Dangerous Speech and Dangerous Ideology: An Integrated Model for Monitoring and Prevention

Jonathan Leader Maynard  
*University of Oxford*

Susan Benesch  
*American University; Harvard University*

Follow this and additional works at: https://scholarcommons.usf.edu/gsp

Recommended Citation
DOI: http://dx.doi.org/10.5038/1911-9933.9.3.1317

Available at: https://scholarcommons.usf.edu/gsp/vol9/iss3/8

This Article is brought to you for free and open access by the Open Access Journals at Scholar Commons. It has been accepted for inclusion in Genocide Studies and Prevention: An International Journal by an authorized editor of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Dangerous Speech and Dangerous Ideology: An Integrated Model for Monitoring and Prevention

Jonathan Leader Maynard  
The University of Oxford  
Oxford, United Kingdom

Susan Benesch  
American University  
Washington, D.C., USA

Abstract: There is considerable agreement amongst scholars and international actors that ideologies and speech play a critical role in the path of escalation towards mass atrocity crimes. Speech features prominently in the jurisprudence of the U.N. war crimes tribunal for Rwanda, for example, and in historical accounts of the months and years preceding many other genocides. Nonetheless, this is one of the most underdeveloped components of genocide and atrocity prevention, in both theory and practice. This paper draws together the authors’ independent past work on dangerous speech and the ideological dynamics of mass atrocities by offering a new integrated model to help identify the sorts of speech and ideology that raise the risk of atrocities and genocides. We suggest that this model should inform monitoring activities concerned with the risk of genocides and mass atrocities, and prevention efforts at the strategic and targeted levels.

Keywords: dangerous speech, ideology, atrocities, genocide, prevention, justification, discourse

Introduction

On 12 March 1941, the German Reichskommissar of the occupied Netherlands, Arthur Seyss-Inquart, addressed collected German officials in Amsterdam, using archetypal pre-genocidal language. Reaffirming a chronic claim of Nazi ideology—that Jews sought to annihilate the Germans—he stated:

The Jews are the enemy of national socialism and the national socialistic Reich. From the moment of their emancipation, their methods were directed to the annihilation of the common and moral worth of the German people... They are the enemies with whom we can neither come to an armistice nor to peace...”

Seyss-Inquart’s language has since been echoed by a great many propagandists in a striking diversity of countries and cultures, to cultivate fear and hatred, and incite mass violence against civilians. This is a particular sort of speech that scholars and policymakers have begun to focus on for its apparent role in promoting mass atrocities. As paragraph 21 of the United Nations Secretary-General’s 2013 report on State Responsibility and Prevention affirms:

[T]he underlying motivation for targeting a community...is an additional risk factor for atrocity crimes. This motivation is often demonstrated through the use of exclusionary ideology and the construction of identities in terms of “us” and “them” to accentuate differences. These differences can be even further accentuated through hate speech or propaganda campaigns that depict the targeted community as disloyal, as a security or economic threat or as inferior in order to justify action against the community.

Such discourse is indeed typical. But advocates of mass violence also use a variety of other, quite different rhetorics in calling for the killing of civilians. Their justifications do not always rely on hatred or blatant exclusionary discourse. Twenty three years before Seyss-Inquart’s words, Vladimir Ilyich Lenin published one of the most detailed justifications of the fledgling Soviet

The regime’s ongoing ‘Red Terror’ in the Russian Civil War. In a letter in the 22 Aug 1918 issue of Pravda, the official newspaper of the Bolshevik party, Lenin denounced critics of Soviet violence, arguing:

To blame us for the “destruction” of industry, or for the “terror”, is either hypocrisy or dull-witted pedantry; it reveals an inability to understand the basic conditions of the fierce class struggle…. for in revolutionary epochs the class struggle has always, inevitably, and in every country, assumed the form of civil war, and civil war is inconceivable without the severest destruction, terror and the restriction of formal democracy in the interests of this war. Only uncouth parsons… cannot see, understand and feel this necessity. Only a lifeless “man in the muffler” can shun the revolution for this reason instead of plunging into battle with the utmost arduity and determination at a time when history demands that the greatest problems of humanity be solved by struggle and war.¹

Lenin does not justify terror here via hatred of those it targeted, but in light of a theoretically dense claim that revolutionary violence is necessary. This constituted a powerful and highly violent worldview which endorsed and rationalized atrocities by the new Soviet government against its own civilians. And Lenin’s arguments here are not idiosyncratic outliers. Assertions that violence is inevitable and unavoidable, and the displacing of responsibility for it onto others (or simply onto abstract historical forces), are recurring features of the ideologies behind many different atrocities.

Most such claims are quite different from the hate speech or dehumanization that tend to dominate current discussions of the role of speech, ideology and propaganda in catalyzing mass violence. So while it is often important, hate can be something of a distraction.² Research on perpetrators by Christopher Browning, Lee Ann Fujii, John Mueller, Scott Straus and others suggests that hatred of another group often isn’t what drives a person to kill.³ Some scholars erroneously conclude from this that ideology or speech in general are not catalysts for violence —assuming that ideology and speech can encourage violence only by inculcating hatred. This is a mistake. Instead, a broad repertoire of justifications for atrocity can help explain why many perpetrators commit violence even though they do not match classic stereotypes of the wild-eyed fanatic, enthusiastically slaughtering civilians. Both speech and the ideology that underpins it can be dangerous (in the sense of promoting violence) without being hateful, and can also be hateful without being dangerous. Less ardent beliefs matter: cool but anxious perceptions that other groups have radical or threatening ambitions, for instance, or a sense that terror is the only option for resolving a particular social problem, or a belief that violence is honourable and necessary.

Critically, such justifications are patterned across cases. The forms of speech and ideology that catalyze mass violence, and the ways in which they do so, are strikingly similar across different cases. The same sorts of justification recur with uncanny similarity across cases from widely differing time periods, across vast geographical distances, and in spite of contrasting social, political and cultural contexts.⁴ This should not obscure differences—sensitivity to the context-specific ideological and  

---

discursive dynamics of particular atrocities is crucial, and the recurring patterns do not tell the whole story. But the fact that there are such patterns is not only important but useful, holding open the possibility that discerning them can support efforts to prevent mass atrocities.

This breadth in the way atrocities are justified, patterns in such justification, and the implications this creates for monitoring and prevention activities, have not been adequately addressed in the literature.7 Speech and ideology feature prominently in the jurisprudence of the U.N. war crimes tribunal for Rwanda, the risk analysis framework of the Office of the Special Advisor for the Prevention of Genocide, the report of the Genocide Prevention Task Force and a wide range of academic research.8 Yet references to “hate”, “dehumanization” and related forms of blatantly vicious discourse still dominate such documents.9 Whilst scholars have generated considerable insights regarding the ideological and discursive dynamics of genocides and mass atrocities, especially in deep, case-specific research,10 analysis of the role played by speech and ideology remains one of the more underdeveloped components of the comparative theorising of genocides and mass atrocities, and the lacuna is largest with respect to monitoring and prevention.

In this paper, we begin to fill these gaps, in order to enhance both the theory and practice of mass atrocity risk monitoring and prevention. We do this in three steps. First, we provide a brief theoretical overview of the relationship between speech and ideology, and explain how these two phenomena relate to violence. Second, we combine two existing frameworks that we have independently formulated elsewhere—Benesch’s ‘Dangerous Speech’ framework11 and Leader Maynard’s ‘Six Justificatory Mechanisms’ framework.12 Together, these provide an integrated model for identifying the contextual and content-based risk factors associated with certain sorts of speech and ideology. They can thus provide a more comprehensive basis for monitoring activities that might support efforts to assess the risk of mass violence. Finally, we consider how this integrated model can inform attempts to counter dangerous speech and ideology, and help to prevent genocide and mass atrocities.

Analysing Speech and Ideology in Mass Violence

The relationship between speech and ideology (or, more generally, thought) has long been a subject of theoretical reflection in linguistics, philosophy, intellectual history, political theory and social theory. Yet it generally lacks theoretical clarification in contemporary scholarship on mass violence. This is unfortunate, since speech and ideology are vitally interrelated.

Following recent trends in the study of ideology in the social sciences, we conceptualize ideology broadly, defining it as a distinctive system of normative and/or purportedly factual ideas, typically shared by members of groups or societies, which underpins individuals' understanding of their political world and guides their political behaviour. As systems of ideas, ideologies are stored in memory, providing cognitive resources for thinking processes, including decision-making processes and, thereby, shaping agents' behaviour. Each person's ideology is unique (their 'personal ideology'), but social scientists also analyze 'group ideologies': analytical constructs describing important similarities across the personal but heterogeneous ideologies of group members. Ideologies can also acquire important social dimensions as they are embedded in institutions and recognized in political discourse. By speech or 'speech act,' we mean any act of human communication, not only in the form of verbal discourse but also including non-verbal communication such as images, gestures, music, rituals, and so forth.

Ideologies are communicated by speech and, consequently, constructed and altered through speech. They are also produced through an individual's own thinking, but genuinely creative thinking is arduous. Most people avoid it most of the time, instead picking up and reusing prominent and familiar ideas from social discourse. The bulk of ideological development in a group or society therefore tends to be driven by the thinking of a relative minority, though these may not necessarily conform to the presumed 'intellectual elite', and various forms of bottom up...
activism are often ideologically powerful.\textsuperscript{20} Just as speech communicates and alters ideology, so ideology essentially underpins speech. Speech acts often express ideological claims, and draw on the speakers’ underlying ideologies. And the impact the speech acts have on others is necessarily mediated by the existing ideologies of those individuals—since their ideologies provide the existing assumptions, beliefs and interpretative frameworks used to understand, reflect on and evaluate such speech.

Ideological claims might represent \textit{instrumental} moves that serve a variety of interests. They might, for example, be ways for community or political leaders to legitimate themselves in the minds of domestic constituencies and audiences, or they might be used to solicit financial or symbolic support from external actors, like great power patrons or affiliated diasporas.\textsuperscript{21} Ideological claims can be used in this way, regardless of whether those espousing them truly believe them. But ideological speech might also genuinely represent actors’ personal ideologies—here ideology plays a \textit{constitutive} role in providing actors’ actual beliefs and understandings.\textsuperscript{22} These could be \textit{evaluative} beliefs about the permissibility or impermissibility of attacking certain sorts of targets, or of the moral praiseworthiness or reprehensibility of violence in general. Or they could be \textit{descriptive} beliefs, about the threat posed by a certain group, or the likelihood that a shocking atrocity will deter future aggression against the in-group, which make violence appear necessary or desirable.\textsuperscript{23}

Consequently, speech can matter since it offers evidence of the ideologies of actors producing it, or because it is likely to create certain effects in audiences. Scholars cannot properly assess a risk of violence by examining speech without considering ideology, or ideology without considering speech. One cannot acquire any deep understanding of actors’ ideological positions without examining their actual discourse (broadly conceived). Familiar big ideological labels like ‘communist’, ‘Islamist’, or ‘nationalist’, so often used to code the ideological positions of key actors for social scientific analysis, encompass far too many distinct strains of belief-system to reliably predict or analyze behaviour. Describing actors’ ideologies more specifically and accurately is possible, but only by consulting what they actually say. Conversely, successfully interpreting what actors meant by certain speech, discerning how that speech connects to other speech acts, evaluating the likelihood that it is sincere, and assessing its likely impact on audience behaviour, requires consideration of the probable ideology of the speaker and the ideological content expressed in the speech.\textsuperscript{24} And predicting what impact the speech is likely to have on an audience is therefore impossible without some knowledge of the ideological environment, including recent history of the context in question, and the probable ideological commitments of the audience(s).

Ideas can be held by perpetrators with varying levels of consciousness and conviction. They might be explicit and fanatical beliefs that drive fervent participation in violence. Yet this is typically not the case. Much of the time, key ideas might be believed ambivalently, regretfully, presumptively, unthinkingly, or for lack of apparent alternatives. These weaker, more nuanced levels of endorsement can still be sufficient for the ideas to generate participation in violence in a variety of roles.\textsuperscript{25} Frequently, the key burden for success for a particular ideology in turning a group toward violence is not how many people entirely internalize the ideology and become devotees or even perpetrators, but how many \textit{partially internalize} the ideology enough to see violence as permissible or even desirable. Although perpetrators constitute only a small proportion of the total population of an in-group, even in cases of mass violence with high levels of civilian participation, they rely on the support and approval of other members of their social groups.

\textsuperscript{20} Tarrow, Sidney. 2013. \textit{The Language of Contention}. Cambridge: Cambridge University Press.


\textsuperscript{22} Ibid, 220-222.

\textsuperscript{23} Though evaluation and description are intertwined.


Ideas also influence behaviour in different ways. Whether deeply or shallowly felt, they can provide key motives for violence, or they can provide key legitimations for it, shaping perceptions of its permissibility even if they do not provide the primary reasons for wanting to engage in it. Or they may do both. They might sincerely drive the original decisions to engage in violence, or function as a post-hoc rationalization for acts conducted for other reasons (though such rationalizations might still be crucial for maintaining willingness to participate in lengthy, reiterative campaigns of violence). And ideas can also, of course, restrain violence.²⁶ The balance between the availability and power of ideological justifications and restraints in a particular context determines the overall ideological impetus towards violence—and this is crucial for thinking about how to counter such justifications, as we do in Part 3.

We do not suggest that dangerous speech and ideology are the only catalysts for mass violence, nor that they deterministically produce it. Human beings have autonomy to choose among actions, and their decisions are affected by myriad external factors, including a large variety of speech to which they are exposed over time.²⁷ Therefore it is rare to find definitive evidence that a person engaged in violence because of a particular act of speech, though exceptions include orders to kill or be killed, orders given by commanders to their troops, and certain inflammatory false rumours.²⁸ We can make educated and systematic estimations of the effects of speech given the best possible data on its context, content, and the likely ideological dispositions of speakers and audience, but our analytical tools often cannot reliably measure the impact of individual acts of speech on individual people, and we do not pretend to be able to predict them accurately. Indeed Benesch has criticized some of the judges of the international war crimes tribunal for Rwanda for asserting that speech ‘caused’ mass killings.²⁹

Scholars have begun to search for evidence of links between speech and intergroup violence at scale. One study on the well-known case of the Rwandan radio station Radio Télévision Libre des Milles Collines (RTLM) used econometrics and transmission patterns to suggest that killing was more intense in Rwandan villages that received RTLM broadcasts, than in others that did not.³⁰ A similar study on the effects of an intervention by ActionAid International Nigeria found that anti-violence campaigning, conversely, substantially reduced levels of subsequent political violence surrounding the 2007 national and state-level elections in Nigeria.³¹ Still, the complex, multifarious links between speech, ideology and violence require greater, broader and deeper investigation.³²

As is widely recognized in the literature, genocides and mass atrocities do not occur spontaneously, but represent the culmination of lengthy and often non-linear radicalizing trajectories.³³ Mass violence may recur with tragic frequency in human history,³⁴ but it is...
nevertheless unusual, not a normal condition of human existence. Some degree of radicalization from the norm is thus necessary. Many parts of such radicalizing journeys are not principally ideological: existing policies are replaced by successively more extreme and provocative options, low level violence psychologically habituates perpetrators for more extensive campaigns, complex bureaucratic arrangements develop progressively and expand a group’s capacity to organize violence on a mass scale, and so forth.

But changes in speech, and the changes in ideologies they reflect and cause, are part of the radicalizing trajectory. In the months or years preceding mass intergroup violence, relevant groups of perpetrators and their supporters come increasingly to see violence as permissible and even necessary. Repertoires of ideas and arguments that encourage such a perception become increasingly formulated and disseminated through speech—spoken, written or otherwise—allowing greater numbers of potential perpetrators to see violence as increasingly thinkable, possible, and justified, and to possess the justificatory resources necessary to convince others. Such trajectories are complex, may be halting and uneven, and involve a mix of ‘top-down’, ‘bottom-up’ and ‘horizontal’ processes. Although explicit and easily identifiable propaganda campaigns can be powerful, they are not the only form that matters. Much dissemination and radicalization occurs through diffuse and gradual processes of political, cultural and intellectual change.

The mix of dissemination and radicalization processes is visible in a number of classic cases of mass violence. In the Soviet Union, for example, security policies towards perceived internal threats progressively radicalized from the early 1930s onwards, leading to the mass Stalinist violence of the 1937-8 period. Much of this was fairly centralized and elite-level. The party leadership around Stalin articulated securitized diagnoses of economic failings and alarmist interpretations of domestic conditions, producing a series of reports that were discussed amongst the Bolshevik Party Central Committee, and then propagated downwards throughout the extensive party apparatus (though there was also positive feedback up the system, and varying forms of mass participation).

In post-World War I Germany, by contrast, key ideas that would provide the foundations for Nazi atrocities initially circulated in a range of veterans networks and paramilitary organizations, that pulled in many young Germans, popularized contempt for Jews and the Weimar Republic, and glorified violence. After the Nazi seizure of power the principal dynamics of dissemination shifted to the extreme, propagandistic indoctrination of large sections of German society through private and state media, the Nazi Party, a range of affiliated Nazi organizations, and the military. In Rwanda, Hutu extremism developed amongst key elites many decades prior to the 1994 genocide. Their radical ideas were increasingly disseminated through the education system, quasi-official publications, and radio. In the months before the genocide itself, radio networks like Radio Rwanda and RTLM disseminated radicalizing Hutu Power notions, and social and political leaders disseminated such ideas in speeches that were further propagated among friends, neighbors, and even family members.


relatives by word of mouth. Through such dissemination and radicalization, as Stuart Kaufman summarizes, “the extremists were able to make their genocidal program accepted as part of the normal political landscape.”

Such processes of radicalization consistently include speech that justifies violence in characteristic ways, using recurring ideological and rhetorical patterns. To understand how this encourages violence, we need a model of the features of speech, and the ideologies behind it, that are dangerous.

Dangerous Speech and Dangerous Ideology: An Integrated Model

The force of a speech act, and specifically its capacity to encourage approval of violence in the minds of listeners, is the product of both the context and content of what is said (and these feed into and overlap with one another). Even the most inflammatory or powerful justification of violence may have little chance of leading to atrocities if uttered in a context where it cannot be disseminated, where the audience strongly disapproves of violence, or where the speaker has no influence or authority. Conversely, a highly authoritative speaker engaging with a volatile audience and using powerful mechanisms of dissemination will not raise the risk of atrocities unless the content of what s/he says portrays violence as permissible or desirable, producing ideological notions that allow an individual to see atrocities as justified.

Identifying when speech is in danger of causing violence must therefore be a two-part inquiry. In this section we propose an analytical framework that describes the features of both context and content that lead to speech which may increase the risk of atrocities.

Context

The context can be described systematically with reference to four of its features or aspects, any or all of which can confer greater force on the speech act, i.e. make it more influential: the speaker, the audience, the socio-historical environment, and the means of dissemination.

First, in many cases the speaker or source of the message may be a powerful contextual element. A speaker might have authority over the audience derived from political office or de facto leadership. In other cases (and in some of the same ones, since one person may have multiple bases of influence), a speaker derives authority from religious position or status, or from their status as a public performer, such as an actor or athlete. Or they may carry particular ‘epistemic’ authority in light of audiences’ dependence on them for information that is deemed credible. Even where there is no vocational basis for influence, a speaker may acquire influence with charisma, or a persuasive speaking style that skillfully deploys rhetoric and mobilizes an audience’s affective inclinations.

The second contextual element is the *audience* itself. No audience is homogeneous, and individual members of the audience vary in their propensity to commit violence, for any number of reasons. Analysis can and should focus on audiences, or audience members, most likely to respond violently. In almost any human group or society, young men are more likely to attack other people than other demographic subgroups. In addition, people of any age and gender can become more likely to resort to violence – or to condone it when it is committed on their behalf – when they are afraid of being seriously harmed by another person or group, and especially when they believe themselves – or their dependents – to be subject to an existential threat. The extent to which an audience is susceptible to dangerous speech is thus, itself, partly a question of the existing audience ideology.

The third element consists of the *social and historical context*, which can increase the chance that inflammatory messages will be accepted by the audience even where they are exaggerated or entirely false, as in the example with which we began this article. Other scholars have described social and structural characteristics that increase the chances of intergroup violence; in fact there has been so much work on cataloguing and examining the relative weight of such factors (often for early warning efforts) that we do not attempt to repeat it all here. Some of the most relevant points are longstanding grievances between the relevant groups, a weak or dysfunctional justice system, competition among groups for scarce basic resources, and land disputes. And again, assessment of the social and historical context is in part a question of the extant ideological environment. Notions of grievance and memories of historical injustice form part of the existing ideological landscape that justifiers of atrocity seek to capitalize on and intensify. We have also observed some other historical factors that can make speech more dangerous, such as previous episodes of violence following similar inflammatory speech. Such episodes put both the speaker and the audience on notice that speech can indeed be catalytic of violence.

In some cases, language itself is an influential feature of relevant social and historical context. A single word can play this role, by taking on a fearsome meaning, as when verbs meaning “to eat” and “to wash” were used to refer to killing in the run-up to elections in June 2015 in Burundi. Such expressions were all the more powerful because the same words were used in the same ways during past killing in Burundi, and in neighboring Rwanda, before the 1994 genocide.

The last contextual element is the *means of dissemination* of speech. Where a community or audience relies predominantly on one source of news or information, the fact that a message comes from that source may, itself, render the message more influential. This was the case for broadcasts over the notorious Rwandan radio station RTLM, according to witnesses who testified at the Rwanda war crimes tribunal. Similarly, the International Criminal Court chose to indict justifiers of atrocities at the Rwanda war crimes tribunal. Similarly, the International Criminal Court chose to indict

---


Joshua arap Sang, a Kenyan radio presenter, for crimes against humanity on the theory that his broadcasts over the airwaves of the Kalenjin-language station KASS-FM were very powerful, given the great influence of that station among Kalenjins in Kenya’s Rift Valley where massacres occurred in early 2008. Perpetrators can easily become “epistemically dependent”—reliant on particular disseminators for their information, and thereby internalize even very radical notions where there is a practical monopoly on local public discourse.55 This allows disseminators to saturate a given context with atrocity-justifying ideas, allowing even fantastical and abhorrent notions to appear as ‘common sense’, and something that ‘everybody says’.56

Social media are starting to play a significant role in dissemination of dangerous speech and ideology. Increasingly, certain websites and online platforms are dominant sources of information for some people, who can easily come to believe in dangerous ideological claims when disseminated on such platforms via particular, often extremist, online networks: ‘silos’ or ‘echo chambers.’57 Even the choice of language itself can be a critical form of dissemination: the same message communicated in the “mother tongue” of an ethnic group can have more force than if it were delivered in a language shared by other groups, since this reinforces the sense of solidarity within the group, and may encourage a feeling of impunity given the presumption that one will only be understood by co-linguals.58

Two important clarifications are in order. First, although one of us is a legal scholar, this analytical framework is not intended to list the elements of a crime. Instead it articulates salient dangers created by certain contextual features, as a tool for monitoring and preventing mass violence. We believe that criminal prosecution for speech cannot be relied upon, especially alone, for the prevention of atrocities. It may even backfire in some cases, by making a speaker and his or her speech better known and therefore more influential than they would have been, without legal intervention.

Second, speech may be ‘dangerous’ even when only some of the prongs of our contextual analysis are relevant. For example, in some cases a speech act is highly dangerous even though the speaker is not particularly influential; indeed the speaker is sometimes unknown to the audience, especially when messages are disseminated online. As an example, in some communities in Pakistan and Afghanistan, to accuse a group of people of burning or defacing the Quran is such a dangerous speech act that it can easily trigger mob violence, even if the source of the accusation is neither influential nor authoritative.59

Only two elements are indispensable: the content of the speech act itself must be dangerous (see below), and the audience must be susceptible to incitement. An earlier version of this framework was mistakenly constrained by some as a set of necessary conditions.60 Certainly a very dangerous speech act is one in which all five prongs are maximized. A highly influential speaker would address a susceptible audience in highly inflammatory terms, in a volatile social and historical context, using a powerful means of dissemination. This would be prototypical dangerous speech, but it is far from the only form that it takes. Not only are the prongs not all necessary in every case, but even where they are all relevant, they weigh quite differently in proportion to one another.

---

55 Baumann, “Rational Fundamentalism?”
58 Author interviews (Benesch), Kenya, 2011 & 2012.
For example, where speech is produced by a very powerful or influential speaker, that speaker’s identity may weigh heavily, and the medium by which the speech was disseminated may not carry much weight. In another case, the opposite could be true.

Content
As noted in our introduction the claims and arguments used to justify mass atrocities—the actual content of dangerous speech and dangerous ideology—are (like dangerous contextual factors) so similar from case to case that they form characteristic patterns. Based on past work, we identify six justificatory mechanisms, of varying degrees of theoretical originality. Collectively, these provide a framework for thinking about the major recurring ways in which violence is made to look permissible, desirable and even necessary before, during, and after mass atrocities, and thus describe the content of dangerous speech that makes it likely to catalyze violence. These six mechanisms overlap heavily, so the borders between them are blurry, and individual speech acts tend to involve multiple mechanisms at once. Together, the six mechanisms provide a framework for more easily identifying dangerous speech and ideology in practice in monitoring and prevention efforts.

1. Dehumanization
Targets of dangerous speech are described in a variety of ways that deny or diminish their humanity, reducing the moral significance of their future deaths, or the duties owed to them by potential perpetrators. Violence against victims has been variously justified by describing them as either biologically subhuman (“cockroaches”, “microbes”, “parasites”, “yellow ants”), mechanically inhuman (“logs”, “packages”, “enemy morale”), or supernaturally alien (“devils”, “Satan”, “demons”). Dehumanising discourses and conceptions have been identified in almost all major mass atrocities, prominently including those of Nazi Germany, Stalinist Russia, Rwanda, Yugoslavia, Cambodia, Indonesia, and the Japanese occupation of China. Often, outgroup members (or victims-to-be) are even compared with toxins, microbes, or cancer, suggesting that they are polluting, despoiling, or debilitating the entire in-group—leading to particularly prominent recurring demands to ‘purify’ groups or societies from the supposedly toxifying elements. Multiple mechanisms explain this effect. Dehumanization can consciously defeat normative concerns about violence by allow perpetrators to believe that the targets of violence lack moral protections. Like guilt attribution and threat construction, dehumanization moves out-group members into a social category in which conventional moral restraints on how people can be treated do not seem to apply. They are now “outside the universe of [moral] obligation,” as Helen Fein has described it. But dehumanization can also work at a much less conscious level – perceptually eroding affective moral concern for certain

---

62 Haslam, “Dehumanization.”
63 Sémelin, Purify and Destroy, 38 & 243; Weitz, Century of Genocide, 156.
68 Fein, Accounting for Genocide, 4-9.
categories of person, and encouraging emotional responses of revulsion or antipathy. A series of influential psychological studies by Albert Bandura and colleagues thus highlight the power of dehumanization by finding that even single dehumanizing words heard casually can frame outgroup members in such a way so as to substantially increase levels of aggression and violence against them.69

2. **Guilt Attribution**

Individuals, or an entire group, are said to be guilty of heinous past crimes against the in-group (or audience of the speaker), which warrant a violent response. Violence is presented as just punishment (and perhaps vengeance)70 for their wrongdoing. Such claims of guilt are often at the core of anti-civilian ideologies, and elites go to extensive efforts to cement them in the minds of killers.71 In the ongoing Syrian civil war, for example, "security services and official media [have] spread blood-curdling, often exaggerated and sometimes wholly imaginary stories of the protesters’ alleged sectarian barbarism."72 Similarly, the Rwandan Hutu nationalist magazine *Kangura* stated: "The unspeakable crimes of the Inyenzi ['cockroaches', referring to Tutsi] of today... recall those of their elders: killing, pillaging, raping girls and women, etc."73 Indeed, rumours or accusations of rape are particularly prominent—and play on sexual anxieties that fuel macho nationalism on the part of young male perpetrators.74 Critically, victims are often deemed guilty as a group, deserving collective punishment for the specific crimes of some of their "members".75 Such processes expose the enormous atrocity-justifying potential of demonyms or group-labels, as predicted by prominent psychological research on social categorization.76 Once real or imagined crimes committed by specific, individual human beings are repeated discussed in terms of "Jews", "Tutsis", "Muslims", "Japs" or similar pluralized identities, perpetrators prove remarkably willing to conceive of all members of the referenced category as equally and uniformly complicit.

3. **Threat Construction**

A speaker asserts that the in-group faces serious and often mortal threats from the victims-to-be, which makes violence seem defensive, and therefore proper and necessary. Tales of future wrongdoing (threat construction) can be even more powerful than tales of past crimes (guilt attribution), since the future is frightening. The significance of such processes is uncontroversial across research on genocide, mass atrocities, and armed conflict. What


73 Midlarsky, *Killing Trap*, 177.


©2016 Genocide Studies and Prevention 9, no. 3 http://dx.doi.org/10.5038/1911-9933.9.3.1317
is critical and not as familiar is that notions of threat are frequently a product of ideological construction.\textsuperscript{77}

This justification is so powerful since it is the collective analogue of the only iron-clad defence to homicide in myriad systems of penal law and cultural norms: self-defence. Where dehumanization makes atrocities seem acceptable, threat construction takes the crucial next step of making them seem necessary. Whether through the passionate assertion of enemy machinations to destroy the in-group, or the more ambivalent circulation of rumours, anecdotal incidents, and vague accusations, perpetrators can successfully propagate beliefs that the threat is real. Such beliefs can sometimes capitalise on genuine threats or at least reasonable insecurities—they need not be complete madness. But they provide reasons to believe that certain groups need to be targeted with violence. As Martin Shaw explains, mass killings are thus accompanied by a “construction of civilian groups as enemies, not only in a social or political but also in a military sense, to be destroyed”, a process often conducted by key military, security and political authorities.\textsuperscript{78}

In a strange yet common form of threat construction, a speaker accuses another group of planning to engage in the sort of violence that the speaker wants to see perpetrated against them, instead. This has been dubbed (originally in a Hutu propaganda manual discovered after the 1994 genocide) \textsuperscript{79} “accusation in a mirror.” \textsuperscript{80} Examples of the technique are legion. In the speech with which we opened this article, Arthur Seyss-Inquart accused Jews of planning to annihilate the German people—a baseless claim that in fact mirrored what the Nazis planned and would attempt to do to the Jews. The idea that Jews would wipe out the German \textit{volk}—if Germans did not pre-empt that effort—was a relentless feature of Nazi propaganda, of which Seyss-Inquart’s speech was just a typical example. Speeches and articles by Hutu leaders before that genocide similarly warned that the Tutsi were planning to annihilate the Hutu. The same technique is also widely used in less famous cases of intergroup violence: examples found by Kenneth Marcus include the claim that Coptic Christians in Egypt were amassing weapons with which to attack the Muslim majority in that country.\textsuperscript{81}

The degree of psychological projection involved can be remarkable—as a Viennese Austrian who participated in the slaughter of two thousand Jews in Mogilev in the Soviet Union wrote to his wife: “By the tenth car, I was aiming calmly and shooting dependably at the many women, children and babies. Bearing in mind that I have two babies at home, I knew they would suffer exactly the same treatment, if not ten times as bad, at the hands of these hordes.”\textsuperscript{82} As this quote illustrates, a strong degree of future-bias (see below) is also often present in threat-construction, with a rhetoric of absolute certainty in the future behaviour of others, perhaps in decades hence, key to the reasoning.

Several justificatory mechanisms work together to confer force, or persuasiveness, on accusation in a mirror. It is a powerful form of threat construction, first of all. It also attributes guilt to the victims-to-be for the crimes that they are said to be planning. Finally, it destroys an audience’s perception that it has alternative courses of action, since a group convinced that it faces a mortal or existential threat can see no alternative other than to fend off that threat by any means available. It is in this sense that accusation in a mirror


\textsuperscript{78} Shaw, \textit{What is Genocide?}, 86 & 111.


\textsuperscript{81} Marcus, “Accusation in a Mirror.”

is a collective analogue of the doctrine of self-defence as a near-universal moral and legal justification for killing.

4. **The Destruction of Alternatives**

Violence is presented as inevitable, as necessitated by forces beyond the control of human agency, or as the only choice open to perpetrators. Decision-making always occurs within a delimited field of possibilities: human beings do not consciously consider every one of the infinite options in any situation, but focus decision making on those choices that appear salient.83 This ‘menu’ of alternatives can be significantly expanded or contracted by ideology and speech. Eliminating the perceived viability of alternatives to violence is therefore *just as critical* as are assertions of the desirability or permissibility of violent options themselves. There are various ways in which this destruction of alternatives84 can occur. It its most grandiose form, violence might be presented as a historical necessity, as an ineradicable feature of ‘racial struggle’, ‘class conflict’, ‘human progress’, ‘the nature of war’ and so forth. The Holocaust, Communist mass killings, and colonial genocides against native peoples, were all prominently justified by their perpetrators as simply mandated by iron laws of nature and historical change. Genocide thus became seen, in the words of one American perpetrator of genocidal policies against the Native Americans, “as ultimately beneficial as it was inevitable”.85 Hitler asked his dinner guests: “In any case, is it we who created nature, established its laws? Things are as they are, and we can do nothing to change them”.86 And citizens of the Soviet Union later testified that they thought the violence of the Stalinist era was simply the only possible and thus necessary path to Communism: “I had my doubts about the Five Year Plan... But I justified it by the conviction that we were building something great...a new society that could not have been built by voluntary means”.87 But often the destruction of alternatives is more contingent that this: alternatives might be destroyed by normatively disparaging them and associating them with reprehensible moral qualities (see ‘Virtuetalk’ below); or they might be presented as inefficacious or impractical, often as a matter of assertion or taken-for-granted self-evidence; or they may be framed as unacceptable given particular situations – ‘supreme emergencies’ or ‘states of exception’88—so that mass atrocities seem the only possible course. Frequently, violence is imputed as inevitable in ‘war’ or international anarchy, and thus pointless to criticize. Enthusiastically championing mass aerial bombardment of German civilians in World War II, British newspapers proclaimed to their readers that: “This is the only policy. This is the only effective method available to us in self-defence... The invention of the bombing plane abolished chivalry for ever. It is now ‘retaliate or go under’. We are not dedicated to passive and polite martyrdom. We must hit back...”89 Irrespective of the method, the destruction of alternatives serves to ‘deagentify’ the violence: making it appear to be the product of irresistible inhuman forces rather than conscious choices by policymakers and perpetrators, and thereby promoting moral disengagement from their acts.90 This can occur at a more macro or micro level. With macro-level destruction of alternatives, individual perpetrators might accept that they faced a choice to *participate* in the violence as an individual, but the violence itself takes on

---

83 Foucault, “The Subject and Power.”
90 Bandura, “Moral Disengagement in the Perpetration of Inhumanities.”
the appearance of a simple dictate of the situation or of natural laws, rather than a chosen policy selected by the perpetrators.\textsuperscript{91} Often, of course, the destruction of alternatives occurs at the micro level too—many perpetrators affirm that they personally had no choice, perhaps due to coercion or due to sincere beliefs in specific situational pressures.\textsuperscript{92} This is more likely to be rooted in individual conceptions than in speech and ideology, but these could play a role—situational pressures do not speak for themselves but are interpreted in light of how they are framed in speech and perceived in light of ideological assumptions.\textsuperscript{93} Speech that inculcates fears about punishment for non-conformity to violent campaigns, that frames individual choice as absent, or that spreads intense peer-pressure and notions of required behaviour can be crucial to the destruction of alternatives at the individual level.

5. Virtuetalk

The valorization of violence by associating it with a range of praiseworthy characteristics, and the parallel denigration of resistance or non-participation as indicating a lack of proper character traits, a deplorable “weakness”, or a range of other deficiencies. As a now substantial literature in psychology emphasizes, most ordinary political and ethical thinking is not intellectualized philosophy, but is guided by vague but quick and often deeply felt impressions and intuitions regarding what “looks” or “feels” good or bad.\textsuperscript{94} Researchers have also long recognized that human beings are powerfully motivated to draw a satisfactory mental image of themselves, a positive moral self-identity, often shaped by notions of ideal group-identities, that produces considerable self-esteem.\textsuperscript{95} Virtuetalk capitalizes on these psychological tendencies, directing and regulating individuals’ moral emotions, by reconstructing violence as admirable. Violence is linked through verbal expressions, symbolism and imagery to a range of praiseworthy qualities: duty, honour, courage, toughness, manliness, and so forth. Not to participate in violence is to expose oneself to internal shame or social ridicule as an improper or inadequate member of the in-group, or (often) not a proper or adequate man—since virtuetalk tends to be heavily gendered, targeting the insecurities prominent amongst the young men who might join the ranks of perpetrators.\textsuperscript{96} Typically the virtues chosen represent established social values, that carry considerable existing cultural status. Sometimes, though, virtuetalk may involve an explicit harshening of the dominant ideological discourse of virtue—a ‘revalencing’ of qualities like mercilessness, brutality, and ruthlessness as positive traits. In Nazi Germany, one SS general declared that “every man should be trained to be a fanatical hater”\textsuperscript{97} and a Nazi police battalion instruction manual likewise stated that “he behaves correctly who, by setting aside all possible impulses of personal feeling, proceeds ruthlessly and mercilessly”.\textsuperscript{98} The Hutu Ten Commandments, disseminated in the Kangura newspaper in Rwanda in December 1990, likewise demanded that Hutu “stop having mercy on the Batutsi”,\textsuperscript{99} whilst a Radio Rwanda broadcast a month earlier stated that “since their goal is to exterminate and enslave us, we must not feel any mercy to them”.\textsuperscript{100} The

\textsuperscript{92} Fujii, \textit{Killing Neighbours}, Ch. 6.
\textsuperscript{93} Newman, “A ‘Social-Psychological’ Account,” 60-62.
\textsuperscript{95} Tajfel, “Social identity,” 67-72.
\textsuperscript{97} Mann, \textit{Dark Side of Democracy}, 200.
\textsuperscript{98} Browning, \textit{Ordinary Men}, 183.
\textsuperscript{99} Fujii, “Transforming the Moral Landscape,” 102.
\textsuperscript{100} Ibid, 104.
Khmer Rouge similarly applauded “seething hatred and blood rancour against national and class enemies”. And in parallel to such praise, virtuetaol presents the failure to engage in violence as shameful—an indicator of treason, dishonour or weakness—so that in the words of one former Bolshevik: “I would still my doubts the way I had learned to... the concepts of conscience, honour, humaneness we dismissed as idealistic prejudices, ‘intellectual’ or ‘bourgeois’ and, hence, perverse.” Virtuetaol is crucial for explaining how perpetrators come to see violence as permissible and even desirable, and has not received sufficient analysis in scholarship thus far.

6. Future-Bias

The confident anticipation of future goods that will be accrued through violence, and which are so extensive and so enduring in a relatively certain future that they easily outweigh the moral costs of victims’ deaths in the here and now. Violence is rarely justified without any reference to its consequences, many of which are implicit in the other justificatory mechanisms (it is assumed that good consequences will flow from the elimination of threats or the punishment of guilty parties). The future goods can thus be fairly basic—ensuring that no out-group dare threaten the in-group again, promising that military victory in a campaign will be achieved through extreme violence, or anticipating economic or scientific benefits from atrocities. Himmler justified Nazi massacres of Russian prisoners of war and civilians with the claim that: “Communism is a tremendous danger for the future. We must get away from the standpoint of soldierly comradeship. The Communist is from first to last no comrade. It is a war of extermination. If we do not regard it as such, we may defeat the enemy, but in thirty years’ time we will again be confronted by the Communist enemy.” One German doctor, in a research plan for experiments on handicapped children in the spring of 1942, was confident that these acts would produce results: “…reaching far beyond other scientific discussion and research in the field of psychiatry, at last the most practical and immediate questions affecting the health of the nation can be most comprehensively resolved because thanks to the [euthanasia] programme a rapid anatomical and histological clarification can be achieved.” But the anticipated benefits can also be extravagant and utopian—promises that a positive transformation of society will be brought about through a temporary violent transition, or that national unity and prosperity for a long-mistreated people can be obtained. In light of the expectation that Soviet violence would protect the revolution and usher in Communist utopia, Lenin assured his followers that in the future “the cruelty of our lives, imposed by circumstance, will be understood and pardoned. Everything will be understood, everything.” Soviet ideology and the justification of massive violence and cruelty in the name of a promised future society of abundance convinced millions in the Stalinist era. The novelist Boris Pasternak wrote in a letter in 1935: “The fact is, the longer I live the more firmly I believe in what is being done, despite everything. Much of it strikes one as being savage [yet] the people have never before looked so far ahead, and with such a sense of self-esteem, and with such fine motives, and for such vital and clear-headed...
reasons.” Many individuals can be persuaded that even seriously harmful moral acts can be justified in certain circumstances if a significant enough need or benefit flows from them. Dangerous speech can lead perpetrators to become convinced that they “know” that their actions serve certain ends, or produce certain benefits, often simply because their actions have been defined as aiming at those ends. Often, though, there are deeper foundations for confidence, with expectations for the positive future rooted in underlying ideologies: the assuredly improved future of a racially pure nation, a theologically-correct caliphate, a ‘final solution’ to a correctly diagnosed security threat, and so forth. Either way, the actual uncertainty (or even complete fantasy) of promised future goods is not factored into the consequentialist calculations, and the real and established suffering of victims in the here and now is measured up on an equal basis (or worse) with anticipated and far from certain future benefits.

This integrated model of the contextual and content-based factors can form the foundation for more sophisticated monitoring of dangerous speech and dangerous ideology, and thereby support efforts to assess the risks of genocide and other mass atrocities. Data collection can be difficult but is not an insuperable challenge. The uploading of videos of dangerous speech by supporters or witnesses online, the migration of dangerous speech and ideology itself online, especially into social media networks, and the typical need for perpetrators to engage in widespread dissemination of their ideological speech via print, video and radio media means that dangerous speech and ideology are more easily documented than ever before. In addition, the increasing deployment of crowd-sourcing technology and partially-automated online monitoring is opening up new methods for data-collection. What is needed, though, is an analytical framework that can permit consistent, structured and theoretically informed examination of the data. Already, an earlier version of Benesch’s Dangerous Speech framework was used to support the Umati project, monitoring dangerous speech in Kenya before and after the 2013 national elections there. This new integrated model now offers the sort of framework that we believe can significantly advance the sophistication and accuracy of monitoring.

Countering Dangerous Speech and Ideology
This framework can also support prevention efforts, since efforts to counter dangerous speech and ideology will be more effective when they are based on a detailed understanding of when they are dangerous, and why. This is difficult, of course, and in many cases it is impossible for outside actors to do it successfully. There are no easy options in genocide and atrocity prevention, however, and efforts to monitor and counter dangerous speech and ideology are not especially costly or risky when compared with other measures such as military operations, targeted sanctions, legal indictments, or on-the-ground fact-finding missions. Moreover, prevention efforts are not an all or nothing endeavour—and even if interventions targeting speech and ideology do not prevent genocides or mass atrocities outright, they may reduce participation rates and thus save lives.

---

108 Figes, Whisperers, 190.
111 Awori, Benesch, and Crandall, “Umati: Kenyan Online Discourse to Catalyze and Counter Violence.”
Efforts to counter dangerous speech and dangerous ideology, like other preventive effects, can be conducted at either the long-term ‘systemic’ level, aiming to strengthen obstacles to violence in contexts with a general risk of mass violence, or through more short-term targeted interventions, in response to escalating risks of immediate outbreaks of atrocities. Existing work in the former category includes the programs of Radio La Benevolencia in West and Central Africa to inoculate audiences against inflammatory and hateful speech, and the reconciliation and peace-building projects of organizations like Search for Common Ground and Fondation Hirondelle. Past efforts in the latter category include the extensive saturation of Kenyan society with pro-peace messages in anticipation of the 2013 elections. In Kenya, for example, the cast of the popular television drama Vioja Mahakamani produced four episodes in 2012 on dangerous speech, designed to inoculate audiences against such speech. They were independently evaluated by scholars at the University of Pennsylvania’s Annenberg School of Communications, who found that Kenyans who watched the episodes were more sceptical about inflammatory speech than those who had not watched the episodes. And as noted earlier, a parallel campaign in Nigeria analyzed by Paul Collier and Pedro Vicente appeared to substantially reduce levels of actual violence in the 2007 elections.

Our model can support identification of the sorts of speech and ideology that ought to be countered, and the ideas and interventions that might stand a chance of succeeding at this. Scholars have already suggested appeals or norms that might obstruct the ideas and discourses that serve to justify mass atrocities: stressing, for example “humanism”, “non-divisionism”, “humanization”, and “universalism”. We support these suggestions, but our integrated model highlights how such ideas would target only a limited part of the justificatory practices that can make atrocities look permissible or desirable. One can promote humanistic norms and the perception of universal rights successfully, for example, but if a society is still permeated by perceptions of violence as admirable and virtuous and a compelling narrative of threat-construction asserting the existence of hidden enemies within, atrocities will remain more than a theoretical possibility. Humanism, nondivisionism, humanization and universalism frustrate some avenues through which mass atrocities can be justified, but leave others open. As well as highlighting this problem, our integrated model can help address it, and facilitate a more comprehensive strategy. Efforts to counter dangerous speech and ideology need to encourage scepticism that certain people are subhuman, collectively guilty, or pose an existential threat. In addition, they should cultivate antipathy towards violent virtuethinking, facilitate awareness of alternatives to violence, and encourage doubts over claims that violence will bring great benefits. To accomplish this, they should train political and civil society actors to identify and counter the hallmarks of dangerous speech and ideology that we identify here.

And prevention efforts are not simply a matter of countering the recurring content of justifications of violence, but also, as our model highlights, about eroding the contexts within which such justifications can be powerful. Long-term systemic efforts must address entrenched ideological components of the socio-historical context on which justifications capitalize, and push for diversity in the networks of information and news, undermine epistemic dependence by audiences on unreliable or even malevolent sources of discourse. Short-term targeted efforts should seek to undermine the authority and credibility of those disseminating dangerous ideology, contest audience sympathies, and seek to sway reluctant perpetrators, especially those at the elite level who could potentially


118 Collier and Vicente, “Votes and Violence”.


120 Leader Maynard, “Combating Atrocity-Justifying Ideologies.”
be lobbied into resisting rather than supporting policies that lead to genocide or mass atrocities. At both the systemic and targeted level, online communities of influential actors committed to peace should be mobilized to counter ideas conducive to the justification of mass atrocities.

What clearly emerges from the complexity of the contextual and content-based components of dangerous speech and ideology is that preventive efforts of a long-term systemic or short-term targeted nature require sophisticated planning, strategizing and campaigning. The contextual rootedness of many ideas and appeals means that effective interventions cannot be conducted by external actors. Local expertise, credibility, and persuasive capacity is crucial.

Conclusion
Efforts to monitor and counter dangerous speech and dangerous ideology remain at a very early stage of theory and practice. In this paper, we have advanced an integrated model of the contextual and content-based risk factors that define dangerous speech and dangerous ideology: those communicative acts, and the sets of ideas they reflect and disseminate, that make genocide and other mass atrocities seem permissible and even necessary. This is only a first step in formulating more sophisticated monitoring and preventive strategies. But it is a vital one, if such strategies are going to be built on a more holistic appreciation of the role played by speech and ideology, and work to monitor and counter the full panoply, and not merely a narrow subset, of the words and ideas that encourage the worst of humanity’s crimes.

Bibliography
Asia Pacific Centre for the Responsibility to Protect. 2012. “A Common Standard for Applying R2P.” In *APC R2P Ideas in Brief*.
http://dx.doi.org/10.1093/acprof:oso/9780199288427.001.0001


http://dx.doi.org/10.1111/j.1559-1816.1999.tb00134.x


http://dx.doi.org/10.1093/oxfordhb/9780199747672.013.0004

http://dx.doi.org/10.1111/ecoj.12109

http://dx.doi.org/10.1080/713999448


http://dx.doi.org/10.4324/9780203809068.ch1

©2016 Genocide Studies and Prevention 9, no. 3 http://dx.doi.org/10.5038/1911-9933.9.3.1317


Foucault, Michel. 1982. “The Subject and Power.” *Critical Inquiry* Vol. 8, No. 4: pages 777-795. [http://dx.doi.org/10.1086/448181](http://dx.doi.org/10.1086/448181)


Goldman, Wendy Z. 2011. *Inventing the Enemy: Denunciation and Terror in Stalin’s Russia*. Cambridge: Cambridge University Press. [http://dx.doi.org/10.1017/CBO9780511994906](http://dx.doi.org/10.1017/CBO9780511994906)


Hall, Todd H., and Andrew A.G. Ross. 2015. “Affective Politics after 9/11.” *International Organization* Vol. 69, No. 4: pages 847-879. [http://dx.doi.org/10.1017/S0020818315000144](http://dx.doi.org/10.1017/S0020818315000144)


Milgram, Stanley. 2010 *Obedience to Authority: An Experimental View*. London: Pinter & Martin Ltd.


©2016 Genocide Studies and Prevention 9, no. 3 http://dx.doi.org/10.5038/1911-9933.9.3.1317


