort to operationalise it. These are the audiences receiving it, the goals of practitioners who employ it in relation to these audiences, the tactics used to reach these audiences, and the effects it has in relation to its audiences. The following sections first examine the audiences, goals, and tactics of counterspeech and then its messaging and effects.

**Audiences, goals, and tactics**

Audiences are the different groups who are exposed to hate speech and counterspeech messages but with clear differences in their relationship to these messages. Understanding the differences between audiences is critical for the effectiveness of counterspeech in achieving its goals. In general, counterspeech aims to reduce the likelihood of audiences accepting and spreading hate speech and increase the willingness of audiences to challenge and speak out against such speech (Brown 2016). In this section, four different audiences and associated counterspeech goals are considered. The audiences examined are hate groups, violent extremists, the vulnerable, and the public. Individuals can potentially transition between these groups as a result of exposure to both hate speech and counterspeech, although such shifts often involve other factors beyond message exposure.

While those creating and sharing hateful messages are likely to hold a variety of motivations for their actions, a core goal of such groups is often to grow and strengthen their in-group. The first audience that can be distinguished in this analysis, therefore, is hate groups. In the United States, the Southern Poverty Law Center (n.d.) defines hate groups as organisations with ‘beliefs and practices that attack or malign an entire class of people, typically for their immutable characteristics’. Hate groups are often voluntary social groups that vary in size, strength, and organisation. While some prominent groups, such as the Ku Klux Klan, have had organised and formal structures with thousands of members (at their peak), many other hate groups are diffuse and
Countering hate speech

informally organised around a common ideology. Many white supremacist hate groups in recent years, for example, can be characterised as leaderless, without formal structure, and loosely organised around a common ideology (Berger 2019; Allen 2020). In such cases, there is often no official membership, and other terms indicating affiliation, such as supporters and followers, may be more accurate. The central goal of counterspeakers and counterspeech is to reduce the size and strength of hate groups collectively and to shift the discourse and ultimately the beliefs of the individuals producing hate speech (Benesch 2014).

The internet and social media create affordances to connect individuals with similar grievances to organise much more efficiently, leading many to link the growth in hate groups to the new media ecology. Furthermore, as digital media is increasingly adopted earlier in life and consumes a greater proportion of one's time, messages received in the digital media ecosystem can begin to penetrate young people at an early age when their core beliefs and identities are forming. Within this context, hate groups use a variety of forums to reach disaffected youth and offer them kinship and a sense of belonging they may lack (Kamenetz 2018).

In recent decades, hateful ideology and state power have merged at different times and places to implement discriminatory and violent policies against perceived enemies, leading to mass atrocities and even genocide in worst-case scenarios. Hateful policy against the Tutsi and moderate Hutu in Rwanda in 1994 and against the Rohingya in Myanmar in 2017, for example, led to the extermination of 800,000 in the former case and the ethnic cleansing of 700,000 in the latter. But even outside official power, hate groups often have members willing to operationalise their beliefs with acts of violence against target groups. Hate-driven massacres in 2019 against Muslims in Christchurch, New Zealand, and against Latinos in El Paso, Texas (United States), were, for example, conducted by ‘lone wolves’ who carried out hate-induced acts of terror. This leads to the second audience – violent extremists – who are often sub-groups within hate groups willing to carry out violence in the name of their cause. What turns or radicalises a member or supporter of a hate group to turn into a violent extremist is a matter of much concern and research. A key goal of counterspeech is to understand these triggers and use communication to offset them as much as possible so that hate does not turn into violence.

The final group in this analysis includes everyone else who may inadvertently or intentionally come across hateful rhetoric online. This audience is called the public in this chapter. For this analysis, we exclude members of the group targeted by hate speech and counterspeakers as part of the public and instead only focus on third-party individuals who are otherwise unaffiliated with any group or audience already mentioned. The public will usually be much larger than any of the other audiences. The goal of counterspeakers is to get this audience to pay attention and ideally to engage in constructive interventions as a civic duty to assist counterspeakers in their goals. It is important to note that hate groups are not just a random set of individuals but often claim to represent a larger community with common grievances. These commonalities can be based on immutable factors like the groups targeted for hate speech, creating an ‘us versus them’ dynamic. By highlighting common grievances and identifying targets to blame, hate groups hope to attract support from the larger group they claim to represent. The third audience, therefore, are referred to as the vulnerable – those who have immutable similarities with hate groups and are potential future supporters. Counterspeech aims to prevent this vulnerable audience from joining or supporting hate groups through a number of different messaging strategies that try and limit them to fringe movement within the larger group they hope to represent. When discussing tactics, this chapter refers to the types of observations in which researchers or practitioners are intentionally attempting to employ counterspeech to understand its effects and prevent harm from hate speech, respectively.
To reach these audiences online, Wright et al. suggest four tactics or vectors of counterspeech: one-to-one, in which a single counterspeaker converses with another person sharing hateful messages; one-to-many, in which an individual counterspeaker reaches out to a group that is using a particular hateful term or phrase through, for example, using a hateful hashtag; many-to-one, in which many respond to a particular hateful message that may have gone viral; and finally, many-to-many, in which a conversation involving many breaks out, often over a timely or controversial topic (2017). One example of many-on-many involved the hashtag #KillAllMuslims, which trended on Twitter but was then taken over by counterspeakers who reacted on mass to challenge it, with one particular countermessage shared over 10,000 times (Wright et al. 2017).

When examining counterspeech research, it is important to distinguish amongst the different means by which such activity is observed and understood and the degree of intervention. At the one end of the spectrum are naturalist studies that involve no direct intervention. In these studies, researchers observe organic conversations between hateful speakers and counterspeakers (neither of whom may identify themselves by these labels) through gathering data on real conversations from social media feeds. At the other end of the spectrum are full experimental research studies in which one or both sides are recruited and observed communicating in an artificial environment. However, between these spaces, activists and NGOs sometimes engage in coordinated counterspeech interventions to try and reduce the perceived negative impacts of hate speech. One notable example of this type of work is the activist group #ichbinhier and #jagärhär (German and Swedish for #Iamhere). This group, which operates in a number of countries, engage in a series of activities in this regard, including counterarguing against hateful posts (Porten-Cheé 2020).

The tactical choices for reaching hateful speakers online depend to some degree on which audiences counterspeakers want to influence. In this regard, there are at least three choices. If the goal is to reach hate group members and potential violent extremists in order to change their views and behavior, going into the ‘hornet’s nest’ and finding the spaces where they congregate is an obvious option. This can involve particular website such as 8kun (formerly 8chan) or hate groups on social media platforms, although many have been removed based on recently updated community standard guidelines (Facebook, n.d.). The second tactic involves going to mainstream news websites and social media pages and monitoring these prominent locations for hateful comments, with the goal of catching them early and limiting influence and possible recruitment of vulnerable audiences and the broader public. The third approach involves countering event-driven hate, which can surge during planned events such as elections or unexpectedly after high-profile crimes or acts of terrorism. In such scenarios, emotions can be particularly elevated as the political stakes are high, so hateful rhetoric can spread rapidly amongst not just hate groups but also vulnerable audiences who might share common grievances or prejudices. In such scenarios, counterspeakers can play a calming role and try and prevent hateful rhetoric from turning into violence. This can involve countering hateful hashtags as they emerge following critical events (Wright et al. 2017) or sending preventive ‘peace text’ messages promoting calm during key events or responding to false rumours. These latter activities were employed by the NGO Sisi Ni Amani Kenya during the 2013 Kenyan elections to try and prevent a repeat of the violence that marred the 2007–2008 election (Shah and Brown 2014).

Messages and effects

So far, this chapter has examined the differences between four audiences and the goals of counterspeakers for each. It has also described some tactical issues faced by counterspeakers. This
section now turns to the limited counterspeech research literature and highlights some key findings. It presents these through the lens of the final two counterspeech dimensions – the messaging or content employed by counterspeaker and the effects on audiences, which are presented as five key findings.

1 Changing online norms

As scholars of media and persuasion studies have found over decades of research, it is difficult to change opinions on core beliefs through communication. This is also true for those holding prejudiced and hateful attitudes. Those who are members or supporters of hate groups, and especially those willing to share their views online, are often deeply committed to their positions. As a result, responding with views that challenge their message is unlikely to change their views. But what counterspeech appears to do, especially when it gains momentum through growing numbers of counterspeakers, is to change the online norm and reduce the tendency of people espousing hate to continue the practice in forums with less support and more opposition (Miškolci et al. 2020). For those creating and sharing hateful messages, one attraction of going online and especially of participating in forums with like-minded views, is the support, reinforcement, and sense of community. When this changes, the new environment can become uncomfortable and confrontational and more akin to the offline world in which prejudicial views are largely unpopular. In the offline world, the ‘spiral of silence’ theory posits that people tend to avoid sharing their opinions when such views are perceived to be unpopular due to fear of social isolation (Noelle-Neumann 1974). In a similar manner, it appears that some notable percentage of those engaging in hate speech online disengage when they sense a change in general sentiment in the online forum in which they were participating (Miškolci et al. 2020).

Furthermore, when counterspeech is introduced, others who were not part of the counterspeech efforts but are likely to oppose bigotry and hate are more likely to engage and post comment in line with the counterspeakers (Foxman and Wolf 2013; Miškolci et al. 2020; Costello et al. 2017). This may include members of the public who come across online forums with a mix of hate speech and counterspeech. It may also include vulnerable audiences who may be on the fence but ultimately swayed in the direction of what appears to be the more popular sentiment. Various studies show that those entering online spaces are influenced by the norms already present (Kramer et al. 2014; Cheng et al. 2017; Kwon and Gruzd 2017; Molina and Jennings 2018). This has been described as ‘emotional contagion’ in which exposure to the volume of positive or negative expression results in posts in the same emotive direction (Kramer et al. 2014).

While counterspeech, as mentioned earlier, generally involves a response to existing hate speech, there is some concern that reacting to hate speech through counterspeech might draw attention to the original message and inadvertently amplify it. To prevent this, pre-emptive counterspeech has been proposed to condition the conversation context in a way that disables possible future hate speech (Lepoutre 2019). This could involve interventions before critical events such as forums known to draw hate speech.

2 Constructive approaches are more effective

Research on the nature of counterspeech finds that ‘constructive’ communication is more effective at garnering engagement than disparaging responses that involve name calling and insulting hateful speakers (Benesch et al. 2016) and attempting to invoke shame or combativeness (Bruneau et al. 2018). On the one hand, a constructive approach is about tone, so casual
and sentimental tone and the use of humour when appropriate can disarm the serious nature of hateful rhetoric, making those espousing it open up and feel more comfortable engaging. On the other hand, constructive approaches include particular types of content such as personal stories and calling attention to the negative consequences of hate speech (Bartlett and Krasodomski-Jones 2015; Frenett and Dow 2015; Benesch et al. 2016). A constructive approach requires a combination of critical thinking and ethical reflexivity to understand the context and underlying assumptions, biases, and prejudices of hateful speakers in formulating effective responses (Gagliardone et al. 2015).

3 **Self-reflective and hypocrisy messaging are particularly effective**

When it comes to counterspeech content, it is often assumed that messages that rehumanise targeted groups are most effective (Bahador 2015). This can involve individualising the target group members by challenging assumptions of homogeneity or countering stereotypes by showing, for example, that out-groups thought to ‘hate us’ actually have affection for our side (Bruneau et al. 2018). However, research by Bruneau et al. finds that the most effective messaging for reducing collective blame and hostility towards an out-group, in their study looking at anti-Muslim sentiment in the US, involved messaging that exposed the in-group’s hypocrisy (Bruneau et al. 2018). This was done through an ‘intervention tournament’ in which subjects were randomly assigned to ten groups (with nine treatments involving watching a video with different messages and one control group). The ‘winning’ intervention involved collective blame hypocrisy messaging, which highlighted hypocrisy as individuals blamed Muslims collectively for terrorist acts committed by individual group members but not white Americans/Christians for similar acts committed by individual members. Exposure to this type of messaging resulted in reductions in collective blame of Muslims and anti-Muslim attitudes and behavior (Bruneau et al. 2018). This finding appears to lead to change because it triggers cognitive dissonance, in which individuals first support a position and then are made aware that they advocate for action that contradicts that position. One way to redress this dissonance, therefore, is to change one’s prejudiced position to create cognitive consistency (Festinger 1962; Aronson et al. 1991; Bruneau et al. 2018).

4 **Source credibility matters**

The importance of source credibility in persuasion is not unique to counterspeech and has been advocated as far back as at least Aristotle, who saw the credibility of the speaker to the audience (‘ethos’) as a central component of effective rhetoric. A key part of the critical thinking needed to construct an effective counterspeech message campaign, therefore, involves having messengers who are credible to the targeted audiences one seeks to influence (Briggs and Feve 2013; Brown 2016; Munger 2017). In many cases, this requires a deep understanding of the local context as hate speech is only impactful and dangerous within particular contexts, and different speakers also hold different levels of credibility in different locations. In one study of white nationalists, for example, it was found that more response was garnered when the speaker was conservative, suggesting it was someone with some similarities to them (Briggs and Feve 2013).

5 **Fact-checking can moderate views**

As mentioned earlier in this chapter, hate speech against groups is almost always based on dis- or misinformation. That is because the entire groups targeted for hate are almost never wholly
Countering hate speech

guilty of the purported negative actions or characteristics allocated to them. To remedy various
types of misinformation, including hate speech, counterspeech based on fact-checking appears
to be an obvious solution. While there is some concern that under some circumstances, chal-
lenging the views of those holding misinformed hateful views can bring more salience to them
(Lepoutre 2019) or even backfire and strengthen them, there is growing evidence that such cases
are rare, and fact-checking, in fact, makes people’s beliefs more specific and factually accurate
(Porter and Wood 2019; Nyhan et al. 2019). As a result, fact-checking could be helpful amongst
the arsenal of other tools used in crafting counterspeech messages. More than likely, as other
research has shown, those with deeply held beliefs will be less likely to moderate their views
based on fact-based corrections. However, others who are not as deeply committed, or who
hold other values that contradict the underlying values of hate speech, may be susceptible to
being affected by such counter-messages.

Conclusion

This chapter has defined hate speech and counterspeech and examined the latter concept
through a review of five factors that help disaggregate it – audiences, goals, tactics, messaging,
and effects. Through an examination of emerging counterspeech research and practice, the
chapter identifies five key findings that show that counterspeech can have some effect on audi-
ences under particular circumstances. While these findings are novel within this context, they
tap into research findings on media effects and persuasion that are already well established, such
as the spiral of silence theory and cognitive dissonance. These findings are particularly relevant
today as there is growing social and political concern over the role of hate speech in individual
and collective acts of violence from New Zealand to Myanmar. For practitioners seeking to
develop new programmes to counter hate speech without infringing on free speech rights,
engaging with this new body of research can be particularly helpful, and this chapter has aimed
to highlight some of the latest findings in this regard.

References

Aronson, E., Fried, C., and Stone, J. (1991) Overcoming Denial and Increasing the Intention to Use Con-
Bahador, B. (2012) Rehumanizing Enemy Images: Media Framing from War to Peace. In K.V. Korostelina
New York: Palgrave Macmillan.
Bartlett, J., and Krasodomski-Jones, A. (2015) Counter-speech: Examining Content that Chal-
pdf (accessed December 18, 2018).
Counterspeech. Dangerous Speech Project. At: https://dangerousspeech.org/considerations-for-
successful-counterspeech/ (accessed November 29, 2019).


Countering hate speech


