Deconstructing Hate Speech in the DRC:
A Psychological Media Sensitization Campaign

Johanna Vollhardt[1], Marie Coutin[2], Ervin Staub[1],
George Weiss[2], and Johan Deflander[2]

[1] University of Massachusetts, Amherst
[2] Stichting Radio La Benevolencija/Humanitarian Tools Foundation

Correspondence concerning this article should be addressed to Johanna
Vollhardt, Department of Psychology, University of Massachusetts at
Amherst, 135 Hicks Way, Tobin Hall, Amherst MA 01003. E-mail:
jvollhar@psych.umass.edu.

ABSTRACT

This article describes an interdisciplinary and theory-based radio cam-
paign that has been developed to counteract, and sensitize citizens to hate
speech in the Democratic Republic of the Congo (DRC). The article pro-
vides a brief overview of the instrumentalization of hate speech and the
violent effects it has had in the Great Lakes region of Africa. A summary
of the most recent events in the DRC is given. Here, hate speech was
used in the presidential election campaigns in 2006, contributing to a
polarization of the country and giving the campaign an ethnic underpin-
ning. A radio program developed to counteract hate speech during the
election campaigns is described. Its theoretical basis, the application of
Staub's (1989) theory of the evolution of mass violence to hate speech, is
presented. Based on this and other relevant psychological concepts, char-
acteristics and psychological aspects of hate speech are summarized, and
markers and guidelines are provided that allow listeners to detect and
counteract hate speech.

I. Hate Speech in the Great Lakes Region of Africa: Historical Background and Recent Events in the Democratic Republic of Congo (DRC)

Over the last nearly fifteen years, the media—and radio broadcasting
in particular—has been instrumental in instigating and sustaining rebel wars
and ethnic conflicts in the Great Lakes region of Africa, especially in
Burundi, Rwanda, and the eastern parts of the Democratic Republic of
Congo. In this region, the media has become an effective tool in propagat-

ing hatred and ethnic divisions, thereby increasing existing tensions between and within the countries by reinforcing nationalistic sentiments, or heightening and politicizing ethnic identities. Commonly known as hate radio stations, these media are identified as "encouraging violent activities, tension or hatred between races, ethnic or social groups, or countries for political goals and/or to foster conflict by offering one-sided and biased views and opinions, and resorting to deception" (Radio Netherlands Media Network). They disseminate hate speech, which is characterized by unsubstantiated and intense negative views of and expressions directed toward a group or a group's representative (Tsesis 2002). A central characteristic of hate speech is that it is dehumanizing and defaming, and is used to denigrate and harm the target (ibid.; also see definition below).

The most drastic and well-known example of hate speech communicated through the media that has brought world-wide attention to the phenomenon in this region was disseminated by the radio station Radio Télévision Mille Collines (RTLM) during the genocide in Rwanda in 1994. The broadcasts of this government-owned station, which incited the Hutu majority to murder Tutsis and opponents of the regime, are commonly recognized as having played a major role in this genocide (des Forges 1999; Gourevitch 1998). Furthermore, hate radio has continued to exist even after the genocide, and has resulted in further conflict across borders. After Rwandan Patriotic Front troops succeeded in driving the genocidal government forces out of the capital of Rwanda in July 1994, RTLM used mobile radio transmitters to broadcast disinformation from inside the French-controlled zone on the border between Rwanda and Zaire (now DRC). This caused millions of Hutus to flee toward refugee camps, where they could be recruited as freedom fighters. Hutu extremists then began to stage raids into Rwanda from Congolese territory. Thereby, the ground for future conflict and war between Rwanda and the DRC was created, which discord continues to have a negative impact on the relations between the two countries and the lives of people (Gourevitch 1999; Nzongola-Ntalaja 2003).

In sum, hate speech and hate media have a historical and symbolic meaning in the region, and have played a crucial role in violent conflicts both between and within states. Within the DRC, the effects of hate speech have not been as drastic as in its neighboring country Rwanda. Nevertheless, hate media has been and is very present here as well, and continues to play a destructive role in the political events of the country.

During the civil war in the Congo, which took place in 1998-2003 and caused four million deaths due to fighting and disease, hate propaganda was used, for example, in the eastern part of the country. Here it fueled ethnic conflict, for example between Hema and Lendu, two ethnic groups in the Ituri region. Based on the classification of dehumanizing speech as one of

eight stages leading to genocide (Genocide Watch a), the occurrence of hate speech was among the factors that led international NGOs to warn against a potential genocide in the region (Genocide Watch b).

However, even since the official end of the war, the media in the DRC rarely has been neutral or objective in presenting events. Since many newspapers, radio and TV stations are owned by presidential candidates or their associates, news tends to be strongly biased. The media often resort to personal attacks against political opponents and engage in ethnically charged hate speech (International Crisis Group).

This tendency in the Congolese media has peaked in the recent campaigns for the first free presidential elections in the country in 2006. Strengthened through media campaigns and public claims of opinion leaders, a schism between the eastern and western part of the country has increased. Language has been the major line of division, symbolizing the classification and distinction between the two major political opponents, and linking it to ethnicity and origin. Specifically, presidential candidate Bemba, who speaks Lingala, has been supported primarily in the capital, Kinshasa, and western DRC, where this language is spoken by the majority of the population. Conversely, candidate Kabila speaks Swahili, and has been supported in the eastern part of the country, where this language is more common. In addition, Kabila grew up in exile in Tanzania, and is the son of previous president Laurent Kabila, who marched into Kinshasa in 1996, backed by Ugandan and Rwandan forces. This background has led his political opponents to denounce him as a foreigner, a Rwandan or Tanzanian. Moreover, his followers, mainly Swahili-speaking Congolese in the east, are considered foreigners as well by many in the western part of the country. Given the recent history of the civil war that was started by an invasion of eight foreign nations (including Rwanda and Uganda), these accusations are particularly loaded with negative associations, and prone to create hatred and fear (ISN Security Watch). Kabila's opponent Bemba has capitalized on this, running his campaign with the slogan "100 percent Congolese." Telling citizens, for example, to "vote for the chicken and not for the bird, as the bird will fly away" (speech by Bemba, cited in International Crisis Group, 5), he has repeatedly alluded to Kabila's alleged foreign nationality. A nationalistic and xenophobic ideology has developed that is referred to as "Congolité." Bemba's attacks have been countered by derogatory speeches against Bemba made by Kabila's supporters, who claim that Bemba was born as a result of extramarital relations between his mother and the former dictator Mobutu.

In response to the dangerous presence of hate speech in the Congo, the High Media Authority (Haute Autorité des Médias, abbreviated HAM) was created as one of the "institutions of support for the transition" under the

2002 peace accords by the UN, comprising representatives from all parties in the transition government. The HAM is a public institution that has the authority to impose sanctions on media outlets for the use of political and ethnic propaganda and other unethical journalistic practices (see Committee to Protect Journalists a, b). During the election campaign, several media outlets (including those owned by Kabila and Bemba, and the governmental Radio Television Nationale Congolaise) were sanctioned for inappropriate programming and hate speech, some radio stations were banned from airwaves, and the most offensive ads were removed from radio and TV (see reports in Human Rights Watch; International Crisis Group; Monuc).

However, these attempts have been only partially successful in banning hate speech in the DRC, and in response have even led to violent rallies of Bemba supporters in front of the HAM headquarters on July 27, 2006, during which offices were destroyed and, allegedly, two police officers were burned alive (MSNBC Newsweek). These images were later used by television stations close to Kabila for repeated and emotionally charged broadcastings. A station owned by Bemba broadcasted interviews encouraging attacks on the head of the electoral commission; it also showed pictures of dead bodies following actions by Kabila's troops and of the bombing of populations in the Equator province during the war (International Crisis Group). This and similar hate propaganda was considered a contributing factor to violence that broke out in Kinshasa after the first round of the elections in August 2006. The violence was preceded by rumors that had been spread by the media, claiming that Kabila had won the election, attempting to avoid the second round of elections which had to be held since neither of the candidates had the required absolute majority. During three days of riots more than twenty-four people died, and Bemba was attacked during a meeting with the UN and ambassadors by militias allied with Kabila (ISN Security Watch).

In sum, images and words used in the media to defame and generate hostility continue to pose a serious security threat in the DRC. The prevention of hate speech has been identified as essential in order to prevent further violence and achieve security and stability in the Congo (International Crisis Group). However, as these examples also have shown, even a structure that has been created, such as an official organization dedicated to the banning of hate speech like the HAM, is "often unable to prevent politicians from manipulating the press. It lacks power" (International Crisis Group, 6). Therefore, it is crucial to complement structural and political interventions against hate speech with interventions on the individual level, including psychological campaigns that sensitize citizens and create resistance to hate speech. Campaigns that focus on the individual citizen can also take into account that "the unfairness of the media here was not particularly to one

candidate or the other, it was to the public. . . . the people who suffer the most in this case are the people of Congo who are supposed to make a clear decision" (Voice of America News).

Human beings have the need for self-determination (Deci and Ryan 2000). Therefore, when people are aware of manipulation attempts, most will have the desire to resist them. Furthermore, many citizens in war-torn countries have a strong wish for positive change. In campaigns against hate speech, it is crucial to provide citizens with media literacy and tools for a critical assessment of political broadcasts that empower them and enable them to analyze, detect, and deconstruct hate speech in the media.

II.   COUNTERACTING HATE SPEECH WITH PEACE MEDIA

In considering methods to provide sensitization against hate speech for citizens, it seems particularly powerful to counteract hate speech the way it is disseminated–through the media. There is reason to assume that a tool powerful enough to facilitate genocide has the potential to be an instrument of positive change as well. This idea has been captured in the concept of "peace media" (for a definition see Radio Netherlands Media Network). Media can support peaceful motivations. It can address people's fears constructively. It can contribute to reconciliation by educating the population about the roots of violence and its prevention. The devastating effects of hate media can also be transformed by enhancing a spirit of "never again" among audiences who have suffered from past violence. Furthermore, specific knowledge and media literacy for societies in conflict can be provided, such as analytic tools to detect and counteract hate speech in its early stages.

Realizing this potential, the Dutch NGO Stichting Radio La Benevolencija/Humanitarian Tools Foundation (La Benevolencija 2005a, b), directed by George Weiss and in collaboration with psychologists Ervin Staub and Laurie Anne Pearlman, started a large-scale media campaign in Rwanda in 2003. The campaign consisted of several different reconciliation radio programs (see Staub and Pearlman 2006; Staub 2006). The programs were based on an innovative combination of an approach to healing, reconciliation, and the prevention of new violence that has been developed and implemented in Rwanda by Staub and Pearlman (see Staub et al., 2003, 2005; Staub and Pearlman 2001, 2006), and "edutainment" (education and entertainment) methodologies developed at the Johns Hopkins Center for Communication Programs, which had hitherto been used primarily for health education and behavior change campaigns (see de Fossard 1996; John Hopkins University Center for Communication Programs). After an evaluation showed measurable positive effects of the combined approach

(Levy-Paluck 2006), these programs were later extended to Burundi and the DRC. In this article we will focus on the programs that have been added in 2006 to address the issue of hate speech during the elections in the DRC.

A.   Background and Objectives of La Benevolencija's Media Campaign against Hate Speech in the Great Lakes Region of Africa

The central goal of the media campaign is the empowerment of groups and individuals that have been the target of hate speech and ensuing acts of violence, including extreme violence such as the genocide in Rwanda. Based on a psychological theory of group violence developed by Staub (1989) and described below, one central assumption of the media campaign is that understanding the roots of violence will enhance violence prevention and reconciliation. The goals include healing from the complex trauma (Pearlman and Saakvitne 2005) that such violence creates and promoting justice processes in post-conflict societies (Staub 2004; Staub and Pearlman 2006; Staub et al. 2005).

Staub (1989) has proposed a multicausal theory of the roots of mass violence that emphasizes the interplay of psychological and group processes as well as facilitating societal, structural, and cultural conditions. In other words, rather than concentrating on individual characteristics that predispose to aggression and violence, the theory takes into account a number of external, situational factors that can be influenced and changed. Staub's theory focuses on the psychological effects of objective conditions in a society, that is how individuals and groups react to economic, historical, and political events. A central tenet of the theory is that violence evolves gradually. Mass violence does not erupt suddenly, but is preceded by less obvious acts of derogation, exclusion, scapegoating, and other elements of destructive ideologies which often start out in a rather subtle manner. However, these steps prepare individuals to engage in more direct and physical forms of violence, which become more and more severe as smaller acts of violence lower the threshold for following, more extreme violence. By describing how violence evolves gradually, the theory provides indicators of early stages that can be counteracted in order to prevent mass violence.

Specifically, Staub (1989) has suggested that often difficult life conditions and sudden changes–such as economic deterioration, societal chaos, group conflict, and war–are among the factors that frustrate basic human needs for security, control, a positive identity, connection to others, and understanding of one's world and one's own place in the world. When these psychological needs are not fulfilled constructively and within the current societal arrangements, individuals will seek to fulfill them in alternative ways that are often destructive. In other words, under such conditions they

will be more likely to turn to sources and alleged solutions that give rise to violence. For example, individuals will align themselves with a group in order to fulfill the need for security, and with leaders who promise immediate solutions to their problems. In the quest for solutions and in order to fulfill the need for understanding what is happening with and around them, individuals and groups will become susceptible to scapegoating. This provides a simple interpretation of the existing troubles. For the same reason, destructive ideologies become attractive. They offer a vision of a better, positive future for their own group; however, they do so by eliminating the scapegoat or enemy. In the course of the gradual development from blaming to violence, individuals justify their actions by becoming increasingly extreme in their devaluing attitudes. This allows them to maintain a positive image of themselves as moral and good human beings, while the others are seen as evil. Dehumanization also justifies and facilitates further harmdoing. The increasingly negative views of the victims give rise to greater violence on a continuum of destruction. Along the way, the passivity of bystanders (i.e. members of society witnessing the extreme and violent development) allows perpetrators to carry out their harmful acts without repercussions and reinforces them in their views and actions. Passive bystanders themselves change, and distance themselves from victims, sometimes joining the perpetrators.

Staub (1989) has also identified conditions and characteristics of society that facilitate the evolution to extreme forms of violence. Among these are a history of discrimination against the targeted group, unhealed psychological wounds from collective trauma experienced by the perpetrator group in the past, recently experienced violence-induced trauma, the presence of war, and an authoritarian, monolithic political system that lacks pluralism and excludes devalued groups from public discussion.

On the basis of these theories, as a tool in support of reconciliation and justice processes in the Great Lakes region, radio programs have been developed and carried out in Rwanda, eastern DRC, and Burundi that employ "edutainment" techniques to provide psycho-education on these topics. Through a number of fictional soap operas as well as factual and discussion programs (see La Benevolecija 2005 a for a description), citizens in these vulnerable societies are provided with information about how to resist manipulation to violence, how to intervene and act as positive bystanders in the presence of violence and injustice (Staub 1989), and how to cope with trauma in post-conflict societies (Staub and Pearlman 2006). In sum, the focus of the media campaign is on reconciliation after violent conflict, as well as prevention of future violence.

The project is based on the collaboration of an international team of psychologists and communication professionals, assembled by the organi-

zation La Benevolencija, who train and then work with local writers and journalists to implement the goals in radio programs based on the psychological theory described above. The programs are broadcasted in local languages, Kinyarwanda (in Rwanda and Burundi) and Kiswahli (in the DRC). All programs are edited not only by journalists and communication experts, but also by a team of academic consultants who give feedback on the scripts from the perspective of the underlying psychological theories. These additional steps ensure the theory-based approach and quality of the programs. In addition, an extensive and rigorous impact evaluation of the entire campaign is conducted on a regular basis, and has revealed initial positive results (Levy-Paluck 2006).

B.   *Counteracting Hate Speech: Hate Speech Sensitization Programs in the DRC*

In Rwanda, reconciliation radio programs have immense historical and symbolic significance. Only a decade after about one million people were murdered during the genocide, the same technical support tool that was an instrument of the genocide, the radio, is now being used to counteract violence and to promote a culture of peace and reconciliation.

Although in the DRC the historical context of hate speech is far less drastic, the dangers are nevertheless present. Furthermore, given the current political tensions and use of hate speech as described above, radio programs counteracting it have an important and urgent function. Therefore, the most recent expansion of the media campaign in the eastern DRC has been the broadcasting of a series of programs explicitly focusing on hate speech. The first of these programs were broadcasted in October 2006 on Radio Okapi, a local peace radio station set up by the UN. The immediate, short-term goal of these programs was to intervene directly against hate speech before the second round of elections on October 29, 2006, providing Congolese listeners with knowledge to analyze and deconstruct the hate speech they were being exposed to during the election campaign. Specifically, a series of four weekly programs was broadcasted over Radio Okapi all over the country throughout the month of the elections. In these live programs, questions gathered through brief interviews with listeners all around the country ("vox-pops") were answered by experts of the La Benevolencija team. These questions addressed, primarily, (1) the definition of hate speech and the markers that can be used to detect it and distinguish it from more neutral speech; (2) the role of politicians, the media, and citizens in developing and counteracting hate speech; and (3) what Congolese citizens can do to resist and counteract hate speech. The answers to these questions were based on an extension of the existing psychological theory of group violence that has

been used in the programs so far and is described above (Staub 1989), as well as on other psychological theories relevant to hate speech. Common underlying mechanisms and psychological processes were outlined, and the use of hate speech was placed on the continuum of violence, as delineated below. Overall, the primary goal was not the communication of mere theoretical analysis, but instead of practical knowledge that increases competence in detecting, deconstructing, and counteracting hate speech.

In addition to furthering the short-term goal of counteracting hate speech during the election campaign with the series of programs broadcasted weekly in October 2006, La Benevolencija has also planned a long-term media campaign throughout the country to educate citizens and warn against the dangers of incitement to violence. Specifically, in cooperation with Radio Okapi, a year-long series of factual programs about the psychology of incitement to violence is planned, based on the theory of how violence evolves that has been described earlier. Furthermore, for mid-2007 a similar, year-long series of event programs on major Congolese radio and television channels is planned, with the goal of reaching a majority of the audience throughout the country, including the leadership and the intellectual elite. All of these programs include psychoeducation on the elements and effects of hate speech as a crucial step on the continuum of violence.

III. The Roots and Functions of Hate Speech on the Continuum of Violence

A more general definition of hate speech characterizes it as "any form of expression directed at objects of prejudice that perpetrators use to wound and denigrate its recipient" (Bockmann and Turpin-Petrosino 2002). Hate speech expresses irrational, unsubstantiated, and unjustified antagonism toward a group or a representative of a group, frequently entailing consistently disapproving, hypercritical, and reiterated generalizations (Tsesis 2002). In other words, the basis of denigration is often an actual or perceived difference between the speaker and the target of his/her hate speech (Bockman and Turpin-Petrosino 2002), a difference that compels the speaker to draw a sharp distinction between "us" and "them." Based on these differences, members of outside groups are delegitimized, demonized, or depicted as inferior.

Despite differences in the context, intensity, and specific manifestations of hate speech, general patterns can be observed that apply across cultures and contexts. This knowledge can be the starting point for overcoming the destructive force of hate speech. In working with victims of ethnic violence, Staub and colleagues (e.g. Staub et al. 2003, 2005) have noted that it has significant meaning, and can be transformative, for survi-

vors to learn about similar occurrences in different parts of the world. Information about shared experiences and universal human processes increases the comprehension of the events and can create the hope that a solution for this shared problem, as well as future prevention, is possible (ibid.). Therefore, we chose to illustrate the structures and dynamics of hate speech with a case from a region and historical period different from those discussed so far, in an attempt to exemplify commonalities and parallels.

A less known historical example of the construction of differences in hate speech is the antisemitic campaign that occurred in Poland in 1968. During this campaign, which was largely carried out in the media, Polish citizens of Jewish origin (who had lived in the country for several hundred years and were well integrated in society, frequently actively participating in the political system) were referred to as Zionists and were accused of being disloyal to the Polish state, of having a greater affiliation with Israel than with Poland, and of being a "fifth column" in the country (Stola 2005). In other words, they were depicted as foreigners, implying that they did not belong to Polish society. Portraying individuals or groups as foreigners, for example by tracing back names, origin, and family histories, is one way of constructing negative images and views of the other that have far-reaching consequences.

While the presentation of the other as foreign is often far-fetched or completely untrue, it stigmatizes and places the individual or the groups labeled this way in the category of an outgroup member. This gives rise to all the negative consequences of social categorization (Turner 1987; Tajfel and Turner 1986) that have been demonstrated in the research literature in social psychology, such as increased stereotyping, decreased sharing of resources, and other manifestations of discrimination (see e.g. Brewer 1979; Gaertner and Dovidio 2000). Statements alleging disloyalty of the "other" can create mistrust against the stigmatized individual or group, and even incite irrational fear of the ostensibly subversive group within one's society.

In the political realm, the label "foreigner" is frequently used to discredit opponents and delegitimize political participation. This is among the central functions of hate speech, which is often an instrument to influence and persuade audiences, with the goal of maintaining or gaining political power (Tsesis 2002). For example, one of the main motivations behind the antisemitic media campaign in Poland in 1968 was to gain power in the rivalry between opposing fractions within Poland's ruling Communist Party. Following an intense phase of hate speech in the media, this was achieved by a purge of Polish Jews from their jobs and party positions in the course of the campaign (Stola 2005; Wolak 2004). However, while hate speech often has such instrumental use and value to those who practice it, it can also be an expression of deeply held negative views and feelings toward

another. Such negative views and feelings intensify in the course of the evolution that has been described.

In order to manipulate the audience with hate speech, gain support, and achieve the goal of political power, the speaker needs to communicate efficiently with his or her audience. Such communication is facilitated by drawing on concepts and images that are familiar to the audience and embedded in society and culture. Therefore, hate speech often builds on existing stereotypes, societal beliefs, cultural meanings, and other preconceptions about the targeted groups (Tsesis 2002). In Poland, for example, antisemitism had a long history, and was shared by large parts of the society, at least to some extent (Wolak 2004). Powerful stereotypes about Jews were widespread. For example, the neologism *Żydokomuna* ("Jewish Communism") captured the belief that Jews were responsible for the introduction of the unpopular communist system (Szaynok 2005).

Citizens are more likely to offer support when they are provided with simple solutions to problems that concern them personally. Therefore, audiences are particularly vulnerable to hate speech in times of societal chaos and difficult life conditions, which is, as described above and according to Staub (1989), one important starting point of the evolution of violence. The simple solutions that are offered to the listeners often consist of blaming a group or political opponent for the existing problems, implying that once the offender is removed from power and without political influence, the problems will cease to exist. Intense blaming and scapegoating are therefore further frequent characteristics of hate speech. By offering these simple solutions, the speaker provides listeners with hope and fulfills the need for security. As a result, political support for the speaker is likely to increase. If the speaker is in power at the time of the existing problems, blaming another group also shifts responsibility for these conditions and problems in society, so that further justifications are not needed.

Hate speech often begins in a subtle and hardly noticeable manner. This is important because a too-sudden onset of intense communication of hatred is likely to alienate many listeners. Therefore, perpetrators of hate speech frequently start by using existing stereotypes and widespread beliefs that appear acceptable in a given society and are questioned by only a few.

The degree of derogation gradually becomes more extreme. Using hate speech in public, for example in the media, is essential for eliciting broad societal support for exclusionary and destructive ideologies (Tsesis 2002). Through continuous repetition, certain messages become acceptable, a normal part of everyday life. In this way, listeners habituate to dehumanizing language and destructive ideologies. This allows for a gradual increase of the extremity of hate speech, and constitutes a progression on the continuum of destruction as behavior is carried out that justifies the expressed

JOURNAL OF HATE STUDIES   [Vol. 5:15

beliefs and vice versa (Staub 1989, 2005). Throughout all historical cases, these structures and dynamics of hate speech are, in principle, comparable and therefore universal. Differences exist, of course, in the extent and scope of the resulting violence. In Poland, the primary effect was structural, leading to the mass emigration of almost the entire Jewish population that had remained in the country after the Holocaust (Stola 2005; Wolak 2004). As mentioned earlier, in the Great Lakes Region of Africa hate speech has been an important element in moving the society to direct, intense violence at the very end of the continuum–genocide. This most extreme form of violence is regularly preceded by dehumanization and hate speech (Chalk and Jonassohn 1990; Genocide Watch). Along the way, less visible but very harmful effects such as intense psychological distress among members of the targeted group can occur, even predicting increased suicide rates in targets of hate speech (Mullen and Smyth 2004).

To summarize, hate speech plays a crucial role in the evolution of violence and the steps leading to destruction. Hate speech entails devaluing, dehumanizing, and scapegoating the target, who is frequently used by leaders to explain difficult life conditions and gain support. Hate speech transports destructive ideologies. For listeners exposed to hate speech on a regular basis, for example through mass media, these destructive ideologies and dehumanizing messages become normal. The habituation allows a gradual increase in the extremity of expressed hatred, preparing for violence that can lead from structural and political exclusion all the way to genocide.

Hate speech is, of course, only one of a number of important influences that act together to move a society to significant violence against another group. Among other influential factors are lesser forms of direct or structural violence, which change people and facilitate the perpetration of more severe violence (Staub 1989; 2005). However, hate speech has a particularly important role in generating and increasing feelings of hate in large groups.

IV.   DETECTING AND DECONSTRUCTING SUBTLE FORMS OF HATE SPEECH: EARLY WARNING SIGNALS

While blatant hate speech is easy to identify, it is more difficult to detect subtler forms. However, the use of one of these less overt forms can be one of the first steps along a continuum of destruction as described above (Staub 1989). It is therefore crucial to know how to detect not only obvious expressions of hatred on a more advanced stage of the continuum of violence, but also subtler forms of hate speech. This allows it to be counteracted at an early stage, and prevents its gradual increase and normalization, which would otherwise result in violence of different forms and

intensity. This more detailed knowledge is also important as research has shown that even a simple sensitization to the practice of hate speech can decrease support for its use (Cowan et al. 2002).

Ideally, this sensitization should entail making oneself familiar with a broad range of the characteristics of hate speech. For the purpose of this article, we focus on two aspects of hate speech, and introduce a relevant concept and related sets of criteria for each. These criteria can be applied in order to analyze and identify hate speech in early stages. They describe (a) different forms by which dehumanization occurs, and (b) standards describing violations of argumentative integrity in communication.

A.   *Detecting Subtle Forms of Dehumanization*

Dehumanization is a frequent element of hate speech, and is used to vilify the target (Genocide Watch). It also has the effect of disinhibiting violence (Bandura 1990; Bandura et al. 1975). Thus, dehumanizing descriptions of individuals and groups are an alarming signal of hate speech and the danger of future violence, and its early detection is crucial. Dehumanization can take on blatant forms, such as labeling other groups or individuals with animal names. During the genocide in Rwanda, for example, Tutsis were labeled as *inyenzi*, "coakroaches" (see des Forges 1999). However, dehumanization can also play out in various other, subtler ways.

Haslam (2006) has developed a classification of different forms of dehumanization, and has made the distinction between animalistic and mechanistic dehumanization. The first category is present when individuals or groups are denied characteristics that constitute human uniqueness. These include civility, refinement, moral sensibility, rationality or logic, and maturity. Accordingly, we can speak of "animalistic dehumanization" when others are labeled with any of the following characteristics: a lack of culture, coarseness, amorality or lack of self-restraint, irrationality, predominance of instincts, or childlikeness (p. 258). A second kind of dehumanization occurs when individuals or groups are denied characteristics that constitute human nature, such as emotional responsiveness, interpersonal warmth, cognitive openness, agency or individuality, and depth. Thus, "mechanistic dehumanization" is present when others are labeled as inert, cold, rigid, passive, fungible, or superficial (p. 258).

As this analysis shows, dehumanization occurs in ways that are subtler and more surprising than one might commonly think. However, once an individual or group is referred to repeatedly or labeled publicly in one of these ways, the inhibition against and threshold for using more extreme forms of dehumanizing hate speech are both lowered.

B.   *Detecting Violations of Argumentative Integrity in Communication*

Another central aspect of hate speech is that it is simplistic and unbalanced communication. It violates standards of argumentative integrity by using speech acts that cannot be perceived as fair according to implicit norms of communication. These norms can be represented by 11 standards of (un)fair argumentation that have been delineated by Schreier and colleagues (1995). We will present the six that seem particularly relevant to hate speech.

Specifically, those that involve a direct reference to other individuals overlap with derogating and dehumanizing speech acts. This entails discrediting of others (violation of standard 8) and the expression of hostility (standard 9). Others are discredited when the speaker uses personal attacks and devaluation rather than objective arguments. The expression of hostility increases this attack as the other is treated as if s/he "were your personal enemy" (Schreier et al. 1995, 284).

Violations of fair argumentation in everyday communication also include a number of other aspects that can play an important role in the genesis of hate speech. The violation of standards of fair argumentation (developed and validated by Schreier et al. 1995) is perceived as a breach of the sincerity and cooperativeness in argumentative communication. Such violations occur on four dimensions, namely faulty arguments (I), insincere contributions (II), unjust arguments (III), and unjust interactions (IV) (Schreier et al. 1995, 282). Discrediting is a prototypical violation on the dimension of unjust arguments, while the expression of hostility lies on the dimension of unjust interactions (Schreier et al. 1995, 284). Within dimension III, the distortion of meaning (standard 6) also plays an important role in the development of hate speech. One way in which this plays out is, for example, when arguments of the targeted group are misrepresented so that they will be rejected more forcefully. Within the dimensions of faulty arguments (I) and insincere contributions (II) there are also several standard violations within which hate speech frequently operates. On the level of insincere contributions, for example, shifting of responsibility (standard 4) is present when stigmatized and persecuted groups are blamed for political events and societal problems for which they do not have true or objective responsibility. Likewise, the pretense of truth (standard 3) can be viewed as a general instrument of hate speech, since subjective attitudes are presented as objective truth and even falsehoods are communicated with the certainty of an objective claim of truth. On dimension I, refusal of justification (standard 2) is particularly relevant for hate speech. Strong ideological statements are usually slanted to such a degree that they can be maintained only by refusing rational discussion when confronted with counterarguments.

The dimensions outlined here demonstrate that problematic features of everyday communication are relevant in the development of hate speech and can be used as a guideline to detect it in early stages. Future research will have to clarify how often and to what extent each of these standards is violated in hate speech. This research will also provide the further development and empirical validation of tools that can be used to detect characteristics of hate speech and sensitize listeners to this issue (see also Christmann et al. 2000).

V. SUMMARY: DETECTING AND DECONSTRUCTING HATE SPEECH

To summarize, hate speech plays a crucial role in the evolution of violence. It can manipulate listeners and distract from constructive solutions to existing problems in society. However, the analysis presented here suggests ways in which listeners can analyze information to detect hate speech and resist its influence. Based on our review, we would like to propose that the following elements are central characteristics of hate speech that should alarm the listener.

(1) *The communication contains instigating elements of the continuum of violence.*
   a) The core element is a distinction between "us" and "them." Individuals or groups are referred to by using their group membership, and information about their (alleged) origin is used to label them as foreigners. Frequently this is achieved by pointing to their affiliation with a region, nationality, religion, or language group different from that of the majority of listeners.
   b) These individuals or groups are blamed for the misfortune of the country in terms of historical or present difficulties.
   c) These individuals or groups are accused of disloyalty, treachery, alliance with other countries (in particular with the enemy) or the previous regime, thereby implying threat and appealing to the listeners' emotions.

(2) *The communication is derogatory and violates standards of (argumentative) integrity.*
   a) Personal insults and attacks on the integrity of an individual are involved, and the communication is defaming and derogatory.
   b) The arguments are unbalanced and are not objectively verifiable with facts from other sources or standards of a rational argumentative debate.
   c) The legitimacy and ability of an individual or group to hold political

power and influence is questioned, or it is claimed that this person or group has too much power.

    d) The targeted group or individual is denied distinct characteristics of human nature.

    e) An individual or group is threatened, for example with revenge.

(3) *The suggested strategies do not offer real or constructive solutions to existing problems, and serve self-interests of the speaker and/or his or her group only while harming another group.*

    a) The speaker attains direct political gain and an increase in power by harming the target.

    b) There is a focus on individuals or groups instead of on issues.

    c) By focusing on one alleged source of problems and blaming the targeted group or individual, the accuser offers solutions that are simplistic and do not take into account the complexity and multifaceted nature of societal problems. The promised solutions are therefore not real solutions to the existing situation.

    d) The offered solutions are destructive rather than constructive in nature, as they are based on the exclusion of certain individuals or groups from political power or the society in general.

    e) The communicated ideas and suggested solutions for problems are not inclusive of all in society, but instead benefit a specific group while excluding others.

TABLE 1. CHARACTERISTICS OF HATE SPEECH VERSUS NEUTRAL SPEECH

Hate speech	Neutral speech
Arguments can be shared/supported by only one group	Arguments can be shared/supported by many groups
Biased view, one-sided and distorted perspectives	Balanced view, multiple perspectives
Destructive spirit and solutions	Constructive spirit and solutions
Focus on blame, personal attacks (on the integrity of one person or group)	Focus on issues and facts
Solutions benefit only one person or group	Solutions benefit all of society
Exclusive/excluding of others	Inclusive of all
Simple	Complex
Derogatory language	Neutral, respectful language
Emotionally charged (especially anger, fear)	Neutral and objective communication

These characteristics of hate speech are summarized in table 1. While not all must be present in a given piece of communication in order to define it as hate speech, this classification provides a tool that allows us to analyze any given statement, speech, or article for elements that typically distinguish neutral communication from hate speech. It must be pointed out that this dichotomy is a simplification that has been drawn for the sake of illustration. In reality, the boundaries are more fluid. However, we believe that this contrasting depiction provides a useful heuristic that will enable listeners to detect signals of hate speech in the early stages.

VI. Conclusion: Counteracting Hate Speech

Once hate speech has been detected, sensitized and motivated listeners can engage in activities that counteract its destructive effects in several ways. In discussions with family, friends, neighbors, or other members of their communities who are recipients of hate speech, they can question the soundness and truth of the faulty arguments and insist on justifications and facts that would support the arguments presented. Sometimes common sense, the use of a critical consciousness, and already existing knowledge among those who are participants in such discussions can disconfirm the veracity of hate speech. At other times, multiple media outlets and diverse sources (including foreign and independent press, or information from impartial, non-governmental organizations) can be used in order to provide evidence of the bias in the views presented, as well as the objective facts that challenge them. In discussions about the groups or individuals that are targeted by dehumanizing hate speech, descriptions can be used that humanize them. This can be achieved, for example, by emphasizing their characteristics that constitute human uniqueness as described above, or by pointing out other aspects of shared humanity, such as common fate, or shared goals and emotions.

In communication with members of the groups targeted by hate speech, it is crucial to provide emotional support and show solidarity. It is important to show awareness and address the nature of surrounding issues and the fact of hate speech, even though such discussion is sensitive and likely to cause unease, for example because of the sudden appearance of differences through hate speech where before perhaps none had been perceived. Specifically, it should be expressed that the nature of the accusations and derogations has been recognized, and is not shared. Exploratory research (Vollhardt 2006) has shown that such solidarity and support from members of groups that are not targeted can alleviate the harmful effects on the target. Additionally, when solidarity and rejection of the derogations is expressed by someone who belongs to the same social group as those who

are using hate speech, it makes it less likely that the entire group will be perceived as antagonistic, and therefore reduces the potential for cycles of violence.

In conclusion, on the one hand it is evident that hate speech is a destructive political tool that must be fought on the societal level, with legislation and other structural and political interventions. However, on the other hand it is also important to keep the active role of the recipient in mind, and the individual's ability to apply knowledge to detect and analyze hate speech even in early stages. Educational programs and sensitization campaigns, such as the reconciliation radio programs in Rwanda, Burundi, and the DRC, are essential for increasing this knowledge. They reinforce the listener's ability to resist manipulative influence and strengthen the individual's role as a powerful contributor to counteracting hate speech in society.

REFERENCES

Bandura, Albert. 1990. Mechanisms of moral disengagement. In *Origins of terrorism: Psychologies, ideologies, theologies, states of mind,* ed. Walter Reich, 161-91. New York: Cambridge University Press.

Bandura, Albert, Bill Underwood, and Michael Fromson. 1975. Disinhibition of aggression through diffusion of responsibility and dehumanization of victims. *Journal of Research in Personality* 9: 253-69.

Boeckmann, Robert, and Carolyn Turpin-Petrosino. 2002. Understanding the harm of hate crime. *Journal of Social Issues* 58: 207-25.

Brewer, Marilyn. 1979. Ingroup bias in the minimal group situation: A cognitive-motivational analysis. *Psychological Bulletin* 86: 307-24.

Chalk, Frank, and Kurt Jonassohn. 1990. *The history and sociology of genocide: Analyses and case studies*. New Haven: Yale University Press.

Christmann, Ursula, Christoph Mischo, and Jürgen Flender. 2000. Argumentational integrity: A training program for dealing with unfair argumentational contributions. *Argumentation* 14: 339-60.

Committee to Protect Journalists a. Fragile Freedom (Report by Julia Crawford). http://www.cpj.org/Briefings/2004/DRC_9_04/DRC_9_04.html.

Committee to Protect Journalists b. Attacks on the press in 2005: Democratic Republic of Congo. http://www.cpj.org/attacks05/africa05/drc_05.html.

Cowan, Gloria, Miriam Resendez, Elizabeth Marshall, and Ryan Quist. 2002. Hate speech and constitutional protection: Priming values of equality and freedom. *Journal of Social Issues* 58: 247-63.

Deci, Edward, and Richard Ryan. 2000. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry* 11: 227-68.

De Fossard, Esta. 1996. *How to write a radio serial drama for social development:*

2006/07]   DECONSTRUCTING HATE SPEECH   33

A script writer's manual. Baltimore, MD: John Hopkins University Center for Communications Programs. Available online: http://www.jhuccp.org/pubs/fg/3.

Des Forges, Allison. 1999. *Leave none to tell the story: Genocide in Rwanda*. New York: Human Rights Watch.

Gaertner, Samuel, and John Dovidio. 2000. *Reducing intergroup bias: The common ingroup identity model*. Philadelphia: Psychology Press.

Genocide Watch a. The eight stages of genocide. http://www.genocidewatch.org/8stages2006.htm.

Genocide Watch b. Genocide emergency: Ituri, Eastern Congo. http://www.genocidewatch.org/GenocideEmergencyIturi.htm.

Gourevitch, Philip. 1998. *We wish to inform you that tomorrow we will be killed with our families: Stories from Rwanda*. New York: Picador.

Haslam, Nick. 2006. Dehumanization: An integrative review. *Personality and Social Psychology Review* 10: 252-64.

Human Rights Watch. D.R. Congo: Halt growing violence ahead of elections. http://hrw.org/english/docs/2006/10/25/congo14441.htm.

International Crisis Group. Securing Congo's elections: Lessons from the Kinshasa showdown. Crisis Group Africa Briefing No. 42, 2 October 2006. http://www.crisigroup.org/library/documents/africa/central_africa/b042_securing_congos_elections.pdf.

ISN Security Watch. Tensions high as DRC awaits poll results. www.isn.ch/news/sw/details.cfm?id=16871.

John Hopkins University Center for Communication Programs. http://www.jhuccp.org.

La Benevolencija. 2005a. Stichting Radio La Benevolencija/Humanitarian Tools Foundation (Radio Benevolencija). http://www.labenevolencija.org.

La Benevolencija. 2005 b. *Great Lakes Reconciliation Radio. Radio programs and grassroots activities as instruments of prevention of conflict, reconciliation and restoration in the Great Lakes region (Rwanda, Burundi, Kivu - DRC)* [Unpublished project proposal]. Amsterdam, Kigali: Stichting La Benevolencija.

Levy-Paluck, Elizabeth. 2006. *The second year of a "New Dawn." Year two evidence for the impact of the Rwandan reconciliation drama Musekeweya*. Unpublished evaluation report for La Benevolencija.

MSNBC Newsweek International Editions. As Congo votes, U.N. put to the test. http://www.msnbc.msn.com/id/15473851/site/newsweek.

Monuc (UN Mission in DR Congo). Monitoring of 2nd November 2006, by Yulu Kabamba. http://www.monuc.org/news.aspx?newsID=12936.

Mullen, Brian, and Joshua Smyth. 2004. Immigrant suicide rates as a function of ethnophaulisms: Hate speech predicts death. *Psychosomatic Medicine* 66: 343-48.

Nzongola-Ntalaja, Georges. 2003. *From Zaire to the Democratic Republic of the Congo.* 2nd ed., rev. Uppsala: Nordiska Afrikainstitutet.

Pearlman, Laurie Anne, and Karen Saakvitne. 1995. *Trauma and the therapist: Countertransference and vicarious traumatization in psychotherapy with incest survivors*. New York: W. W. Norton.

Radio Netherlands Media Network. Counteracting Hate Radio: Definitions. http://www.radionetherlands.nl/features/media/dossiers/definitions.html.

Schreier, Margit, Norbert Groeben, and Ursula Cristmann. 1995. "That's not fair!" Argumentational integrity as an ethics of argumentative communication. *Argumentation* 9: 267-289.

Staub, Ervin. 1989. *The roots of evil. The origins of genocide and other group violence.* New York: Cambridge University Press.

Staub, Ervin. 2005. The origins and evolution of hate, with notes on prevention. In *The psychology of hate*, ed. Robert Sternberg, 51-66. Washington, DC: American Psychological Association.

Staub, Ervin. 2006. Reconciliation after genocide, mass killing or intractable conflict: Understanding the roots of violence, psychological recovery and steps toward a general theory. *Political Psychology* 27: 867-94.

Staub, Ervin, and Laurie Anne Pearlman. 2001. Healing, reconciliation and forgiving after genocide and other collective violence. In *Forgiveness and reconciliation: Religion, public policy and conflict transformation*, ed. Raymond Helmick and Rodney Petersen, 205-29. Radnor, PA: Templeton Foundation Press.

Staub, Ervin, and Laurie Anne Pearlman. 2005. Psychological recovery and reconciliation after the genocide in Rwanda and in other post-conflict settings. In *Psychological interventions in times of crisis,* ed. Laura Barbanel and Robert Sternberg, 213-43. New York: Springer.

Staub, Ervin, Laurie Anne Pearlman, Alexandra Gubin, and Athenase Hagengimana. 2005. Healing, reconciliation, forgiveness and the prevention of violence after genocide or mass killing: An intervention and its experimental evaluation in Rwanda. *Journal of Social and Clinical Psychology* 24: 297-34.

Staub, Ervin, Laurie Anne Pearlman, and Vachel Miller. 2003. Healing the roots of genocide in Rwanda. *Peace Review* 15: 287-94.

Stola, Dariusz. 2005. Fighting against the shadows: The Anti-Zionist campaign of 1968. In *Antisemitism and its opponents in modern Poland*, ed. Robert Blobaum, 284-300. Ithaca: Cornell University Press.

Szaynok, Bożena. 2005. The role of antisemitism in postwar Polish-Jewish relations. In *Antisemitism and its opponents in modern Poland*, ed. Robert Blobaum, 265-83. Ithaca: Cornell University Press.

Tajfel, Henri, and John Turner. 1986. A social identity theory of intergroup behavior. In *Psychology of intergroup relations,* ed. Stephen Worchel and William Austin, 7-24. Chicago: Nelson-Hall.

Turner, John. 1987. A self-categorization theory. In *Rediscovering the social group: A self-categorization theory*, ed. John Turner, Michael Hogg, Penelope

Oakes, Stephen Reicher, and Margaret Wetherell, 42-67. Oxford: Basil Blackwell.

Tsesis, Alexander. 2002. *Destructive messages: How hate speech paves the way for harmful social movements*. New York: New York University Press.

Voice of America News. DRC observers seek transparency in vote counting (by Nico Colombant, November 01, 2006). http://www.voanews.com/english/2006-11-01-voa38.cfm.

Vollhardt, Johanna. 2006. *Victim consciousness and interpretations of ethnopolitical persecution: A qualitative interview study with Polish Jews about their experiences of the March events in 1968*. Unpublished data, University of Massachusetts.

Wolak, Arthur. 2004. *Forced out: The fate of Polish Jewry in Communist Poland*. Tucson: Fenestra.