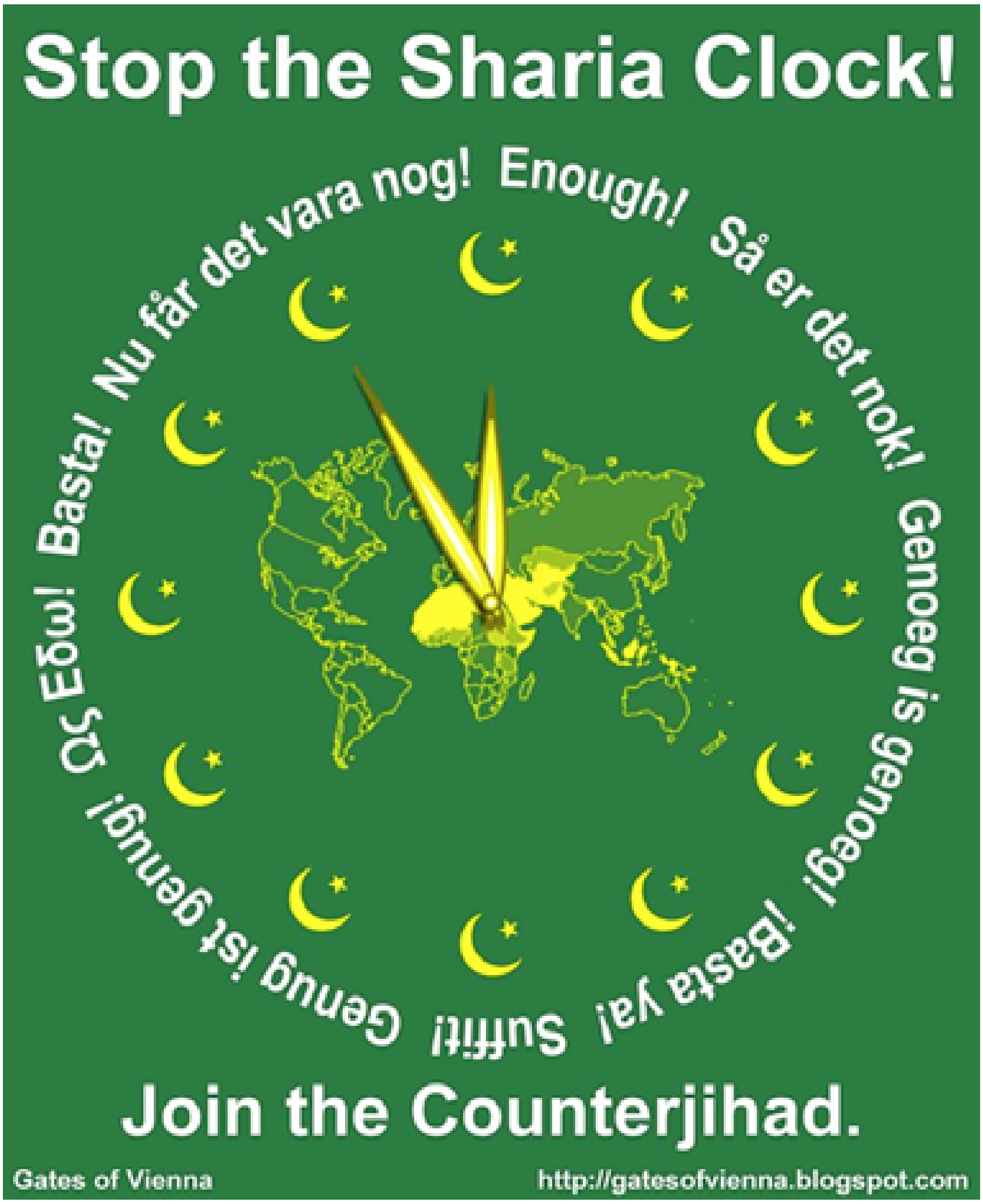
What does the Internet add?

Studying extremism and counter-jihadism online



Reader
Digital Methods International Workshop and Data Sprint
New Media & Digital Culture
University of Amsterdam
14-18 January 2013

CONTENTS

Who Controls the Internet? lllusions of Borderless World

JACK GOLDSMITH
TIM WU

OXFORD UNIVERSITY PRESS

one 1

Introduction

Yahoo!

Marc Knobel is a French Jew who has devoted his life to fighting neo-Nazism, a fight that has taken him repeatedly to the Internet and American websites. In February 2000, Knobel was sitting in Paris, searching the Web for Nazi memorabilia. He went to the auction site of yahoo.com, where to his horror he saw page after page of swastika arm bands, SS daggers, concentration camp photos, and even replicas of the Zyklon B gas canisters. He had found a vast collection of Nazi mementos, for sale and easily available in France but hosted on a computer in the United States by the Internet giant Yahoo.[1]

Two years earlier, Knobel had discovered Nazi hate sites on America Online and threatened a public relations war. AOL closed the sites, and Knobel assumed that a similar threat against Yahoo would have a similar effect. He was wrong. AOL, it turned out, was atypical. Located in the Washington, D.C. suburbs, AOL had always been sensitive to public relations, politics, and the realities of government power. It was more careful than most Internet companies about keeping offensive information off its sites.

Yahoo, in contrast, was a product of Silicon Valley's 1990s bubble culture. From its origins as the hobby of Stanford graduate students Jerry Yang and David Filo, Yahoo by 2000 had grown to be the mighty "Lord of the Portals." At the time, Yahoo was the Internet entrance point for more users than any other website, with a stock price, as 2000 began, of $475 per share.[2] Yang, Yahoo's billionaire leader, was confident and brash—he "liked the general definition of a yahoo: 'rude, unsophisticated, uncouth.'"[3] Obsessed with expanding market share, he thought government dumb, and speech restrictions dumber still. Confronted by an obscure activist complaining about hate speech and invoking French law, Yang's company shrugged its high-tech shoulders.

Mark Knobel was not impressed. On April 11, 2000, he sued Yahoo in a French court on behalf of the International League against Racism and Anti-Semitism and others. Yahoo's auctions, he charged, violated a French law banning trafficking in Nazi goods in France. "In the United States [these auctions] might not be illegal," said Knobel, "but as soon as you cross the French border, it's absolutely illegal"[4] Ronald Katz, a lawyer representing the French groups, added, "There is this naïve idea that the Internet changes everything. It doesn't change everything. It doesn't change the laws in France."[5]

Yahoo received a summons from Le Tribunal de Grande Instance de Paris, Judge Jean-Jacques Gomez presiding. "The French tribunal wants to impose a judgment in an area over which it has no control," reacted Jerry Yang.[6] Yang's public relations team warned of the terrible consequences of allowing national governments to control content on the Internet. If French laws applied to a website in America, then presumably so would German and Japanese regulations, not to mention Saudi and Chinese law. "It is very difficult to do business if you have to wake up every day and say 'OK, whose laws do I follow?'" said Heather Killen, a Yahoo vice president. "We have many countries and many laws and just one Internet."[7]

Jerry Yang embraced 1990s conventional wisdom in thinking that Judge Gomez could legitimately exercise power only in France, and could not control what Yahoo put on its servers in California. French officials, he thought, simply had no authority over a computer in the United States.

Yahoo's Nazi web pages also seemed hard for French officials to stop at the French border. "The volume of electronic communications crossing territorial boundaries is just too great in relation to the resources available to government authorities," wrote David Post and David Johnson, two proponents of a "sovereign" Internet.[8] Even if French officials identified and blocked the offending offshore website, the same information could be posted on mirror sites outside France.

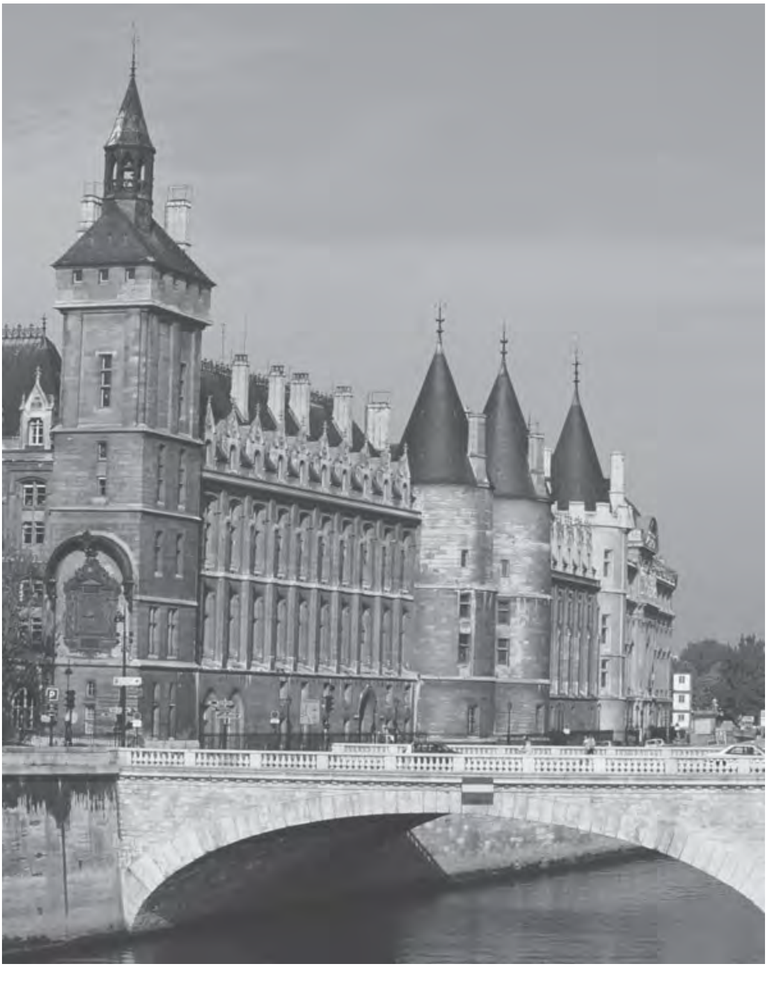
Moreover, the Internet's decentralized routing system was designed to carry messages from point to point even if intermediate communication exchanges are blocked, damaged, or destroyed. "The net interprets censorship as damage, and routes around it," John Gilmore famously declared.[9] To keep out the Nazi pages France would need to shut down every single Internet access point within its borders—seemingly an impossible task. And even this wouldn't have worked, because determined users in France could access the Net by a telephone call to an Internet access provider in another country.

For these reasons, the Internet seemed in the 1990s to have shattered the historical congruence between individual conduct and government power. Some, like Jerry Yang, were sanguine about this development. But many were alarmed. In the midst of the Yahoo trial, Paul Krugman wrote a *New York Times* column about the Net's threat to traditional copyright and tax laws. Internet technology is "erasing boundaries" and undermining government power, he warned. "Something serious, and troubling, is happening—and I haven't heard any good ideas about what to do about it."[10] In the late 1990s, there was broad agreement that the Internet's challenge to government's authority would diminish the nation-state's relevance. "It's not that laws aren't relevant, it's that the nation-state is not relevant," argued Nicholas Negroponte, the co-founder and director of MIT's Media Lab. "The Internet," he concluded, "cannot be regulated."[11]

Yahoo's fearlessness before Judge Gomez thus seemed justified. By the standards of the day, Knobel's effort to stop Yahoo from violating French law seemed dated, ridiculous, and destined to fail.

Paris's Tribunal de Grande Instance is on the Ile de la Cité, the cradle of Parisian civilization, just a few blocks from the Notre Dame Cathedral. It is housed in the beautiful but haunting Palais de Justice, where Marie Antoinette and thousands of others were incarcerated before being guillotined during a different revolution. It was in this ancient building that Yahoo's lawyers would defend the Internet's conventional wisdom against the tradition and glory of the French State.

In Judge Gomez's courtroom, it became clear that the irrelevance of the nation-state would not go uncontested. Knobel's lawyers asserted that France had the sovereign right to defend itself from the sale of illegal Nazi merchandise from the United States, and asked



Palais de Justice, where the Yahoo case was litigated (Martial Colomb/Getty Images)

Yahoo to explain why it ought be exempt from French law. As one anti-Nazi lawyer put it, "French law does not permit racism in writing, on television or on the radio, and I see no reason to have an exception for the Internet."[12]

This simple argument threw Yahoo on its heels. If Yahoo caused harm in France, why should it be any more immune from regulations in different nations than "real-space" multinational firms? The Ford Motor Company must obey the varying safety and environmental laws of the many countries in which it sells cars. Why should Yahoo be exempt from laws in the countries where it does business?

Yahoo responded with an "impossibility" defense. If Ford found French environmental regulations too costly, it could stop selling cars in France without suffering harm in other markets. But Yahoo claimed that its situation was different. It maintained a French-language website (yahoo.fr) that complied with French law. But it also had a U.S. website that the French could visit. And unlike Ford, Yahoo argued, it had no power to identify where in the world its "customers" were from and thus no control over where in the world its digital products go. Were Yahoo forced to comply with French law, it would need to remove the Nazi items from its U.S. server, thereby depriving Yahoo users everywhere from buying them, and making French law the effective rule for the world.

On May 22, 2000, Judge Gomez issued a decision that, on a preliminary basis, rejected Yahoo's arguments. He ruled that Yahoo's U.S. websites violated French law, and he ordered Yahoo "to take all necessary measures to dissuade and make impossible" visits by French web surfers to the illegal Yahoo Nazi auction sites on yahoo.com.[13]

But Yahoo remained defiant. "We are not going to change the content of our sites in the United States just because someone in France is asking us to do so," reacted Jerry Yang.[14] The trial wasn't

Jerry Yang (Robyn Beck/AFP/Getty Images) over yet, and the ability of Yahoo to filter its users by geography would be the key issue. And on this issue, Yahoo felt confident. Said Yang, "Asking us to filter access to our sites according to the nationality of web surfers is very naïve."[15]

Yahoo's "impossibility" argument reflected turn-of-century assumptions about the architecture of the Internet. The Net was not built with physical geography in mind. Neither Internet Protocol Addresses (each computer's Internet ID), nor Internet domain names (such as mcdonalds.com or cnn.com), nor e-mail addresses, were designed to dependably indicate the geographical location of computers on the Net. Even domain names and e-mail addresses with geographical clues—such as toystore.co.fr, or tonyblair@gov.uk—were unreliable. The toy store web page might be located on a computer in Germany (and the data might be cached in dozens of nations), or might be sold or re-assigned to an entity outside France. Prime Minister Blair, meanwhile, could have been reading his e-mail on vacation in Italy, or while visiting the United States.

These architectural "facts" meant that users of 1990s Internet technology could not know where in the world their e-mail messages and web pages were being viewed, and thus what laws in which nations they might be violating. "In Cyberspace, physical borders no longer function as signposts informing individuals of the obligations assumed by entering a new, legally significant, place," said Johnson and Post in 1997.[16] One reason why it seemed unfair for France to apply its laws to Yahoo was that Yahoo didn't know where particular users were, and thus didn't know which laws it should be complying with.

France's attempt to govern Yahoo seemed unfair for another reason. Internet firms and users confronted with a bevy of conflicting national laws could reasonably be expected to comply with the strictest among them in order to avoid legal jeopardy. The ultimate effect of territorial control of the Net thus seemed to be a tyranny of unreasonable governments. "We now risk a race to the bottom," said Alan Davidson of the Center for Democracy and Technology about the *Yahoo* case. "The most restrictive rules about Internet content—influenced by any country — could have an impact on people around the world."[17]

There's an old European joke that captures the problem. In heaven, the joke goes, you find French cooks, English government, Swiss trains, and Italian lovers. In Hell, by contrast, you find French government, Italian trains, English chefs, and Swiss lovers. Territorial control of the Internet seemed to promise a parallel version of legal hell: a world of Singaporean free speech, American tort law, Russian commercial regulation, and Chinese civil rights.

Judge Gomez gave Yahoo two months to figure out how to block French surfers. During this recess, Cyril Houri, the founder of a fledgling American firm called Infosplit, contacted the plaintiff's lawyer, Stephane Lilti, and told him that he had developed a new technology that could identify and screen Internet content on the basis of its geographical source. Houri flew to Paris and demonstrated his technology on Lilti's computer. The men blinked and peered into the screen, astonished. Yahoo's servers, which the firm had claimed were protected by the U.S. First Amendment to the U.S. Constitution, were actually located on a website in Stockholm! Yahoo had placed a constantly updated "mirror" copy of its U.S. site in Sweden to make access to the site in Europe faster.[18]

When the trial resumed on July 24, Yahoo lawyers again asserted that it was technically impossible to identify and filter out French visitors to the firm's U.S.-based websites. Lilti responded by discussing Houri's geo-location technology in the courtroom. Yahoo auctions in France, he argued, were not in fact coming from servers in the United States. The assumption that every web page was equally accessible to every computer user everywhere in the world, Lilti claimed, was simply wrong. If Yahoo could target French users from Swedish servers, it could potentially identify users by geography and, if it liked, screen them out.

Judge Gomez responded cautiously to this seemingly audacious claim and appointed three Internet experts—Vinton Cerf, the "father" of the Internet, Ben Laurie, a British Internet expert, and Francois Wallon, a French technologist—to assess the extent to which Yahoo could block transmissions into France. The experts' report was devastating. It relied on the state of technology in late 2000—namely Houri's IP-identification technology, and self-reporting about nationality—and concluded Yahoo could effectively screen out 90 percent of French users.[19]

Based on this report, Judge Gomez issued a landmark final decision on November 20, 2000, that reaffirmed that Yahoo had violated

French law by allowing Nazi goods to appear for sale on web pages there.[20] The judge determined that the French court had power over Yahoo and its servers because the company had taken conscious steps to direct the prohibited Nazi auction pages into France. He pointed out that Yahoo greeted French visitors to its U.S. website with French-language advertisements. This showed both that Yahoo was tailoring content for France, and that it could to some extent identify and screen users by geography.[21] The court acknowledged that 100 percent blocking was impossible, and ordered Yahoo to make a reasonable "best effort" to block French users.[22]

Yahoo remained indignant. It announced that it would ignore Judge Gomez's decision unless a U.S. court made it do otherwise.[23] A month after the decision, it filed a counter-lawsuit in the United States meant to block the French judgment. "We hope that a U.S. judge will confirm that a non-U.S. court does not have the authority to tell a U.S. company how to operate," said Yahoo France's managing director Philippe Guillanton.[24]

But the company had a problem. While Yahoo thought it would be impossible for a French court to exercise power in the United States, Yahoo also had assets in France, including income from a sizeable French subsidiary, at risk of seizure.[25] Judge Gomez warned the firm that it had until February 2001 to comply before facing fines of 100,000 francs (about $13,000) per day.[26] Yahoo executives, who make frequent trips to Europe and who would be subject to legal process there, began to think things through.

On January 2, 2001, Yahoo abruptly surrendered. It pulled all Nazi materials from its auction sites, announcing that it "will no longer allow items that are associated with groups which promote or glorify hatred and violence, to be listed on any of Yahoo's commerce properties."[27] It weakly asserted that it was motivated by bad publicity from the Nazi auctions, and not the French ruling. "Society as a whole has rejected such groups," said a Yahoo spokesperson.[28] But the timing and threat of French sanctions suggest otherwise—that Yahoo's will had broken.

Soon after Judge Gomez's decision, Yahoo's resistance to geographical screening began to wane. In June 2001, Yahoo announced a deal with Akamai, a content delivery company, to use the firm's geographical identification technology to deliver geographically targeted advertising, in order to "increase advertising relevance."[29] One of Yahoo's lawyers, Mary Wirth, had the unenviable job of explaining the firm's contradictions on geo-ID. "We argued that . . . it's not a 100 percent accurate solution for the French court order because we would have to identify (French citizens) with 100 percent accuracy, and that's not possible. [However,] the technology is perfectly appropriate for ad targeting purposes."[30]

And then Yahoo took the next step. In 1999, it had established a new venture in a new place: the People's Republic of China. When Yahoo first entered the Chinese market, it announced that Yahoo China would "give Internet users in China easy access to a range of Yahoo's popular services tailored to meet the needs of this audience."[31] But the Chinese government had its own ideas about what its citizens needed. As a condition of market access, it eventually demanded that Yahoo filter materials that might be harmful or threatening to Party rule. The Chinese government, in effect, asked Yahoo to serve as Internet censor for the Communist party.

We do not know if there was a long internal debate at Yahoo, or whether the company searched its libertarian soul before deciding to go forward. But we do know that in 2002, Yahoo was not the brash and confident firm it had been just a few years earlier. By end of the summer of 2002, Yahoo shares, valued at $475 in 2000, were now trading at $9.71.[32] A new and better search engine, Google, whose motto was "don't be evil," had become the new darling of Internet information retrieval. Yahoo had to do something, and the Chinese market looked to be the future.

In the summer of 2002, Yahoo quietly agreed to China's demands. It signed a document called the *Public Pledge on Self-Discipline for the Chinese Internet Industry* in which it promised to "inspect and monitor the information on domestic and foreign Websites" and "refuse access to those Websites that disseminate harmful information to protect the Internet users of China from the adverse influences of the information"[33] Ken Roth, the executive director of Human Rights Watch, criticized Yahoo for promising "to identify and prevent the transmission of virtually any information that Chinese authorities or companies deem objectionable."[34]

By 2005 Yahoo had come full circle. The darling of the Internet free speech movement had become an agent of thought control for the Chinese government. Yahoo today provides Chinese citizens with a full suite of censored products. Its Chinese search engines do not return full results, but block sites deemed threatening to the public order. Yahoo's popular chat rooms feature software filters designed to catch banned phrases like "multi-party elections" or "Taiwanese independence." It also employs human and software censors to monitor chat room conversations. All this led the group Reporters without Borders in 2004 to label Yahoo a "Chinese police auxiliary."[35]

In the fall of 2005, Chinese Journalist Shi Tao sent an e-mail to a democracy website in the United States. He attached to the e-mail a memorandum recording a Communist party meeting that discussed ways to deal with the anniversary of Tiananmen Square. But Shi Tao made a serious mistake—he used his Yahoo e-mail account to send the document. When Chinese authorities discovered it on the website in the United States, they asked Yahoo to help identify its sender. Yahoo complied, and Tao was thrown in prison for ten years. How did Jerry Yang, the one-time champion of Internet freedom, explain his company's new role? "To be doing business in China, or anywhere else in the world, we have to comply with local law," explained Yang. "I do not like the outcome of what happens with these things," Yang added. "But we have to follow the law."[36]

The Yahoo story encapsulates the Internet's transformation from a technology that resists territorial law to one that facilitates its enforcement. But the Internet's challenge to the nation-state was much more profound than the Yahoo story suggests, and the nation-state's response has been much more complex and, at times, tentative. To understand the transformations of the past decade, we must begin by examining why so many people believed that the Internet might transcend territorial law and render the nation-state obsolete. This is the task of part 1.

Author: Sunstein, Cass.
Title: Republic.com
Year: 2001
Publisher: Princeton University Press
Place: Princeton
ISBN: 0-691-07025-3
Pages: 51-88

ARNO: 292609

3 fragmentation and cybercascades

There is a discussion group in cyberspace. The group was started two years ago by about a dozen political activists, who were concerned about the increasing public pressure for gun control and the perceived "emasculation" of the Second Amendment (in the group's view, a clear ban on government restrictions on the sale of guns). But the group was also troubled by the growing authority of government, especially the national government, over the lives of ordinary people, and worried as well about the threat to our "European heritage" and to "traditional moral values" that is posed by the increasing social power of African-Americans and "radical feminist women." The group's members were fearful that the Republican and Democratic parties had become weak-willed "twins," unable and unwilling to take on the "special interests" who were threatening to "take away our constitutional liberties." The group called itself the Boston Tea Party.

The members of the Boston Tea Party now number well over four hundred people, who regularly exchange facts and points of view, and who share relevant literature with one another. For a majority of the participants, the discussion group provides most of the information on which they base their judgments about political issues. Over the last two years the Boston Tea Party's concerns have been greatly heightened. Nearly 70 percent of the members carry firearms, some as a result of the group's discussions. Small but vigorous protests have been planned, organized, and carried out in three state capitols. A march on Washington, D.C., is now in the works. Recent discussion has occasionally turned to the need for "self-protection" against the state, through civil disobedience and possibly through selective "strikes" on certain targets in the public and private sectors. The motivation for this discussion is the widely disseminated view that the "FBI and possibly the CIA" are starting to take steps to "dismember" the group. One member has sent bomb-making instructions to all members of the Boston Tea Party. No violence has occurred as yet. But things are unquestionably heading in that direction.

So far as I know, there is no Boston Tea Party. This story is not true. But it is not exactly false. It is a composite based on the many discussion groups and Websites, less and often more extreme, that can be found on the Internet. Discussion groups and Websites of this kind have been around for a number of years. On March 23, 1996, for example, the Terrorist's Handbook was posted on the Internet, including instructions on how to make a bomb (the same bomb, as it happens, as was used in the Oklahoma City bombing, where dozens of federal employees were killed). On the National Rifle Association's "Bullet 'N' Board," a place for discussion of matters of mutual interest, someone calling himself "Warmaster" explained how to make bombs out of ordinary household materials. Warmaster explained, "These simple, powerful bombs are not very well known even though all the materials can be easily obtained by anyone (including minors)." After the Oklahoma City bombing, an anonymous notice was posted not to one but to dozens of Usenet news groups, listing all the materials in the Oklahoma City bomb and exploring ways to improve future bombs. Hundreds of hate groups are now reported to be communicating on the Internet, often about conspiracies and (this will come as no surprise) formulas for making bombs. Members of such groups tend to communicate largely or mostly with one another, feeding their various predilections. The two students who launched the attack in Littleton, Colorado, actually had an Internet site containing details about how to make a bomb. Often such sites receive and spread rumors, many of them false and even paranoid.

Of course these are extreme cases. But they reveal something about the consequences of a fragmented speech market. In a system with robust public forums and general interest intermediaries, self-insulation is more difficult, and people will frequently come across views and materials that they would not have chosen in advance. For diverse citizens, this provides something like a common framework for social experience. "Real-world interactions often force us to deal with diversity, whereas the virtual world may be more homogeneous, not in demographic terms, but in terms of interest and outlook. Place-based communities may be supplanted by interest-based communities."[1] Let us suppose that the communications market continues to become far more fragmented, in exactly the sense prophesied by those who celebrate the "Daily Me," and in a way that invites the continuing emergence of highly specialized Websites and discussion groups of innumerable sorts.

What problems would be created as a result?

FLAVORS AND FILTERS

It is obvious that if there is only one flavor of ice cream or only one kind of toaster, a wide range of people will make the same choice. (Some people will refuse ice cream and some will rely on something other than toasters, but that is another matter.) It is also obvious that as choice is increased, different individuals, and different groups, will make increasingly different choices. This has been the growing pattern with the proliferation of communications options. Consider the celebratory words David Bohnett, founder of geocities.com: "The Internet gives you the opportunity to meet other people who are interested in the same things you are, no matter how specialized, no matter how weird, no matter how big or how small."[2]

The specialization of Websites is obviously important here; so too for the existence of specialized discussion groups of countless kinds. But other technologies are important as well. Consider the World Wide Web Consortium's platform for Internet content selection (PICS), which serves to rate and filter content on the Internet. The authors of PICS hope to put in place a system in which users can filter out materials of any kind, through choosing ratings systems from their preferred sources. Those who seek the ratings of the Conservative

Coalition could use its ratings system, whereas those who prefer the ratings system of the American Civil Liberties Union could use its ratings system. This is merely an illustration of the multiple ways in which new technologies reduce the "friction" of ordinary life and permit people, with increasing ease, to devise a communications universe of their choosing. But this is not only an occasion for celebration.

To see this point, it is necessary to think a bit about why people are likely to engage in filtering. The simplest reason is that people often know, or think they know, what they like and dislike. A friend of mine is interested in Russia; he subscribes to a service that provides him with about two dozen stories about Russia each day. If you are bored by news stories involving Russia, or the Middle East, or if you have no interest in Wall Street, you might turn your mind off when these are discussed; and if you can filter your newpaper or video programming accordingly, it's all the better. And many people like hearing discussions that come from a perspective that they find sympathetic. If you are a Republican, you might prefer a newspaper with a Republican slant, or at least without a Democratic slant. Perhaps you will be most willing to trust "appropriately slanted" stories about the events of the day. Your particular choices are designed to ensure that you can trust what you read. Or maybe you want to insulate yourself from opinions that you find implausible, indefensible, or invidious. Everyone considers some points of view beyond the pale, and we filter those out if this is at all possible. Consider the fact that after people make automobile purchases, they often love to read advertisements for the very car that they have just obtained. The reason is that those advertisements tend to be comforting, because they confirm the wisdom of the decision.

We can make some distinctions here. Members of some groups want to wall themselves off from most or all others simply in order to maintain a degree of comfort and possibly a way of life. Some religious groups self-segregate for this reason. Such groups are tolerant of pluralism and interested largely in self-protection; they do not have ambitions on others. Other groups have a self-conscious "combat mission," seeking to convert others, and their desire to self-segregate is intended to strengthen their members' convictions in order to promote long-term recruitment plans. Political parties sometimes think in these terms, and they often ignore the views of others, except when they hold those views up to ridicule. My own empirical study of political Websites (discussed below) suggests that when links are provided to other Websites, it is often to show how dangerous, or how contemptible, competing views really are.

OVERLOAD, GROUPISM, AND *E PLURIBUS PLURES*

In the face of dramatic recent increases in communications options, there is an omnipresent risk of information overload—too many options, too many topics, too many opinions, a cacophony of voices. Indeed the risk of overload and the need for filtering go hand-in-hand. Bruce Springsteen's music may be timeless, but his 1992 hit, "57 Channels (and Nothin' On)," is hopelessly out of date in light of the number of current programming options, at least if contemporary television is put together with the Internet. (Contradicting Springsteen, TiVo exclaims, "There's always something on TV that you'll like!") Filtering, often in the form of narrowing, is inevitable to avoid overload, to impose some order on an overwhelming number of sources of information.

By itself this is not a problem. But when options are so plentiful, many people will take the opportunity to listen to those points of view that they find most agreeable. For many of us, of course, what matters is that we enjoy what we see or read, or learn from it, and it is not necessary that we are comforted by it. But there is a natural human tendency to make choices, with respect to entertainment and news, that do not disturb our preexisting view of the world. I am not suggesting that cyberspace is a lonely or antisocial domain. In contrast to television, many of the emerging technologies are extraordinarily social, increasing people's capacity to form bonds with individuals and groups that would otherwise have been entirely inaccessible. E-mail and Internet discussion groups provide increasingly remarkable opportunities, not for isolation, but for the creation of new groups and connections. This is the foundation for the concern about the risk of fragmentation.

Consider some relevant facts about the current communications market. If you take the ten most highly rated television programs for whites, and then take the ten most highly rated programs for African-Americans, you will find little overlap between them. Indeed, seven of the ten most highly rated programs for African-Americans rank as the very *least* popular programs for whites. Similar divisions can be found on the Internet. Some sites are specifically designed for African-Americans and (it is fair to speculate) are not often consulted by others. American Visions, for example, describes itself as "the magazine of Afro-American culture" and as the biggest, if not the first, "Internet site aimed at African-Americans." Afritech was established primarily "as a forum for black professionals and academics to discuss technical issues." Tony Brown Online is said to be, among other things, "a place where blacks can meet one another." Melanet describes itself as offering "the Uncut Black Experience" focused on "peoples throughout the African Diaspora" and provides a number of services, many of them involving African themes. Of course thousands of Websites, probably millions, are written primarily by and for whites (even if their designers were not self-conscious about this). There are sharp divides along lines of gender as well. Only one site (hotmail.com) can be found on both the list of top sites among women over fifty and the list of top sites among men over fifty. Among girls aged twelve to seventeen, the top entertainment sites in 1998 were Eonline.com, Pathfinder.com, and Titanicmovie.com, whereas the top entertainment sites among boys in the same age group were ESPN.com, Playboy.com, and Song Online.

All this is just the tip of the iceberg. Not surprisingly, people of certain interests and political convictions tend to choose sites and discussion groups that support their convictions. "Because the Internet makes it easier to find like-minded individuals, it can facilitate and strengthen fringe communities that have a common ideology but are dispersed geographically. Thus, particle physicists, Star Trek fans, and members of militia groups have used the Internet to find each other, swap information and stoke each others' passions. In many cases, their heated dialogues might never have reached critical mass as long as geographical separation diluted them to a few parts per million."[3] Many of those with committed views on one or another topic—gun control, abortion, affirmative action—speak mostly with each other. In the mid 1990s, a study found "a bleak vision of democratic discourse on the Web," with only 15 percent of partisan sites offering links to opposing viewpoints.[4] The author concludes that "far from fostering deliberative political discourse, most of the surveyed Websites sought to consolidate speech power and served to balkanize the public forum."[5]

My own study, conducted with Lesley Wexler for this book in June 2000, found the same basic picture. Of a random study of sixty political sites, only nine (15%) provide links to sites of those with opposing views, whereas thirty-five (almost 60%) provide links to like-minded sites (see table 3.1).

TABLE 3.1. LINKS TO ALLIES AND ADVERSARIES

Political Orientation	Links to Opposition	No Links to Opposition	Links to Like-Minded Sites	No Links to Like-Minded Sites	Total Number of Sites
Republican	3	7	7	3	10
Democrats	1	11	7	5	12
Conservative	1	20	12	9	21
Liberal	4	13	9	8	17
Total	9	51	35	25	60

One of the most striking facts here is that when links to opposing sites are provided, it is often to show how dangerous, or dumb, or contemptible the views of the adversary really are. Talkleft.com, for example, provides links to several Websites with opposing viewpoints, calling them "Political Sites: The Danger Zone." (There are impressive exceptions. To its credit, the National Organization for Women provides links to Promise Keepers, which it considers an antifeminist organization; the American Conservative Union provides a neutral-sounding set of links to sites of presidential candidates.) Even more striking is the extent to which sites are providing links to like-minded sites. Table 3.1 shows the number of sites that have one or more such links; but in a way it greatly understates what is happening. Several organizations, for example, offer links to dozens or even hundreds of like-minded sites. Consider environmentalhealth.com (a liberal environmental organization with over 200 links), democrats.org/yda (93), dems200.org (69), heritage.org (a conservative group; 93), fundems.com (a liberal group; over 200), and rga.org (a Republican group; 50).

All this is perfectly natural, even reasonable. Those who visit certain sites are probably more likely to want to visit similar sites, and people who create a site with one point of view are unlikely to want to promote their adversaries. Nor is definitive information yet available about the extent to which people who consult sites with one point of view are restricting themselves to like-minded sources of information. But what we now know, about both links and individual behavior, supports the general view that many people are mostly hearing more and louder echoes of their own voices. This may well be damaging from the democratic standpoint.

I do not mean to deny the obvious fact that any system that allows for freedom of choice will create some balkanization of opinion. Long before the advent of the Internet, and in an era of a handful of television stations, people made choices among newspapers and radio stations. Magazines and newspapers, for example, often cater to people with definite interests in certain points of view. Since the early nineteenth century, African-American newspapers have been widely read by African-Americans, and these newspapers offer distinctive coverage of common issues and also make distinctive choices about what issues are important.[6] Whites rarely read such newspapers.

But what is emerging nonetheless counts as a significant change. With a dramatic increase in options, and a greater power to customize, comes a corresponding increase in the range of actual choices, and those choices are likely, in many cases, to match demographic characteristics, preexisting political convictions, or both. Of course this has many advantages; among other things, it will greatly increase the aggregate amount of information, the entertainment value of choices, and the sheer variety of options. But there are problems as well. If diverse groups are seeing and hearing quite different points of view, or focusing on quite different topics, mutual understanding might be difficult, and it might be increasingly hard for people to solve problems that society faces together.

Take some extreme examples. Many Americans now believe that AIDS is a minor problem, one that is diminishing in degree and faced largely by people who have recklessly chosen to take risks. Many other Americans think that AIDS is an extremely serious problem, growing in degree, and fueled by government indifference and perhaps even by deliberate efforts by white doctors to spread the disease within African-

American communities. Many Americans fear that certain environmental problems—abandoned hazardous waste sites, genetic engineering of food—are extremely serious and require immediate government action. But others believe that the same problems are imaginative fictions generated by zealots and self-serving politicians. Many Americans think that most welfare recipients are indolent and content to live off of the work of others. On this view, "welfare reform," to be worthy of the name, consists of reduced handouts, a step necessary to encourage people to fend for themselves. But many other Americans believe that welfare recipients generally face severe disadvantages and would be entirely willing to work if decent jobs were available. On this view, "welfare reform," understood as reductions in benefits, is an act of official cruelty.

To say the least, it will be difficult for people, armed with such opposing perspectives, to reach anything like common ground or to make progress on the underlying questions. Consider how these difficulties will increase if people do not know the competing view, consistently avoid speaking with one another, and are unaware how to address competing concerns of fellow citizens.

A BRIEF NOTE ON HATE GROUPS

As noted, there are hundreds of Websites created and run by hate groups and extremist organizations. They appear to be obtaining a large measure of success, at least if we measure this by reference to "hits." My own informal survey shows that several hate groups have had well over one hundred thousand visitors, and in at least one case well over one million. What is also striking is that many extremist organizations and hate groups provide links to one another, and expressly attempt to encourage both recruitment and discussion among like-minded people.

Consider one extremist group, the so-called Unorganized Militia, the armed wing of the Patriot movement, "which believes that the federal government is becoming increasingly dictatorial with its regulatory power over taxes, guns and land use."[7] A crucial factor behind the growth of the Unorganized Militia "has been the use of computer networks," allowing members "to make contact quickly and easily with like-minded individuals to trade information, discuss current conspiracy theories, and organize events."[8] The Unorganized Militia has a large number of Websites, and those sites frequently offer links to related sites. It is clear that Websites are being used to recruit new members, to allow like-minded people to speak with one another, and to reinforce or strengthen existing convictions. It is also clear that the Internet is playing a crucial role in permitting people who would otherwise feel isolated, or who might move on to something else, to band together and to spread rumors, many of them paranoid and hateful.

There are numerous other examples along similar lines. A group naming itself the "White Racial Loyalists" calls on all "White Racial Loyalists to go to chat rooms and debate and recruit with NEW people, post our URL everywhere, as soon as possible."Another site announces that "Our multi-ethnic United States is run by Jews, a 2% minority, who were run out of every country in Europe. . . . Jews control the U.S. media, they hold top positions in the Clinton administration . . . and now these Jews are in control—they used lies spread by the media they run and committed genocide in our name." Table 3.2 gives a brief sense of what is now happening.

TABLE
3.2. LINKS AMONG "HATE SITES"

Site	Links to Like-Minded Sites	Links to Opposition
Adelaide Institute (holocaust revisionism)	16	6
Aggressive Christianity	0	0
All Men Must Die	5	0
Altar of Unholy Blasphemy	11	0
Aryan Nations	28	0
Crosstar (nationalistic)	29	0
David Duke Online	11	0
God Hates Fags	7	3
Islam Monitor	0	12
KKK.com	72	0
Martin Luther King, Jr. (revisionist view of King)	0	0
Misogyny Unlimited	92	1
National Association for the Advancement of White People	0	0
Skinheads of the Racial Holy War	100	0
Stormfront (white nationalism)	60	5
Voice of Freedom (antisemitic)	27	5
Vote for USA (antisemitic)	17	0
White Aryan Resistance	0	0
World Church of the Creator	11	0
Total (19)	14 with; 5 without	6 with; 13 without

Here in particular, the provision of opposition links is designed to produce not discussion but instead fear and contempt. Holocaust denial organizations, for example, describe their adversaries as "exterminationists" or "Holocaust enforcers" and provide links with the evident goal of discrediting them. With respect to like-minded sites, several hate groups have formal linking agreements: "You link to us and we'll link to you." One such site lists nearly one hundred such groups, each with a link, under the title "White Pride World Wide." The listed sites include European Knights of the Ku Klux Klan, German Skin Heads, Aryan Nations, Knights of the Ku Klux Klan, Siegheil88, Skinhead Pride, Intimidation One, SS Enterprises, and White Future.

We can sharpen our understanding here if we attend to the phenomenon of *group polarization*. This phenomenon raises serious questions about any system in which individuals and groups make diverse choices, and many people end up in echo chambers of their own design.

GROUP POLARIZATION IN GENERAL

The term *group polarization* refers to something very simple: *After deliberation, people are likely to move toward a more extreme point in the direction to which the group's members were originally inclined.* With respect to the Internet and new communications technologies, the implication is that groups of like-minded people, engaged in discussion with one another, will end up thinking the same thing that they thought before—but in more extreme form.

Consider some examples of the basic phenomenon, which has been found in over a dozen nations.[9]

- After discussion, a group of moderately profeminist women will become more strongly profeminist.[10]
- After discussion, citizens of France become more critical of the United States and its intentions with respect to economic aid.[11]
- After discussion, whites predisposed to show racial prejudice offer more negative responses to the question whether white racism is responsible for conditions faced by African-Americans in American cities.[12]
- After discussion, whites predisposed not to show racial prejudice offer more positive responses to the same question.[13]

The phenomenon of group polarization has conspicuous importance to the communications market, where groups with distinctive identities increasingly engage in within-group discussion. Effects of the kind just described should be expected with the Unorganized Militia and racial hate groups as well as with less extreme organizations of all sorts. If the public is balkanized and if different groups are designing their own preferred communications packages, the consequence will be not merely the same but still more balkanization, as group members move one another toward more extreme points in line with their initial tendencies. At the same time, different deliberating groups, each consisting of like-minded people, will be driven increasingly far apart, simply because most of their discussions are with one another.

Note in particular that even if most of us do not use the power to filter so as to wall ourselves off from other points of view, some or many people will do, and are doing, exactly that.

This is sufficient for polarization to occur, and to cause serious social risks. In general, it is precisely the people most likely to filter out opposing views who most need to hear such views. New technologies, emphatically including the Internet, make it easier for people to hear the opinions of like-minded but otherwise isolated others, and to isolate themselves from competing views. For this reason alone, they are a breeding ground for polarization, and potentially dangerous for both democracy and social peace.

There have been two main explanations for group polarization. Massive evidence now supports both these explanations.

The first explanation emphasizes the role of persuasive arguments. It is based on a simple intuition: Any individual's position on any issue is a function, at least in part, of which arguments seem convincing. If your position is going to move as a result of group discussion, it is likely to move in the direction of the most persuasive position defended within the group, taken as a whole.

If the group's members are already inclined in a certain direction, they will offer a disproportionately large number of arguments going in that same direction, and a disproportionately small number of arguments going the other way. As a result, the consequence of discussion will be to move people further in the direction of their initial inclinations. Thus, for example, a group whose members lean against gun control will, in discussion, provide a wide range of arguments against gun control, and the arguments made for gun control will be both fewer and weaker. The group's members, to the extent that they shift, will shift toward a more extreme position against gun control. And the group as a whole, if a group decision is required, will move not to the median position, but to a more extreme point.

On this account, the central factor behind group polarization is the existence of a *limited argument pool,* one that is skewed (speaking purely descriptively) in a particular direction. It is easy to see how shifts might happen with discussion groups on the Internet (consider a group of Democrats, or Socialists, or members of the Unorganized Militia), and indeed with individuals not engaged in discussion but consulting only ideas (on radio, television, or the World Wide Web) to which they are antecedently inclined. The tendency of such discussion groups, and such consultations, will be to entrench and reinforce preexisting positions—often resulting in extremism.

The second mechanism, involving social comparison, begins with the reasonable suggestion that people want to be perceived favorably by other group members, and also to perceive themselves favorably. Once they hear what others believe, they often adjust their positions in the direction of the dominant position. The German sociologist Elisabeth Noell-Neumann has used this idea as the foundation for a general theory of public opinion, involving a "spiral of silence," in which people with minority positions silence themselves, potentially excising those positions from society over time.[14]

Suppose, for example, that people in a certain group believe that they are sharply opposed to affirmative action, feminism, and gun control, and that they also want to *seem* to be sharply opposed to all these. If they are in a group whose members are also sharply opposed to these things, they might well shift in the direction of even sharper opposition after they see what other group members think. In countless studies, exactly this pattern is observed. Of course people will not shift if they have a clear sense of what they think and are not movable by the opinions of others. But most people, most of the time, are not so fixed in their views.

The point has important implications about the effects of exposure to ideas and claims on television, radio, and the Internet—even in the absence of a chance for interaction. Because group polarization occurs merely on the basis of exposure to the views of others, it is likely to be a common phenomenon in a balkanized speech market. Suppose, for example, that conservatives are visiting conservative Websites; that liberals are visiting liberal Websites; that environmentalists are visiting sites dedicated to establishing the risks of genetic engineering and global warming; that critics of environmentalists are visiting sites dedicated to exposing frauds allegedly perpetrated by environmentalists; that people inclined to racial hatred are visiting sites that express racial hatred. To the extent that these exposures are not complemented by exposure to competing views, group polarization will be the inevitable consequence.

THE ENORMOUS IMPORTANCE OF GROUP IDENTITY

For purposes of understanding modern technologies, a particularly important point has to do with perceptions of identity and group membership. Group polarization will significantly increase if people think of themselves, antecedently or otherwise, as part of a group having a shared identity and a degree of solidarity. If they think of themselves in this way, group polarization is both more likely and more extreme.[15] If, for example, a number of people in an Internet discussion group think of themselves as opponents of high taxes, or advocates of animal rights, or critics of the Supreme Court, their discussions are likely to move them in quite extreme directions, simply because they understand each other as part of a common cause. Similar movements should be expected for those who listen to a radio show known to be conservative, or who watch a television program dedicated to traditional religious values or to exposing white racism. Considerable evidence so suggests.[16]

Group identity is important in another way. If you are participating in an Internet discussion group, but you think that other group members are significantly different from you, you are less likely to be moved by what they say. If, for example, other group members are styled "Republicans" and you consider yourself a Democrat, you might not shift at all—even if you would indeed shift, as a result of the same arguments, if you were all styled "voters" or "jurors" or "citizens." Thus a perception of shared group identity will heighten the effect of others' views, whereas a perception of unshared identity, and of relevant differences, will reduce that effect, and possibly even eliminate it.

These findings should not be surprising. Recall that in ordinary cases, group polarization is a product of social influences and limited argument pools. If this is so, it stands to reason that when group members think of one another as similar along a salient dimension, or if some external factor (politics, geography, race, sex) unites them, group polarization will be heightened. If identity is shared, persuasive arguments

are likely to be still more persuasive; the identity of those who are making them gives them a kind of credential or boost. And if identity is shared, social influences will have still greater force. People do not like their reputations to suffer in the eyes of those who seem most like them. And if you think that group members are in some relevant sense different from you, their arguments are less likely to be persuasive, and social influences may not operate as much or at all.

GROUP POLARIZATION AND THE INTERNET

Group polarization is unquestionably occurring on the Internet. From the discussion thus far, it seems plain that the Internet is serving, for many, as a breeding ground for extremism, precisely because like-minded people are deliberating with greater ease and frequency with one another, and often without hearing contrary views. Repeated exposure to an extreme position, with the suggestion that many people hold that position, will predictably move those exposed, and likely predisposed, to believe in it. One consequence can be a high degree of fragmentation, as diverse people, not originally fixed in their views and perhaps not so far apart, end up in extremely different places, simply because of what they are reading and viewing. Another consequence can be a high degree of error and confusion.

A number of studies have shown group polarization in Internet-like settings. An especially interesting experiment finds particularly high levels of polarization when group members met relatively anonymously and when group identity was emphasized.[17] From this experiment, it is reasonable to speculate that polarization is highly likely to occur, and to be extreme, under circumstances in which group membership is made salient and people have a high degree of anonymity. These are of course characteristic features of deliberation via the Internet.[18]

Consider in this regard a revealing study not of extremism, but of serious errors within working groups, both face-to-face and more importantly online.[19] The purpose of the study was to see how groups might collaborate to make personnel decisions. Resumes for three candidates, applying for a marketing manager position, were placed before the several groups. The attributes of the candidates were rigged by the experimenters so that one applicant was clearly best matched for the job described. Packets of information were given to subjects, each containing only a subset of information from the resumes, so that each group member had only part of the relevant information. The groups consisted of three people, some operating face-to-face, some operating on-line.

Two results were especially striking. First, group polarization was common, in the sense that groups ended up in a more extreme position in line with members' predeliberation views. Second, almost none of the deliberating groups made what was conspicuously the right choice. The reason is that they failed to share information in a way that would permit the group to make an objective decision. In online groups, the level of mistake was especially high, for the simple reason that members tended to share positive information about the emerging winning candidate and negative information about the losers, while also suppressing negative information about the emerging winner and positive information

about the emerging losers. These contributions served to "reinforce the march toward group consensus rather than add complications and fuel debate."[20] In fact this tendency was *twice* as large within the online groups. There is a warning here about the consequences of the Internet for democratic deliberation.

FRAGMENTATION, POLARIZATION, RADIO, AND TELEVISION

An understanding of group polarization casts light on the potential effects not only of the Internet but also of radio and television, at least if stations are numerous and many take a well-defined point of view. Recall that mere exposure to the positions of others creates group polarization. It follows that this effect will be at work for nondeliberating groups, in the form of collections of individuals whose communications choices go in the same direction, and who do not expose themselves to alternative positions. Indeed the same process is likely to occur for newspaper choices. General interest intermediaries have a distinctive role here, by virtue of their effort to present a wide range of topics and views. Polarization is far less likely to occur when such intermediaries dominate the scene. A similar point can be made about the public forum doctrine. When diverse speakers have access to a heterogeneous public, individuals and groups are less likely to be able to insulate themselves from competing positions and concerns. Fragmentation is correspondingly less likely.

Group polarization also raises more general issues about communications policy. Consider the "fairness doctrine," now largely abandoned but once requiring radio and television broadcasters to devote time to public issues and to allow an opportunity for those with opposing views to speak. The latter prong of the doctrine was designed to ensure that listeners would not be exposed to any single view—if one view was covered, the opposing position would have to be allowed a right of access. When the Federal Communications Commission abandoned the fairness doctrine, it did so on the ground that this second prong led broadcasters, much of the time, to avoid controversial issues entirely, and to present views in a way that suggested a bland uniformity. Subsequent research has suggested that the elimination of the fairness doctrine has indeed produced a flowering of controversial substantive programming, frequently expressing extreme views of one kind or another; consider talk radio.[21]

Typically this is regarded as a story of wonderfully successful deregulation. The effects of eliminating the fairness doctrine were precisely what was sought and intended. But from the standpoint of group polarization, the evaluation is far more complicated. On the good side, the existence of diverse pockets of opinion would seem to enrich society's total argument pool, potentially to the benefit of all of us. At the same time, the growth of a wide variety of issues-oriented programming—expressing strong, often extreme views, and appealing to dramatically different groups of listeners and viewers—is likely to create group polarization. All too many people are now exposed largely to louder echoes of their own voices, resulting, on occasion, in misunderstanding and enmity. Perhaps it is better for people to hear fewer controversial views than for them to hear a single such view, stated over and

over again. At least there is a risk, in the current situation, that too many people will be insulated from exposure to views that are more moderate, or extreme in another direction, or in any case different from their own.

IS GROUP POLARIZATION BAD? OF ENCLAVE DELIBERATION

Of course we cannot say, from the mere fact of group polarization, that there has been a movement in the wrong direction. Notwithstanding some of the grotesque examples given here, the more extreme tendency might be better rather than worse. Indeed, group polarization helped fuel many movements of great value—including, for example, the civil rights movement, the antislavery movement, and the movement for sex equality. Each of these movements was extreme in its time, and within-group discussion certainly bred greater extremism; but extremism should not be a word of opprobrium. If greater communications choices produce greater extremism, society may, in many cases, be better off as a result. One reason is that when many different groups are deliberating with one another, society will hear a far wider range of views as a result. Even if the "information diet" of many individuals is homogeneous or insufficiently diverse, society as a whole might have a more richer and fuller set of ideas. This is another side of the general picture of social fragmentation. It suggests some large benefits from pluralism and diversity—benefits even if individuals customize and cluster in groups.

We might define *enclave deliberation* as that form of deliberation that occurs within more or less insulated groups, in which like-minded people speak mostly to one another. The Internet, along with other new communications options, makes it much easier to engage in enclave deliberation. It is obvious that enclave deliberation can be extremely important in a heterogeneous society, not least because members of some groups tend to be especially quiet when participating in broader deliberative bodies. In this light, a special advantage of enclave deliberation is that it promotes the development of positions that would otherwise be invisible, silenced, or squelched in general debate. The efforts of marginalized groups to exclude outsiders, and even of political parties to limit their primaries to party members, might be justified in similar terms. Even if group polarization is at work—perhaps *because* group polarization is at work—enclaves, emphatically including those produced by new technologies, can provide a wide range of social benefits, not least because they greatly enrich the social "argument pool."

The central empirical point here is that in deliberating bodies, high-status members tend to speak more than others, and their ideas are more influential—partly because low-status members lack confidence in their own abilities, and partly because they fear retribution.[22] For example, women's ideas are often less influential and sometimes are "suppressed altogether in mixed-gender groups,"[23] and in ordinary circumstances, cultural minorities have disproportionately little influence on decisions by culturally mixed groups. In light of the inevitable existence of some status-based hierarchies, it makes sense to be receptive to deliberating enclaves in which members of multiple groups may speak with one another and develop their views. The Internet is and will continue to be particularly valuable insofar as it makes this easier.

But there is also a serious danger in such enclaves. The danger is that through the mechanisms of social influence and persuasive arguments, members will move to positions that lack merit but are predictable consequences of the particular circumstances of enclave deliberation. In the extreme case, enclave deliberation may even put social stability at risk. And it is impossible to say, in the abstract, that those who sort themselves into enclaves will generally move in a direction that is desirable for society at large or even for its own members. It is easy to think of examples to the contrary, as, for example, in the rise of Nazism, hate groups, and cults of various sorts.

ENCLAVES AND A PUBLIC SPHERE

Whenever group discussion tends to lead people to more strongly held versions of the same view with which they began, there is legitimate reason for concern. This does not mean that the discussions can or should be regulated. But it does raise questions about the idea that "more speech" is necessarily an adequate remedy for bad speech—especially if many people are inclined and increasingly able to wall themselves off from competing views. In democratic societies, the best response is suggested by the public forum doctrine, whose most fundamental goal is to increase the likelihood that at certain points, there is an exchange of views between enclave members and those who disagree with them. It is total or near-total self-insulation, rather than group deliberation as such, that carries with it the most serious dangers, often in the highly unfortunate (and sometimes literally deadly) combination of extremism with marginality.

To explore some of the advantage of heterogeneity, let us engage in a thought experiment. Imagine a deliberating body consisting not of a subset of like-minded people but of all citizens in the relevant group; this may mean all citizens in a community, a state, a nation, even the world. Imagine that through the magic of the computer, everyone can talk to everyone else. By hypothesis, the argument pool would be very large. It would be limited only to the extent that the set of citizen views was similarly limited. Of course social influences would remain. If you are one of a small minority of people who deny that global warming is a serious problem, you might decide to join the crowd. But when deliberation reveals to people that their private position is different, in relation to the group, from what they thought it was, any shift would be in response to an accurate understanding of all relevant citizens, and not a product of a skewed sample.

This thought experiment does not suggest that a fragmented or balkanized speech market is always bad or that the hypothesized, all-inclusive deliberating body would be ideal. It would be foolish to suggest that all discussion should occur, even as an ideal, with all others. The great benefit of deliberating enclaves is that positions may emerge that otherwise would not, and that deserve to play a larger role both within the enclave and within the heterogeneous public. Properly understood, the case for deliberating enclaves is that they will improve social deliberation, democratic and otherwise, precisely because enclave deliberation is often required for incubating

new ideas and perspectives that will add a great deal to public debate. But for these improvements to occur, members must not insulate themselves from competing positions, or at least any such attempts at insulation must not be a prolonged affair. The effects of group polarization thus show that with respect to communications, consumer sovereignty might well produce serious problems for individuals and society at large—and these problems will occur by a kind of iron logic of social interactions.

NO POLARIZATION AND DEPOLARIZATION

Group polarization is a common phenomenon. But in certain circumstances, it can be decreased, increased, or even eliminated. Recall that no shift should be expected from people who are confident that they know what they think, and who are simply not going to be moved by what they hear from other people. If, for example, you are entirely sure of your position with respect to nuclear power—if you are confident not only of your precise view but of the certainty with which you ought to hold it—the positions of other people will not affect you. People of this sort will not shift by virtue of any changes in the communications market.

With artful design of deliberating groups, moreover, it is possible to produce *depolarization*—shifts, within groups, toward the middle of the extremes. Suppose, for example, that a group of twelve people is constructed so as to include six people who have one view and six people who think the opposite—for example, half of the group's members believe that global warming is a serious problem, while the other half think that it is not. If most of the members do not have entirely fixed positions, there is likely to be real movement toward the middle. The persuasive arguments view helps explain why this is so. By hypothesis, the "argument pool" includes an equal number of claims both ways.

There is a valuable lesson about possible uses of communications technologies to produce convergence, and possibly even learning, among people who disagree with one another. If people hear a wide range of arguments, they are likely to be moved in the direction of those who disagree with them, at least if the arguments are reasonable.

CYBERCASCADES: INFORMATION AS WILDFIRE, AND TIPPING POINTS

The phenomenon of group polarization is closely related to the widespread phenomenon of *social cascades*. No discussion of social fragmentation and emerging communications technologies would be complete without an understanding of cascades—above all because they become more likely when information, including false information, can be spread to hundreds, thousands, or even millions by the simple press of a button.

It is obvious that many social groups, both large and small, move rapidly and dramatically in the direction of one or another set of beliefs or actions.[24] These sorts of cascades typically involve the spread of information; in fact they are usually driven by information. Most of us lack direct or entirely reliable information about many matters of importance—whether global warming is a serious problem, whether there is a risk of war in India, whether a lot of sugar is really bad for you, whether Mars really exists and what it is like. If you lack a great deal of private information, you might well rely on information provided by the statements or actions of others. A stylized example: If Joan is unaware whether abandoned toxic waste dumps are in fact hazardous, she may be moved in the direction of fear if Mary thinks that fear is justified. If Joan and Mary both believe that fear is justified, Carl may end up thinking so too, at least if he lacks reliable independent information to the contrary. If Joan, Mary, and Carl believe that abandoned hazardous waste dumps are hazardous, Don will have to have a good deal of confidence to reject their shared conclusion. And if Joan, Mary, Carl, and Don present a united front on the issue, others may well go along.

The example shows how information travels and can become quite widespread and entrenched, even if it is entirely wrong. An illustration is, in fact, the widespread popular belief that abandoned hazardous waste dumps rank among the most serious environmental problems; science does not support that belief, which seems to have spread via cascade.[25] Some cascades are widespread but local; consider the view, with real currency in some African-American communities, that white doctors are responsible for the spread of AIDS among African-Americans. One group may end up believing something and another group the exact opposite, and the reason is the rapid transmission of information within one group but not the other.

It should be obvious that the Internet, with Websites containing information designed for particular groups, greatly increases the likelihood of diverse but inconsistent cascades. "Cybercascades" occur every day. Many of us have been deluged with e-mail involving the need to contact our representatives about some bill or other—only to learn that the bill did not exist and the whole problem was a joke or a fraud. Even more of us have been earnestly warned about the need to take precautions against viruses that do not exist. And many thousands of hours of Internet time have been spent on elaborating paranoid claims about alleged nefarious activities, including murder, on the part of President Clinton. A number of sites and discussion groups spread rumors and conspiracy theories of various sorts. "Electrified by the Internet, suspicions about the crash of TWA Flight 800 were almost instantly transmuted into convictions that it was the result of friendly fire. . . . It was all linked to Whitewater. . . . Ideas become E-mail to be duplicated and duplicated again."[26] In 2000, an e-mail rumor specifically aimed at African Americans alleged that "No Fear" bumper stickers bearing the logo of the sportswear company of the same name really promote a racist organization headed by former Ku Klux Klan Grand Wizard David Duke. (If you're interested in more examples, you might consult http://urbanlegends.about.com, a Website dedicated to widely disseminated falsehoods, many of them spread via the Internet.)

As an especially troublesome example, consider widespread doubts in South Africa, where about 20 percent of the adult population is infected by the AIDS virus, about the connection between HIV and AIDS. South African President Mbeki is a well-known Internet surfer, and he learned the views of the "denialists" after stumbling across one of their

Websites. The views of the "denialists" are not scientifically respectable—but to a nonspecialist, many of the claims on their (many) sites seem quite plausible. At least for a period, President Mbeki both fell victim to a cybercascade and, through his public statements, helped to accelerate one, to the point where many South Africans at serious risk are not convinced about an association between HIV and AIDS. It remains to be seen to what extent this cascade effect will turn out to be literally deadly.

With respect to information in general, there is even a "tipping point" phenomenon, creating a potential for dramatic shifts in opinion. After being presented with new information, people typically have different thresholds for choosing to believe or do something new or different. As the more likely believers, that is people with low thresholds, come to a certain belief or action, people with somewhat higher thresholds then join them, soon producing a significant group in favor of the view in question. At that point, those with still higher thresholds may join, possibly to a point where a critical mass is reached, making large groups, societies, or even nations "tip."[27] The result of this process can be to produce snowball or cascade effects, as large groups of people end up believing something—whether or not that something is true—simply because other people, in the relevant community, seem to believe that it is true.

There is a great deal of experimental evidence of informational cascades, which are easy to induce in the laboratory[28]; real world phenomena also have a great deal to do with cascade effects. Consider, for example, going to college, smoking, participating in protests, voting for third-party candidates, striking, recycling, filing lawsuits, using birth control, rioting, even leaving bad dinner parties.[29] In each of these cases, people are greatly influenced by what others do. Often a tipping point will be reached. The Internet is an obvious breeding ground for cascades, and as a result thousands or even millions of people, consulting sources of a particular kind, will believe something that is quite false.

The good news is that the Internet can operate to debunk false rumors as well as to start them. But at the same time, the opportunity to spread apparently credible information to so many people can induce fear, error, and confusion, in a way that threatens many social goals, including democratic ones. As we have seen, this danger takes on a particular form in a balkanized speech market, as local cascades lead people in dramatically different directions. When this happens, correctives, even via the Internet, may not work, simply because people are not listening to one another.

A CONTRAST: THE DELIBERATIVE OPINION POLL

By way of contrast to fragmentation and cybercascades, consider some work by James Fishkin, a creative political scientist at the University of Texas, who has pioneered a genuine social innovation: the deliberative opinion poll.[30] The basic idea is to ensure that polls are not mere "snapshots" of public opinion. Instead people's views are recorded only after diverse citizens, with different points of view, have actually been brought together to discuss topics with one another. Deliberative opinion polls have now been conducted in several nations, including the United States, England, and Australia. It is even possible for deliberative opinion polls to be conducted on the Internet, and Fishkin has initiated experiments in this direction.

In deliberative opinion polls, Fishkin finds some noteworthy shifts in individual views. But he does not find a systematic tendency toward polarization. In England, for example, deliberation led to reduced interest in using imprisonment as a tool for combating crime.[31] The percentage believing that "sending more offenders to prison" is an effective way to prevent crime decreased from 57 percent to 38 percent; the pecentage believing that fewer people should be sent to prison increased from 29 percent to 44 percent; belief in the effectiveness of "stiffer sentences" was reduced from 78 percent to 65 percent.[32] Similar shifts were shown in the direction of greater enthusiasm for procedural rights of defendants and increased willingness to explore alternatives to prison.

In other experiments with the deliberative opinion poll, shifts included a mixture of findings, with deliberation leading larger percentages of individuals to conclude that legal pressures should be increased on fathers for child support (from 70% to 85%) and that welfare and health care should be turned over to the states (from 56% to 66%).[33] To be sure, the effect of deliberation was sometimes to create an increase in the intensity with which people held their preexisting convictions.[34] These findings are consistent with the prediction of group polarization. But this was hardly a uniform pattern. On some questions, deliberation shifted a minority position to a majority position (with, for example, a jump from 36 percent to 57 percent favoring policies making divorce "harder to get").[35]

Fishkin's experiments have some distinctive features. They involve not like-minded people, but diverse groups of citizens engaged in discussion after being presented, by appointed moderators, with various sides of social issues. In many ways these discussions provide a model for civic deliberation, complete with reason-giving and political equality. Of course it can be expensive to transport diverse people to the same place. But new communications technologies make the idea of a deliberative opinion poll and of reasoned discussion among heterogeneous people far more feasible—even if private individuals, in their private capacity, would rarely choose to create deliberating institutions on their own. Indeed, Fishkin is now attempting to create deliberative opinion polls on the Internet. There are many efforts and experiments in this general vein.[36]

Here we can find considerable promise for the future, in the form of discussions among diverse people who exchange reasons and who would not, without new technologies, be able to talk with one another at all. If we are guided by the notion of consumer sovereignty, and if we celebrate unlimited filtering, we will be unable to see why the discussions in the deliberative opinion poll are a great improvement over much of what is now happening on the Internet. In short, republican aspirations sharply diverge from the ideal of consumer sovereignty, seeing television as "just another appliance" and dreaming of a future in which, in Gates's words, "you'll be able to just say what you're interested in, and have the screen help you pick out a video that you care about."

The real questions are what sort of ideals we want to animate our choices, and what kinds of attitudes, and regulation, we want in light of that judgment. And here it is important to say that in themselves, new technologies are not biased in favor of homogeneity and deliberation among like-minded people. Everything depends on what people seek to do with the new opportunities that they have. "I've been in chat rooms where I've observed, for the first time in my life, African-Americans and white supremacists talking to each other. . . . [I]f you go through the threads of the conversation, by the end you'll find there's less animosity than there was at the beginning. It's not pretty sometimes . . . [b]ut here they are online, actually talking to each other."[37] The problem is that this remains an unusual practice.

OF DANGERS AND SOLUTIONS

I hope that I have shown enough to demonstrate that for citizens of a heterogeneous democracy, a fragmented communications market creates considerable dangers. There are dangers for each of us as individuals; constant exposure to one set of views is likely to lead to errors and confusions, sometimes as a result of cybercascades. And to the extent that the process entrenches existing views, spreads falsehood, promotes extremism, and makes people less able to work cooperatively on shared problems, there are dangers for society as a whole.

To emphasize these dangers, it is unnecessary to claim that people do or will receive all of their information from the Internet. There are many sources of information, and some of them will undoubtedly counteract the risks I have discussed. Nor is it necessary to predict that most people will speak only with those who are like-minded. Of course many people will seek out competing views. But when technology makes it easy for people to wall themselves off from others, there are serious risks, for the people involved and for society generally.

To be sure, we do not yet know whether anything can or should be done about fragmentation and excessive self-insulation. I will take up that topic in due course. For purposes of obtaining understanding, few things are more important than to separate the question whether there is a problem from the question whether anything should be done about it. Dangers that cannot be alleviated continue to be dangers. They do not go away if or because we cannot, now or ever, think of decent solutions. It is much easier to think clearly when we appreciate that fact.

Intermediaries and Hate Speech: Fostering Digital Citizenship for our Information Age

Danielle Keats Citron
University of Maryland School of Law

Helen Norton
University of Colorado School of Law

No. 2011 - 16

ARTICLE

INTERMEDIARIES AND HATE SPEECH: FOSTERING DIGITAL CITIZENSHIP FOR OUR INFORMATION AGE

DANIELLE KEATS CITRON & HELEN NORTON*

* Danielle Citron is Professor of Law at the University of Maryland School of Law and an Affiliate Fellow of the Stanford Center on Internet and Society and of the Yale Information Society Project. Helen Norton is Associate Dean for Academic Affairs and Associate Professor of Law at the University of Colorado School of Law. This Article grew out of a true partnership with both authors contributing equally. The Authors wish to thank the organizers of the Stanford Center on Internet and Society's Roundtable on Hate Speech; Columbia University Computer Science Department's Distinguished Lecture Series; Stanford Law School's 7th Annual E-Commerce conference; Washington University School of Law's Faculty Workshop Series; and the Progress and Freedom Foundation's panel on Censorship, Hate Speech, and Dissent for the chance to present early versions of this work. Rachel Arnow-Richman, Brad Bernthal, Richard Boldt, Ryan Calo, Miriam Cherry, Julie Cohen, Lisa Fairfax, Lauren Gelman, Don Gifford, Michelle Harner, Melissa Hart, Leslie Meltzer Henry, Deborah Hellman, Sonia Katyal, R.A. Lenhardt, Steven Morrison, Scott Moss, Patrick O'Donnell, Paul Ohm, Frank Pasquale, Neil Richards, Robert Rhee, Wendy Seltzer, Steve Sheinberg, Jana Singer, Catherine Smith, Daniel Solove, David Super, Berin Szoka, Alexander Tsesis, Barbara van Schewick, Chris Wolf, and the participants at University of Colorado and University of Maryland Law School faculty colloquia gave us terrific suggestions. We are grateful to the numerous safety officers, communication directors, and in-house counsel who talked to us about their approach to hate speech – their help was invaluable. Alice Johnson, Ovais Anwar, John Buchanan, Jordan Bunch, David Cline, Adam Farra, Matt Haven, Lindsey Lanzendorfer, and Susan McCarty provided excellent research assistance. Molly Carter and the editors at the *Boston University Law Review* gave us invaluable feedback and support.

No longer confined to isolated corners of the web, cyber hate now enjoys a major presence on popular social media sites. The Facebook group Kill a Jew Day, *for instance, acquired thousands of friends within days of its formation, while YouTube has hosted videos with names like* How to Kill a Beaner, Execute the Gays, *and* Murder Muslim Scum. *The mainstreaming of cyber hate has the troubling potential to shape public expectations of online discourse.*

Internet intermediaries have the freedom and influence to seize this defining moment in cyber hate's history. We believe that a thoughtful and nuanced intermediary-based approach to hate speech can foster respectful and vibrant online discourse. We urge intermediaries to help address cyber hate by adopting accessible and transparent policies that educate users about their rights and responsibilities as digital citizens. Intermediaries' options include challenging hateful speech by responding with counter-speech and empowering community members to enforce norms of digital citizenship.

INTRODUCTION

The Facebook group *Kill a Jew Day* declared July 4, 2010 as the start of an eighteen-day period of violence "anywhere you see a Jew."[1] The group's profile featured a swastika and images of corpses piled on top of one another.[2] Group members commented that they could not "wait to rape the dead baby Jews."[3]

The *Kill a Jew Day* social network group is an example of the more than 11,000 websites, videos, and social network groups devoted to spreading hate.[4]

[1] Yaakov Lappin, *'Kill a Jew' Page on Facebook Sparks Furor*, JERUSALEM POST, July 5, 2010, at 5.

[2] *Id.*

[3] *Id.*

[4] Jesse Solomon, *Hate Speech Infiltrates Social-Networking Sites, Report Says*, CNN (Mar. 15, 2010, 4:37 PM), http://www.cnn.com/2010/TECH/03/15/hate.speech.social.

Neo-Nazi websites allow users to maneuver virtual nooses over digital images of black men.[5] Videos posted online urge viewers to murder "Muslim scum"[6] and to kill homosexuals.[7] Typing "I hate spics" into Google generates 45,300 results.[8]

The greatest increase in digital hate has occurred on social media sites.[9] Examples include the *How to Kill a Beaner* video posted on YouTube, which allowed players to kill Latinos while shouting racial slurs,[10] and the Facebook group *Kick a Ginger Day*, which inspired physical attacks on students with red hair.[11] Facebook has hosted groups such as *Hitting Women*,[12] *Holocaust Is a Holohoax*,[13] and *Join if you hate homosexuals*.[14]

networks/index.html.

[5] Maria Seminerio, *"Hate Filter" Tackles Racist Sites*, ZDNET (Nov. 12, 1998, 3:29 PM), http://www.zdnet.co.uk/news/networking/1998/11/12/us-report-andquothatefilterand quot-tackles-racist-sites-2069870/.

[6] Mark MacAskill & Marcello Mega, *YouTube Cuts Murder Race-Hate Clips*, SUNDAY TIMES (London) (Sept. 28, 2008), http://www.timesonline.co.uk/tol/news/uk/scotland/ article4837923.ece.

[7] Theresa Howard, *Online Hate Speech: It's Difficult to Police*, USA TODAY, Oct. 2, 2009, at 4D.

[8] Petition for Inquiry Filed on Behalf of the National Hispanic Media Coalition at 10, In the Matter of Hate Speech in the Media, Before the F.C.C., Jan. 28, 2009.

[9] *See generally* SIMON WIESENTHAL CENTER, FACEBOOK, YOUTUBE + HOW SOCIAL MEDIA OUTLETS IMPACT DIGITAL TERRORISM AND HATE (2009) (providing screenshots of social media websites promoting hate). Hate groups recruit new members on popular social network sites like YouTube and Facebook. *Social Networks Are New Sites for Hate Speech*, REUTERS, May 13, 2009, http://www.pcmag.com/article2/0,2817,2347004,00.asp.

[10] aborn88, *How to Kill a Beaner*, YOUTUBE (June 1, 2008), http://www.youtube.com/ watch?v=Dq-tUPOGp8w.

[11] Liz Nordlinger, *Cartman Started It*, ST. PETERSBURG TIMES (Fla.), Feb. 25, 2010, at 8; Matthew Moore, *Facebook 'Kick a Ginger' Campaign Prompts Attacks on Redheads*, TELEGRAPH (U.K.) (Nov. 22, 2008, 12:47 AM), http://www.telegraph.co.uk/news/world news/northamerica/canada/3498766/Facebook-Kick-a-Ginger-campaign-prompts-attacks-on-redheads.html.

[12] Phil Bradley, *Facebook Group: Hitting Women*, PHIL BRADLEY'S WEBLOG (Feb. 18, 2010), http://philbradley.typepad.com/phil_bradleys_weblog/2010/02/facebook-group-hitting-women.html (reporting that as of February 10, 2010 the Facebook page *Hitting Women* remained on Facebook); Julie Ross Godar, *Facebook and Hate Speech: Are You a Fan of Hitting Women?*, BLOGHER (Feb. 18, 2010, 5:39 PM). As of December 20, 2010, the Facebook group *Hitting Women* was no longer available.

[13] Corilyn Shropshire, *Facebook Wrestles with Anti-Semitism*, HOUS. CHRON., May 15, 2009, at 6.

[14] David Badash, *Facebook or Hate Book? Facebook Shuts Down Anti-Gay Hate Groups!*, THE NEW CIVIL RIGHTS MOVEMENT (Mar. 9, 2010), http://thenewcivilrights movement.com/facebook-or-hate-book-facebook-shuts-down-anti-gay-hate-groups/successes/2010/03/09/8828.

Groups recognized cyberspace's potential to facilitate hate from its earliest days. In 1984, for example, the Aryan Nation sponsored a Usenet bulletin board featuring a "hit list," which included among its targets Alan Berg, a Jewish radio talk show host who had ignited the anger of the Order, an Aryan Nation spin-off group, by ridiculing the group on air.[15] Members of the Order murdered Berg in his driveway after the posting of the hit list.[16]

Even though cyber hate is not a new phenomenon, its recent growth is startling.[17] No longer isolated in little-known bulletin boards and websites, digital hate appears in the internet's mainstream. Digital hate's prevalence has considerable – and troubling – potential to shape public expectations of online discourse, especially as cyber hate penetrates social media populated with the young and impressionable. We thus face an important point in cyber hate's history and development: norms of subordination may overwhelm those of equality if hatred becomes an acceptable part of online discourse.

For these reasons, some scholars support governmental intervention to combat digital hate.[18] Governmental efforts to regulate hate speech based on its content, however, trigger important First Amendment and other concerns.[19] Given the challenges faced by regulatory solutions to the problem of digital hate, this Article focuses instead on the potential role of online intermediaries – private entities that host or index online content – in voluntarily addressing

[15] *See The Murder of Alan Berg: 25 Years Later*, DENVER POST, June 18, 2009, at A-01.

[16] *Id.*

[17] The Simon Wiesenthal Center has documented the extraordinary increase in online hate over the past ten years. SIMON WIESENTHAL CENTER, *supra* note 9, at 1 (discussing its growth from one website in 1995 to 10,000 today and the 25% increase in sites devoted to hate in the past year alone). This growth mirrors the escalating power and range of the technologies that facilitate the distribution of expression generally, including but not limited to digital hate. *See* Nathan Myhrvold, *Moore's Law Corollary: Pixel Power*, N.Y. TIMES, June 7, 2006, at G3 (explaining that the speed and breath of computing power doubles every eighteen months). As Microsoft founder Bill Gates explains of the information age, "we're always in a time of utter change, maybe even accelerating change." John Markoff, *Gates's Lieutenants Look Ahead, Hoping to Avoid Other Companies' Mistakes*, N.Y. TIMES, June 17, 2006, at C1.

[18] For various proposals to modify the First Amendment standards to be applied to governmental regulation of online hate speech, see Jennifer L. Brenner, *True Threats – A More Appropriate Standard for Analyzing First Amendment Protection and Free Speech When Violence Is Perpetrated over the Internet*, 78 N.D. L. REV. 753, 783 (2002); John P. Cronan, *The Next Challenge for the First Amendment: The Framework for an Internet Incitement Standard*, 51 CATH. U. L. REV. 425, 428 (2002); Nancy S. Kim, *Web Site Proprietorship and Online Harassment*, 2009 UTAH L. REV. 993, 997 (urging courts to impose tort liability upon website sponsors "for creating unreasonable business models" by failing to adopt "reasonable measures" to prevent foreseeable harm of online harassment).

[19] *See, e.g.*, R.A.V. v. City of St. Paul, 505 U.S. 377, 391 (1992) (holding that a city ordinance that prohibited expression that "arouses anger, alarm or resentment in others . . . on the basis of race, color, creed, religion, or gender" impermissibly discriminated on the basis of viewpoint in violation of the First Amendment (internal quotation marks omitted)).

cyber hate and its attendant harms.[20] Internet intermediaries[21] wield considerable control over what we see and hear today, akin to that of influential cable television and talk radio shows. Examples include search engines like Google, Microsoft, and Yahoo!; browsers like Mozilla; social network sites like Facebook, MySpace, and Formspring.me; micro-blogging services like Twitter; video-sharing sites like YouTube; and newsgathering services like Digg.[22] As more and more expression appears online, these intermediaries increasingly impact the flow of information.

Importantly, intermediaries have enormous freedom in choosing whether and how to challenge digital hate, as intermediaries' response to online speech remains largely free from legal constraint in the United States.[23] Not only are intermediaries free from First Amendment concerns as private actors, they are also statutorily immunized from liability for publishing content created by others as well as for removing that content.[24]

[20] Although this Article focuses only on private intermediaries' voluntary responses to cyber hate, we do not discount the possibility that government might have a role to play regarding the perpetrators of digital hate in at least some circumstances. *See, e.g.*, Danielle Keats Citron, *Cyber Civil Rights*, 89 B.U. L. REV. 61, 86-95 (2009) [hereinafter Citron, *Cyber Civil Rights*] (exploring law's coercive role in deterring and remedying cyber harassment); Danielle Keats Citron, *Law's Expressive Value in Combating Cyber Gender Harassment*, 108 MICH. L. REV. 373, 404-14 (2009) [hereinafter Citron, *Law's Expressive Value*] (documenting the expressive value of a cyber civil rights agenda in addressing cyber gender harassment).

[21] *See* David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373, 386 (2010) (suggesting that intermediaries fall into three general categories: communication conduits, content hosts, and search/application providers). This Article focuses on voluntary measures available to a specific subset of internet intermediaries – content hosts and search/application providers – given their unique role in hosting online communities and in linking individuals to them. We leave to the side intermediaries that primarily serve as communication conduits (such as internet service and broadband providers) that we see as more akin to the phone company or the postal service in that they carry, but do not typically mediate, expressive content.

[22] This Article refers to sites that enable the production and sharing of digital content in mediated social settings as "social media." *See* Danielle Keats Citron, *Fulfilling Government 2.0's Promise with Robust Privacy Protections*, 78 GEO. WASH. L. REV. 822, 824 n.12 (2010) (explaining that social media include social-network sites, video-sharing sites, photo-sharing sites, and the like).

[23] This is not necessarily true outside of the United States, as many countries' laws require intermediaries to moderate content and to ensure its compliance with substantive restrictions. *See, e.g.*, Wendy Seltzer, Remarks, *The Politics of the Internet*, 102 AM. SOC'Y INT'L L. PROC. 45, 45-47 (2008) (describing how some countries require internet service providers, search engines, and other intermediaries to prevent in-country users from reaching certain sites).

[24] *See infra* note 111 and accompanying text.

This situation invites important and challenging questions about whether and how intermediaries might thoughtfully exercise their freedom and influence to shape on-line expression. Indeed, a number of intermediaries have begun to consider such questions, motivated by concerns about the potential business, ethical, and instrumental costs of digital hate. This has led many intermediaries to include hate speech prohibitions in their Terms of Service (TOS) agreements and Community Guidelines.

This Article proposes that intermediaries who feel a responsibility to challenge digital hate might also understand that responsibility to include fostering digital citizenship.[25] As we use the term in this Article, a commitment to digital citizenship seeks to protect users' capability to partake freely in the internet's diverse political, social, economic, and cultural opportunities, which informs and facilitates their civic engagement.[26] In short, a commitment to digital citizenship aims to secure robust *and* responsible participation in online life.

Intermediaries can foster digital citizenship by inculcating norms of respectful, vigorous engagement.[27] Just as law can be an "omnipresent

[25] For arguments that intermediaries can also play a central role in responding to defamation and other reputational harms, see David S. Ardia, *Reputation in a Networked World: Revisiting the Social Foundations of Defamation Law*, 45 HARV. C.R.-C.L. L. REV. 261, 264 (2010); Daniel H. Kahn, *Social Intermediaries: Creating a More Responsible Web Through Portable Identity, Cross-Web Reputation, and Code-Backed Norms*, 11 COLUM. SCI. & TECH. L. REV. 176, 195-96 (2010).

[26] *See* Jennifer Gordon & R.A. Lenhardt, *Rethinking Work and Citizenship*, 55 UCLA L. REV. 1161, 1185 (2008) (arguing that citizenship requires a person's ability to participate in society in a meaningful manner).

[27] In Lawrence Lessig's estimation, social norms may often regulate behavior as effectively as law. Lawrence Lessig, *The Law of the Horse: What Cyberlaw Might Teach*, 113 HARV. L. REV. 501, 507 (1999); Lawrence Lessig, *The New Chicago School*, 27 J. LEGAL STUD. 661, 669-70 (1998) [hereinafter Lessig, *The New Chicago School*] (explaining that institutions can shape behavior through the development of social norms, as well as through law, markets, and architecture). Nancy Kim similarly characterizes cyber harassment primarily as a failure of website operators' business norms, suggesting that tort law should encourage operators to engage in a range of preventive behaviors to deter online harassment. *See* Kim, *supra* note 18, at 996.

Here we note, but do not take part in, the debate over whether norms or law are more appropriate or effective in the context of internet governance. *Compare* Ardia, *supra* note 25, at 264 (proposing that online community governance through norms may often better protect users from reputational harms than defamation law), *and* Kahn, *supra* note 25, at 195-96 ("[W]e should account for the coming growth of norms in our decisions about when and how to regulate, as the new growth of norms may sometimes obviate (or occasionally exacerbate) the need for regulation."), *with* Mark A. Lemley, *The Law and Economics of Internet Norms*, 73 CHI.-KENT L. REV. 1257, 1260-61 (1998) (questioning the claim that reliance on norms is more effective than regulation in achieving cyberspace governance), *and* Neil Weinstock Netanel, *Cyberspace Self-Governance: A Skeptical View from Liberal Democratic Theory*, 88 CALIF. L. REV. 395, 401-02 (2000) (concluding that selective teacher,"[28] intermediaries' voluntary actions can educate users about acceptable behavior. Their inaction in the face of online hate plays a similar role: intermediaries' silence can send a powerful message that targeted group members are second-class citizens.[29]

Specifically, we suggest that intermediaries can valuably advance the fight against digital hate with increased transparency – e.g., by ensuring that their efforts to define and proscribe hate speech explicitly turn on the harms to be targeted and prevented. This requires intermediaries to engage in thoughtful conversations with stakeholders externally and internally to identify the potential harms of hate speech (and its constraint) that *they* find most troubling. The more intermediaries and their users understand why a particular policy regulates a certain universe of speech, the more likely they can apply that policy in a way that achieves those objectives.

Not only can well-developed and transparent policies effectively acknowledge and address the meaningful distinctions between hate speech and other expression, intermediaries may also respond to hateful speech in ways other than simply removing it. Indeed, intermediaries' choices among available options – removing speech, responding with counter-speech, and empowering and educating community members to advance norms of digital citizenship themselves – may reflect the varying ways in which their different activities might facilitate the spread of online hate and thus undermine digital citizenship.

To be sure, self-governance is not without its shortcomings.[30] But because regulatory approaches to cyber hate are largely unavailable due to First Amendment constraints, intermediaries' voluntary efforts permit the development of flexible and nuanced solutions tailored to specific contexts.[31]

governmental regulation of cyberspace will better realize liberal democratic ideals than cyberspace self-governance). In evaluating comparative costs and benefits, that debate largely assumes the freedom to choose between law and norms in a particular context. This Article, in contrast, focuses on a context where such a choice is often unavailable because law – i.e., government regulation of online hate speech – is constrained by the First Amendment.

[28] Olmstead v. United States, 277 U.S. 438, 485 (1928) (Brandeis, J., dissenting).

[29] MARY ANN GLENDON, RIGHTS TALK: THE IMPOVERISHMENT OF POLITICAL DISCOURSE 101-05 (1991) (exploring how silence can provide misleading lessons about social responsibility ethos).

[30] *See, e.g.*, Lemley, *supra* note 27, at 1260-61 (discussing limitations of reliance on norms for cyberspace governance); Netanel, *supra* note 27, at 401-02 (discussing the advantages of selective governmental regulation of cyberspace over cyberspace self-governance).

[31] Joanne Scott & Susan Sturm, *Courts as Catalysts: Re-thinking the Judicial Role in New Governance*, 13 COLUM. J. EUR. L. 565, 566 (2006) ("New governance moves away from the idea of specific rights elaborated by formal legal bodies and enforced by judicially imposed sanctions. It locates responsibility for law-making in deliberative processes which are to be continually revised by participants in light of experience, and provides for

Scholars have suggested that "soft" approaches may be especially helpful when addressing issues that are particularly complex and politically intractable.[32] This is certainly true of hate speech, which involves challenging clashes between key commitments to free expression, autonomy, equality, and dignity. Soft approaches also promote solutions that reflect intermediaries' different business models, which offer varying services from which users can choose.[33]

This Article has three Parts. Part I summarizes the internet's potential for deepening civic engagement, as well as the substantial threats to that potential posed by digital hate. After describing the legal and political barriers to regulatory approaches to this problem, it explains that promising solutions nonetheless remain. More specifically, it documents the freedom and influence that intermediaries enjoy in shaping online expression generally and in addressing digital hate specifically.

Part II turns to implementation. It offers a range of recommendations for how intermediaries might exercise their power over cyber hate. We set forth an illustrative spectrum of possible hate speech definitions – grounded in terms of cyber hate's potential threats to digital citizenship as well as other specific harms – from which intermediaries might choose when developing their policies.

Part III then explores the variety of ways in which an intermediary might respond to speech that violates its policy. These include not only removal, but also engaging in or facilitating counter-speech, as well as educating and empowering users with respect to digital citizenship. We conclude that a thoughtful intermediary-based approach to hate speech can foster digital citizenship without suppressing valuable expression.

I. CIVIC ENGAGEMENT, CYBER HATE, AND INTERMEDIARIES' POTENTIAL FOR FOSTERING DIGITAL CITIZENSHIP

This Part starts by briefly recounting the internet's potential for deepening civic engagement and then summarizes the substantial threats to such engagement posed by digital hate. After identifying the legal and political

accountability through transparency and peer review.").

[32] *See id.* at 571 (describing the value of "normatively motivated inquiry and remediation by relevant non-judicial actors" in situations that involve unusual uncertainty or complexity).

[33] *See* Orly Lobel, *The Renew Deal: The Fall of Regulation and the Rise of Governance in Contemporary Legal Thought*, 89 MINN. L. REV. 342, 388, 389, 391 (2004) ("The governance model aims to create a flexible and fluid policy environment that fosters 'softer' processes that either replace or complement the traditional 'hard' ordering of the regulatory model[, such as] . . . social labeling, voluntary corporate codes of conduct, private accreditation, and certification by nongovernmental actors. . . . Flexibility implies variation in the communications of intention to control and discipline deviance. Less coercive sanctions can promote flexibility in implementation and compliance.").

barriers to governmental solutions to this problem, it explains that promising solutions remain in the form of voluntary measures by interested intermediaries.

A. *The Internet's Potential to Deepen Civic Engagement*

Among the many reasons to celebrate the internet's growth is its potential to enhance civic engagement, which in turn facilitates democratic functions. Democracy is often said to work best when citizens build networks of social interaction and trust.[34] Civic engagement allows people to see their lives as entwined with others.[35] In turn, people learn "habits of cooperation and public-spiritedness."[36] Civic engagement reinforces Alexis de Tocqueville's notion of "self-interest well understood" – that is, the capacity to consider the interests of others in addition to one's own[37] – and encourages "responsible citizenship."[38]

Although citizenship often describes a legal status enjoyed by members of a body politic,[39] citizenship can refer more broadly to participation in community life,[40] which may not be explicitly political but may ultimately further political participation.[41] Citizenship extends beyond the legal dimension to include "all the relationships . . . involved in membership in a community."[42] Citizenship "provides what the other roles cannot, namely an integrative experience which brings together the multiple role-activities of the contemporary person and demands that the separate roles be surveyed from a more general point of view."[43]

[34] ROBERT D. PUTNAM, BOWLING ALONE: THE COLLAPSE AND REVIVAL OF AMERICAN COMMUNITY 137-47 (2000).

[35] JOHN STUART MILL, *Considerations on Representative Government*, *reprinted in* THREE ESSAYS 143, 196-98 (Oxford Univ. Press 1975).

[36] PUTNAM, *supra* note 34, at 338.

[37] ALEXIS DE TOCQUEVILLE, DEMOCRACY IN AMERICA 501 (Harvey C. Mansfield & Delba Wintrop eds. and trans., 2000) (1835).

[38] RICHARD DAGGER, CIVIC VIRTUES: RIGHTS, CITIZENSHIP, AND REPUBLICAN LIBERALISM 104 (1997).

[39] *Id.* at 99.

[40] MILL, *supra* note 35, at 196.

[41] Indeed, public participation and civic engagement are often viewed as essential for members of a democracy to form a citizenry. JÜRGEN HABERMAS, BETWEEN FACTS AND NORMS: CONTRIBUTIONS TO A DISCOURSE THEORY OF LAW AND DEMOCRACY 307-08 (William Rehg trans., The MIT Press 1996) (1992); *see also* MILL, *supra* note 35, at 196-97 (explaining that a citizen is someone who develops his faculties through active engagement in public life). For John Dewey, citizen participation in communal life constituted the "idea of a democracy." JOHN DEWEY, THE PUBLIC AND ITS PROBLEMS 148-50 (1927).

[42] John Dewey, *The School as Social Centre*, 3 THE ELEMENTARY SCH. TCHR. 73, 76 (1902).

[43] SHELDON S. WOLIN, POLITICS AND VISION: CONTINUITY AND INNOVATION IN WESTERN POLITICAL THOUGHT 434 (1960).

Online activity can facilitate civic engagement and political participation. Neighborhood communities combine offline activities with online ones.[44] Companies encourage employees to use social network sites to deepen workplace relationships.[45] Workers can therefore discuss issues in person and in online postings.[46] Student organizations meet face-to-face in classrooms and in social network groups.[47]

Mediating institutions like schools, workplaces, churches, and community centers have traditionally given expression to the idea of citizenship.[48] This is especially so for institutions cultivating norms of trust across lines of social division, often referred to as "bridging ties."[49] Tocqueville emphasized the importance of townships and civic associations in enabling citizens to acquire the skills and habits of dialogue.[50] John Dewey found schools uniquely

[44] Amitai Etzioni, *On Virtual, Democratic Communities*, *in* COMMUNITY IN THE DIGITAL AGE: PHILOSOPHY AND PRACTICE 225, 228-29 (Andrew Feenberg & Darin Barney eds., 2004) (explaining that the town of Blacksburg, Virginia has an online community called Blacksburg Electronic Village where various groups and neighborhoods post meetings, share information, and interact). Examples abound of online political engagement, including the use of social media to raise campaign funds and organize voters during the 2008 Presidential election. Miki Caul Kittilson & Russell J. Dalton, *Virtual Civil Society: The New Frontier of Social Capital?*, POL. BEHAV. (Oct. 7, 2010) (forthcoming), *available at* http://www.springerlink.com/content/740r3560j640080t/; *see also* Nathaniel J. Gleicher, *MoneyBombs and Democratic Participation: Regulating Internet Fundraising*, 70 MD. L. REV. (forthcoming 2011) (manuscript at 5-6), *available at* http://ssrn.com/abstract=1695552.

[45] Jacob Christensen, *Managing Mondays: Facebook, a Viable Workplace Tool?*, LINKED 2 LEADERSHIP BLOG (Apr. 5, 2010), http://linked2leadership.com/2010/04/05/mm-facebook-a-workplace-tool/; *Two on Facebook . . . FNN Video and Employee Groups*, ONE DEGREE (Jan. 11, 2008), http://www.onedegree.ca/2008/01/two-on-facebook.html.

[46] For a discussion of the relationship between workplace relationships and civic engagement, see CYNTHIA ESTLUND, WORKING TOGETHER: HOW WORKPLACE BONDS STRENGTHEN A DIVERSE DEMOCRACY. Estlund suggests, however, that the rise of internet technology in the workplace may weaken rather than strengthen these bonds. *Id.* at 36-38.

[47] Popular social network sites like Facebook and MySpace were originally organized around existing institutions like schools, universities, and towns to enhance existing social connections. FELICIA WU SONG, VIRTUAL COMMUNITIES: BOWLING ALONE, ONLINE TOGETHER 22 (Digital Formations No. 54 2009).

[48] BENJAMIN R. BARBER, STRONG DEMOCRACY: PARTICIPATORY POLITICS FOR A NEW AGE 267 (1984).

[49] ESTLUND, *supra* note 46, at 107-08; *see also* BARBER, *supra* note 48, at 268. Not all associations contribute to liberal conceptions of democracy, however. MICHAEL J. SANDEL, DEMOCRACY'S DISCONTENT: AMERICA IN SEARCH OF A PUBLIC PHILOSOPHY 314-15 (1996). Some groups pursue distinctly illiberal aims, as this Article explores.

[50] TOCQUEVILLE, *supra* note 37, at 65, 496-97 (highlighting the importance of townships and civil associations because they allow citizens to "govern society" in the "restricted sphere that is within his reach"); *see also* SANDEL, *supra* note 49, at 333-35, 343 (extolling municipal parks, schools, libraries, community development corporations, and local retail

situated to teach children and adults about the social meaning of community[51] because they brought diverse people together in ways that "introduce deeper sympathy and wider understanding."[52] For Cynthia Estlund, the workplace serves as an important site for the formation of social and political views because it permits informal discourse among people "who are both *connected* with each other, so that they are inclined to listen, and *different* from each other, so that they are exposed to diverse ideas and experiences."[53]

Similarly, online intermediaries have potential to serve as mediating institutions that give expression to the idea of citizenship.[54] They can *extend* workplaces, schoolhouses, and community centers to digital spaces,[55] supplementing real-space exchanges of information and opinion with virtual ones. Online intermediaries also play an indispensable role in bringing together minority or marginalized groups in different geographic locations.[56] As Anupam Chander has noted, cyberspace can help "give members of minority groups a fuller sense of citizenship – a right to a practice of citizenship that better reflects who they are."[57]

establishments because they bring together rich and poor in public places and in public pursuits); Charlotte Garden, *Labor Values Are First Amendment Values: Why Union Comprehensive Campaigns Are Protected Speech*, 79 FORDHAM L. REV. 2617, 2656-58 (2011).

[51] *See* DEWEY, *supra* note 41, at 200.

[52] Dewey, *supra* note 42, at 83; *see* Harry C. Boyte, *A Different Kind of Politics: John Dewey and the Meaning of Citizenship in the 21st Century*. 12 GOOD SOC'Y, no. 2, 2003 at 1, 7. Dewey enlisted schools in the battle against bigotry: intolerance would lose force if exposed to "the ideas of others." Dewey, *supra* note 42, at 77.

[53] ESTLUND, *supra* note 46, at 123. She also emphasized the workplace's potential for enforcing civility and equality, which in turn allows diverse voices to be heard. *Id.* at 121-22.

[54] For an insightful discussion of schools as crucial speech-facilitating institutions, see Joseph Blocher, *Institutions in the Marketplace of Ideas*, 57 DUKE L.J. 821, 856-59 (2008).

[55] *See* SONG, *supra* note 47, at 4 (explaining that social media providers mediate practices of businesses, schools, and associations). A 2007 study found that Facebook cultivates bridging social capital. Nicole B. Ellison et al., *The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites*, 12 J. COMPUTER-MEDIATED COMM. 1143, 1161-62 (2007); *see also* Sebastian Valenzuela et al., Lessons from Facebook: The Effect of Social Network Sites on College Students' Social Capital 33 (Apr. 5, 2008) (unpublished manuscript), *available at* http://online.journalism.utexas.edu/2008/papers/Valenzuela.pdf (finding that Facebook users come from diverse backgrounds, contrary to the popular myth that they are typically female, upper-middle class students). Intermediaries of course also support bonding ties – those involving groups of similar backgrounds.

[56] SONG, *supra* note 47, at 74.

[57] Anupam Chander, *Whose Republic?*, 69 U. CHI. L. REV. 1479, 1481 (2002) (reviewing CASS SUNSTEIN, REPUBLIC.COM (2001)). For an innovative conception of transnational cultural citizenship, see Sonia K. Katyal, *The Dissident Citizen*, 57 UCLA L. REV. 1415, 1467-75 (2010).

In these and myriad other ways, users of online intermediaries can participate in community life.[58] When we speak of "digital citizenship" in this Article, we refer to the ways in which online activity has the potential to deepen civic engagement.[59]

Of course, that the internet carries the promise of fostering digital citizenship does not mean that such promise inevitably will be realized.[60] Online communications can instead foster isolation and disengagement.[61] Timothy Zick explains that networked technologies can interfere with expression in public spaces by distracting people from face-to-face interactions.[62] Robert Putnam questions whether the internet will generate norms of trust given its facilitation of anonymous interactions that lack wider social context.[63] The next Section focuses more specifically on the perils to such engagement posed by cyber hate.

[58] Timothy Fort defines "mediating institutions" to mean "communities which socialize their members," and "require individuals to grasp their responsibilities to others, at least within their group, so that a person's very identity is developed." Timothy L. Fort, *The Corporation as Mediating Institution: An Efficacious Synthesis of Stakeholder Theory and Corporate Constituency Statutes*, 73 NOTRE DAME L. REV. 173, 174-75 (1997); *see also* Andrew Crane et al., *Stakeholders as Citizens? Rethinking Rights, Participation, and Democracy*, 53 J. BUS. ETHICS 107, 108 (2004) (describing various conceptions of corporate citizenship as including "corporations as citizens; corporations as administrators of citizenship; and stakeholders as citizens").

[59] The term "digital citizenship" can mean different things depending upon the community and audience. For instance, some political scientists have used the term to refer broadly to the "ability to participate in society online," arguing that disadvantaged groups cannot fully participate as citizens due to their limited access to the internet. KAREN MOSSBERGER, CAROLINE J. TOLBERT & RAMONA S. MCNEAL, DIGITAL CITIZENSHIP: THE INTERNET, SOCIETY, AND PARTICIPATION 1 (2008). Intermediaries, too, have invoked the concept of digital citizenship as an aspiration for civil online behavior. Indeed, we first encountered this term in conversations with those in the industry who had already identified the facilitation of "digital citizenship" as a goal. Interview with Hemanshu Nigam, former Chief Safety Officer, MySpace (June 22, 2010).

[60] Evgeny Morozov's work explores the related, though distinct, question of how democratic freedoms can be threatened by governmental abuse of networked technologies. *See generally* EVGENY MOROZOV, THE NET DELUSION: THE DARK SIDE OF INTERNET FREEDOM (2011). Anupam Chander has also explored this question with great insight in *Googling Freedom*, 99 CALIF. L. REV. 1 (2011).

[61] SONG, *supra* note 47, at 23.

[62] TIMOTHY ZICK, SPEECH OUT OF DOORS: PRESERVING FIRST AMENDMENT LIBERTIES IN PUBLIC PLACES 304 (2009). Professor Zick explains that new technologies contribute to the phenomenon of "absent presence" where people occupy personal "technology bubbles" and disconnect from others who physically surround them. *Id.*

[63] PUTNAM, *supra* note 34, at 175-76; *see also* Chander, *supra* note 57 at 1480 ("Which of these possible uses of the Internet – the Internet as a tool for discovery and education, or the Internet as an echo chamber – will find more adherents is an empirical question that we may not yet be able to answer.").

B. *Cyber Hate's Potential to Imperil Digital Citizenship*

Online activities can pose dangers that work to undermine civic engagement. The internet facilitates anonymous and pseudonymous discourse, which can just as easily accelerate destructive behavior as it can fuel public discourse.[64] It provides a cheap and easy way to reach like-minded individuals located at disparate geographic locations, removing barriers that often limit group activity. [65] Search engines ensure access to, and the persistence of, online content of all types – including hateful content. Hate groups exploit these and other online attributes to spread, legitimize, and entrench hateful messages that imperil participation in community life.[66]

Cyber threats and calls for violence can undermine political and civic engagement. History[67] and social science[68] confirm that hate speech may

[64] Social science research suggests that people may behave more aggressively when they believe that they cannot be observed and caught. Citron, *Cyber Civil Rights*, *supra* note 20, at 82.

[65] *See* Kyu Ho Youm, *First Amendment Law: Hate Speech, Equality, and Freedom of Expression*, 51 J. COMM. 406, 406 (2001) (book review) (describing reports by Don Black – the "godfather of the Internet racist movement" – that the internet dramatically increased his ability to disseminate his views compared to his previous reliance on traditional print media).

[66] *See* ADAM G. KLEIN, A SPACE FOR HATE: THE WHITE POWER MOVEMENT'S ADAPTATION INTO CYBERSPACE 55 (2009) (describing "information laundering" to mean "the legitimizing factor of an interconnected information superhighway of web directories, research engines, news outlets, and social networks that collectively funnel into and out of today's hate websites"). Klein continues:

> For information-seekers, the result of this funneling process is a wider array of perspectives, and thus, a broader understanding of any given topic. However, for propaganda-providers like the white power movement, the same process inadvertently lends the credibility and reputation of authentic websites to those illegitimate few to which they are nonetheless connected. Such is the case with many of today's leading search engines like Google, that unwittingly filter directly into hate websites, or public networks like YouTube, which host their venomous content everyday.

Id.

[67] *See* Mari J. Matsuda, *Public Response to Racist Speech: Considering the Victim's Story*, 87 MICH. L. REV. 2320, 2352 n.166 (1989) (describing history of escalating racist violence that accompanies racist speech); Alexander Tsesis, *Dignity and Speech: The Regulation of Hate Speech in a Democracy*, 44 WAKE FOREST L. REV. 499, 509-15 (2009) [hereinafter Tsesis, *Dignity and Speech*] (describing history of anti-Semitic and racist speech that incited or escalated violent acts); Alexander Tsesis, *The Empirical Shortcomings of First Amendment Jurisprudence: A Historical Perspective on the Power of Hate Speech*, 40 SANTA CLARA L. REV. 729, 740-55 (2000) (detailing the relationship between hate speech and acts of violence against Jews, Native Americans and African-Americans).

[68] *See, e.g.*, David Kretzmer, *Freedom of Speech and Racism*, 8 CARDOZO L. REV. 445, 463 (1987) (describing social science demonstrating the importance of speech as a precondition to acts of racial violence or scapegoating).

facilitate acts of violence against members of targeted groups.[69] For instance, digital hatred helped inspire the 1999 shooting of African-Americans, Asian-Americans, and Jews in suburban Chicago by Benjamin Smith, a member of the white supremacist group World Church of the Creator (WCOTC) that promotes racial holy war.[70] Just months before the shootings, Smith told documentary filmmaker Beverly Peterson that: "It wasn't really 'til I got on the internet, read some literature of these groups that . . . it really all came together."[71]

More recently, the Facebook group *Kick a Ginger Day* urged members to get their "steel toes ready" to attack individuals with red hair.[72] The site achieved its stated goal: students punched and kicked children with red hair, with dozens of Facebook members claiming credit online for the attacks.[73]

Aside from producing physical harm, online calls for violence and threats can silence members of targeted groups.[74] Consider a California teenager's

[69] Hate speech that takes the form of "fighting words" may sometimes provoke violent responses from its targets in addition to inciting violence against them. *See* Charles R. Lawrence III, *If He Hollers, Let Him Go: Regulating Racist Speech on Campus*, 1990 DUKE L.J. 431, 452; Ronald Turner, *Regulating Hate Speech and the First Amendment: The Attractions of, and Objections to, an Explicit Harms-Based Analysis*, 29 IND. L. REV. 257, 298-300 (1995) (describing violence provoked by use of the n-word or other face-to-face uses of particular racial or religious epithets).

[70] Christopher Wolf, *Racists, Bigots and the Law on the Internet: Internet Hate Speech and the Law*, ANTI-DEFAMATION LEAGUE, http://www.adl.org/Internet/Internet_law3.asp (last visited Apr. 5, 2011). Smith killed former Northwestern University basketball coach Ricky Byrdsong and Indiana University student Won Joon Yoon and wounded six Orthodox Jews and three African-Americans. Elizabeth Brackett, *The Hate Crimes Question*, PBS ONLINE NEWS HOUR (Aug. 11, 1999), http://www.pbs.org/newshour/bb/law/july-dec99/hate_8-11.html. The internet helped make the WCOTC one of the fastest growing hate groups in the United States. *Id.*

[71] *The Consequences of Right-Wing Extremism on the Internet: Inspiring Extremist Crimes*, ANTI-DEFAMATION LEAGUE, http://www.adl.org/internet/extremism_rw/inspiring.asp (last visited Apr. 5, 2011). WCOTC's website operator at the time of the rampage confirmed that Smith sent him "about five" email messages "congratulating" him on the group's websites and indicating that he regularly read them. *Id.*

[72] Nordlinger, *supra* note 11, at 8; Moore, *supra* note 11.

[73] Nordlinger, *supra* note 11, at 8.

[74] *See* Richard Delgado & David Yun, *The Neoconservative Case Against Hate-Speech Regulation – Lively, D'Souza, Gates, Carter, and the Toughlove Crowd*, 47 VAND. L. REV. 1807, 1822-23 (1994) (explaining how our culture has developed a host of narratives about overcoming hurt feelings while ignoring hurtful words that undermine victims' ability to respond and to mobilize effectively against hate); Lawrence, *supra* note 69, at 452 ("When racial insults are hurled at minorities, the response might be silence or flight rather than a fight, but the preemptive effect on further speech is just as complete as with fighting words."); Netanel, *supra* note 27, at 426 ("Individuals may develop deep feelings of attachment and loyalty to virtual communities and may be devastated by perceived wrongs within those communities. In such instances, exit is far from costless."); Steven H. Shiffrin,

experience with internet hate speech. Commenters (later discovered to be students at the teenager's high school) on the student's website repeatedly threatened him in homophobic ways.[75] One wrote: "F-----, I'm going to kill you."[76] Another wrote: "If I ever see you I'm . . . going to pound your head in with an ice pick."[77] Others wrote "You are an oversized f----- . . . . I just want to hit you in the neck" and "I hate f--s . . . . You need to be stopped."[78] The student's father shut down the site and, on the advice of the police, kept his son from attending school during the investigation.[79]

In a similar vein, Kathy Sierra, a well-known programmer, maintained a popular blog on software development called "Creating Passionate Users."[80] In 2007, anonymous posters verbally attacked Ms. Sierra on her blog and two other websites.[81] On her blog, commenters suggested that she deserved to have her throat slit, be suffocated, sexually violated, and hanged.[82] On another blog, posters uploaded doctored photographs of Ms. Sierra: one picture featured her with a noose beside her neck; another depicted her screaming while being suffocated by lingerie.[83] After the attacks, Ms. Sierra canceled speaking engagements and feared leaving her home.[84] As she explained, "my blog was in the Technorati Top 100 [at the time of the attack]. I have not blogged there – or anywhere – since."[85]

Racist Speech, Outsider Jurisprudence, and the Meaning of America, 80 CORNELL L. REV. 43, 86 (1994) ("[R]acial vilification can create a repressive environment in which the speech of people of color is chilled or not heard."); Mike Adams, *Facebook Devolves into Dark Web of Anonymous Hate Speech*, NATURALNEWS (Aug. 26, 2010), https://www.naturalnews.com/029572_Facebook_hate_speech.html (stating that hate speech on Facebook has caused individuals who would otherwise be participating in the public discourse to close their accounts).

[75] Kim Zetter, *Court: Cyberbullying Threats Are Not Protected Speech*, WIRED BLOG: THREAT LEVEL (Mar. 18, 2010, 3:23 PM), http://www.wired.com/threatlevel/2010/03/cyberbullying-not-protected/.

[76] *Id.*

[77] D.C. v. R.R., 106 Cal. Rptr. 3d 399, 405 (Ct. App. 2010).

[78] *Id.* at 407

[79] *Id.* at 446 (Rothschild, J., dissenting).

[80] Dahlia Lithwick, *Fear of Blogging: Why Women Shouldn't Apologize for Being Afraid of Threats on the Web*, SLATE (May 4, 2007, 7:20 PM), http://www.slate.com/id/2165654/.

[81] *Id.*

[82] *Id.*; Greg Sandoval, *Blogger Cancels Conference Appearance After Death Threats*, CNET NEWS BLOG (Mar. 26, 2007), http://news.cnet.com/8301-10784_3-6170683-7.html.

[83] Jessica Valenti, *Women: How the Web Became a Sexists' Paradise*, GUARDIAN (London), Apr. 6, 2007, at 16; Sandoval, *supra* note 82.

[84] *Blog Death Threats Spark Debate*, BBC NEWS (Mar. 27, 2007), http://news.bbc.co.uk/go/pr/fr/-/2/hi/technology/6499095.stm.

[85] Kathy Sierra, Comment to *CCR Symposium: A Behavioral Argument for Stronger Protections*, CONCURRING OPINIONS (Apr. 18, 2009, 2:25 PM), http://www.concurringopinions.com/archives/2009/04/ccr_symposium_a_1.html#comments.

Consider too the posters on a white supremacist website who targeted Bonnie Jouhari, a civil rights advocate and mother of a biracial girl.[86] The site showed a picture of Ms. Jouhari's workplace exploding in flames next to the threat that "race traitors" are "hung from the neck from the nearest tree or lamp post."[87] Posters included bomb-making instructions and a picture of a hooded Klansman holding a noose.[88] Ms. Jouhari and her daughter have withdrawn from public life.[89] They do not have driver's licenses, voter registration cards, or bank accounts for fear of creating a public record of their whereabouts.[90]

Cyber hate can undermine targeted group members' capability for civic engagement in other ways apart from threatening or inciting violence. It can convey the message that a group in the community is "not worthy of equal citizenship."[91] R.A. Lenhardt explains that hate speech undermines group members' ability to belong and participate in processes crucial to community life.[92] Online hate can thus denigrate group members' basic standing in society and deprive them of their "civic dignity."[93]

In this way, cyber hate can inflict serious psychological injury, including fear, stress, feelings of inferiority, and depression.[94] Recall the attacks upon

[86] Ryan Wilson, HUDALJ 03-98-0692-8 ¶¶ 2-3, 6 (July 19, 2000). For an excellent description and analysis of the case, see Catherine E. Smith, *Intentional Infliction of Emotional Distress: An Old Arrow Targets the New Hate Hydra*, 80 DENV. U. L. REV. 1, 35-48 (2002).

[87] Wilson, HUDALJ 03-98-0692-8, at ¶¶ 9-11.

[88] *Id.* at ¶¶ 9, 15

[89] DeWayne Wickham, *They Suffer for Doing Right Thing*, USA TODAY (May 16, 2000, 8:39 AM), http://www.usatoday.com/news/opinion/columnists/wickham/wick093.htm (explaining that Ms. Jouhari and her daughter have moved four times to ensure that posters do not find them).

[90] *Id.*

[91] Jeremy Waldron, *Dignity and Defamation: The Visibility of Hate*, 123 HARV. L. REV. 1596, 1601 (2010).

[92] R.A. Lenhardt, *Understanding the Mark: Race, Stigma, and Equality in Context*, 79 N.Y.U. L. REV. 803, 844-48 (2004); *see also* KENNETH L. KARST, BELONGING TO AMERICA: EQUAL CITIZENSHIP AND THE CONSTITUTION 3 (1989) (offering a principle of equal citizenship that suggests people are "entitled to be treated" as "respected, responsible, and participating member[s]"). Jennifer Gordon and R.A. Lenhardt's theory of belonging focuses on formal and informal pathways to the genuine possession and exercise of citizenship in the United States, including political participation, work, and education. Gordon & Lenhardt, *supra* note 26, at 1186-88.

[93] Waldron, *supra* note 91, at 1607.

[94] *See* Matsuda, *supra* note 67, at 2332 (describing the harm of "[t]he spoken message of hatred"); Shiffrin, *supra* note 74, at 86 (describing how racist speech inflicts harm on its individual victims by inspiring self hatred, isolation, and emotional distress); *see also* Kretzmer, *supra* note 68, at 466 (describing how hate speech may trigger insecurity, self hatred, humiliation, isolation, and other psychological harm); Lawrence, *supra* note 69, at 462 (describing how racial epithets and harassment cause deep emotional scarring in the form of anxiety and fear); *cf.* Citron, *Law's Expressive Value*, *supra* note 20, at 388-90

Ms. Jouhari: Ms. Jouhari suffered headaches and anxiety, and her daughter was diagnosed as suffering from severe post-traumatic stress disorder.[95] Indeed, young people can feel such psychological harms intensely, as electronic media exert a powerful influence on children and teenagers who have not yet reached full cognitive development.[96] Not only are children particularly vulnerable to hate's emotional harms, they are also less able to fight back.[97]

Hate speech may further degrade public discourse by skewing society's assessment of members of certain racial, religious, or other groups and of their ideas.[98] Charles Lawrence, for example, argues that racism "trumps good ideas that contend with it in the market, often without our even knowing it."[99] By devaluing targeted group members' expression, hate speech can produce a process defect in the marketplace of ideas.[100]

Moreover, because hate speech may inspire or deepen prejudice, it can lead to discriminatory decisions about jobs, housing, and other life opportunities.[101] Stigma, often exacerbated or inspired by hate speech, can render targeted group members dishonored and erect significant barriers to full acceptance into the wider community.[102] Not only does such bigotry impose tangible costs on targeted group members who suffer the effects of discriminatory decisions, it more broadly undermines society's commitment to equality and dignity.[103]

(exploring how cyber gender harassment produces anxiety and other forms of emotional distress).

[95] Ryan Wilson, HUDALJ 03-98-0692-8, at 24-25 (July 19, 2000).

[96] Michele L. Ybarra et al., *Linkages Between Internet and Other Media Violence with Seriously Violent Behavior by Youth*, 122 PEDIATRICS 929, 933 (2008).

[97] *See* Richard Delgado, *Words that Wound: A Tort Action for Racial Insults, Epithets, and Name-Calling*, 17 HARV. C.R.-C.L. L. REV. 133, 147 (1982).

[98] As Charles Lawrence powerfully observes, the notions of "racial inferiority of non-whites infects, skews, and disables the operation of the market (like a computer virus, sick cattle, or diseased wheat)." Lawrence, *supra* note 69, at 468.

[99] *Id.*

[100] *Id.* Websites and other online actors may exacerbate this process defect by enabling hate groups to link exclusively to hateful content, creating "echo chambers" of extreme positions, which can harden and encourage the development of even more extreme views. CASS R. SUNSTEIN, REPUBLIC.COM 2.0 145 (2007).

[101] *See* Delgado & Yun, *supra* note 74, at 1813 (maintaining that hate speech feeds discriminatory decision-making by reinforcing stereotypes); Kretzmer, *supra* note 68, at 505. In this way, prejudice and bigotry fostered by hate speech can produce conscious and unconscious behavioral consequences and thus intensify their targets' disadvantage. Jerry Kang, *Trojan Horses of Race*, 118 HARV. L. REV. 1489, 1539-40 (2005) (examining the role played by racial stereotypes in mass media in creating and maintaining biases that result in discriminatory decision-making).

[102] Lenhardt, *supra* note 92, at 844-48; *see also* ERVING GOFFMAN, STIGMA: NOTES ON THE MANAGEMENT OF SPOILED IDENTITY 2-5 (1963) (discussing stigma as creating a spoiled social identity).

[103] *See* Delgado, *supra* note 97, at 142 (explaining that racist speech undermines "society

In turn, search engines ensure the persistence of cyber hate and its costs to digital citizenship. Because search engines reproduce information cached online, targets of hate speech cannot depend upon time's passage to alleviate the damage that online postings cause.[104] For this reason, Jeremy Waldron contends that cyber hate produces a "permanent disfigurement" of group members.[105] In all these ways, cyber hate threatens to undermine digital citizenship.[106]

In our opinion, the threats posed by online hate to digital citizenship are sufficiently substantial to demand a response. Regulatory solutions, however, face considerable constitutional and political barriers. Governmental efforts to regulate hate speech based on its content, for example, trigger significant First Amendment concerns.[107]

as a whole").

[104] Danielle Keats Citron, *Mainstreaming Privacy Torts*, 98 CALIF. L. REV. 1805, 1813 (2010); Jeffrey Rosen, *The End of Forgetting*, N.Y. TIMES, July 25, 2010, at MM30.

[105] Waldron, *supra* note 91, at 1601, 1610.

[106] Although our discussion here focuses on the harm to civic engagement posed by digital hate, we recognize that such hate speech inflicts other moral and instrumental harms as well.

[107] *See, e.g.*, R.A.V. v. City of St. Paul, 505 U.S. 377, 391 (1992) (holding that city ordinance that prohibited expression that "arouses anger, alarm or resentment in others . . . on the basis of race, color, creed, religion, or gender" impermissibly discriminated on the basis of viewpoint in violation of the First Amendment). Indeed, as some thoughtful commentators have observed, regulatory efforts to constrain hate speech not only face constitutional challenges, but may impose instrumental costs of their own. For example, visible hate speech can remind readers and listeners of bigotry's prevalence and the need to enforce existing antidiscrimination laws. Shiffrin, *supra* note 74at 89. It may perform a powerful teaching function in exposing the poverty of a hate group's beliefs. Kingsley R. Browne, *Title VII as Censorship: Hostile-Environment Harassment and the First Amendment*, 52 OHIO ST. L.J. 481, 542 (1991) (arguing that permitting hate speech can contribute to the elimination of prejudice because such speech will expose the poverty of those beliefs). Others suggest that hateful expression may play a role in preventing violence by allowing speakers to let off steam. Vincent Blasi, *The Teaching Function of the First Amendment*, 87 COLUM. L. REV. 387, 408 (1987) (reviewing LEE C. BOLLINGER, THE TOLERANT SOCIETY (1986)). *But see* Dhammika Dharmapala & Richard H. McAdams, *Words That Kill? An Economic Model of the Influence of Speech on Behavior (with Particular Reference to Hate Speech)*, 34 J. LEG. STUD. 93, 132 (2005) (discussing how raising the costs of engaging in hate speech may deter hate crime rather than increase the rate of hate crime). Refraining from regulating hate speech may avoid making martyrs of – and thus invigorating and multiplying – hateful speakers. Graham Hughes, *Prohibiting Incitement to Racial Discrimination*, 16 U. TORONTO L.J. 361, 365 (1996) (suggesting that regulation creates martyrs and converts to the cause of hatred); Larissa Barnett Lidsky, *Where's the Harm?: Free Speech and the Regulation of Lies*, 65 WASH. & LEE L. REV. 1091, 1099-1100 (2008) (concluding that punishing Holocaust denial will paradoxically entrench that view and inspire stronger belief in conspiracy theories). So, too, the expression of hate speech might foster a certain capacity of mind, enabling us to confront our biases, master our irrational passions, and develop further tolerance ourselves. LEE C.

Apart from the First Amendment difficulties confronted by governmental efforts to regulate digital hate, such efforts face considerable political challenges as well, as demonstrated by the experience of those who proposed legislation to address discriminatory conduct far beyond the realm of pure expression. For example, the Hate Crimes Prevention Act - which criminalizes bias-motivated crimes of violence - was enacted only after years of effort.[108] Along the same lines, legislation to prohibit job discrimination on the basis of sexual orientation has been introduced in Congress in various forms since 1975 but has yet to be enacted.[109]

For these reasons, calls for governmental responses to digital hate face substantial challenges. As the next Section explains, however, promising alternatives remain available.

C. *Intermediaries' Freedom to Challenge Digital Hate*

Internet intermediaries enjoy enormous freedom to decide whether and how to shape online expression. The First Amendment, of course, protects speech only from governmental restriction and thus does not govern private actors' decisions to remove or filter online expression.[110] At the same time, federal law immunizes "provider[s] or user[s] of interactive computer services" from liability arising from content created by others and from requirements to remove "offensive" speech.[111] Intermediaries thus enjoy wide latitude to make

BOLLINGER, THE TOLERANT SOCIETY: FREEDOM OF SPEECH AND EXTREMIST SPEECH IN AMERICA 142, 173 (1986).

[108] Matthew Shepard and James Byrd, Jr. Hate Crimes Prevention Act, Pub. L. No. 111-84, 123 Stat 2190 (2009) (to be codified at 18 U.S.C. § 249).

[109] H.R. REP. NO. 110-406 pt. 1, at 2 (2007).

[110] *See, e.g.*, Green v. Am. Online (AOL), 318 F.3d 465, 472 (3d Cir. 2003) (finding that private company AOL is not subject to constitutional free speech guarantees and has not been transformed into a state actor simply because it "provides a connection to the Internet on which government and taxpayer-funded websites are found"); Langdon v. Google, Inc., 474 F. Supp. 2d 622, 631 (D. Del. 2007) (ruling that Google, Yahoo!, and Microsoft are private companies not subject to constitutional free speech guarantees even though they may work with state actors like public universities).

[111] 47 U.S.C. § 230(c) (2006); *see also* Zeran v. Am. Online, Inc., 129 F.3d 327, 330 (4th Cir. 1997) (barring claims against an online service provider under § 230 because defendant did not create the allegedly tortious content). Intermediaries can incur liability for content that they create themselves (e.g., for their own postings that are defamatory or threatening), or for publishing content that violates copyright law. 47 U.S.C. § 230(e)(1); *see also* Wendy Seltzer, *Free Speech Unmoored in Copyright's Safe Harbor: Chilling Effects of the DMCA on the First Amendment*, 24 HARV. J.L. & TECH. 171, 228 (2010) (noting that § 230 "specifically excludes intellectual property and criminal claims from its protections"). For instance, Internet Service Providers (ISPs) and website operators can incur liability under the Digital Millennium Copyright Act for refusing to take down content that they have been notified violates copyright law, whereas they enjoy immunity from liability for defamatory postings created by others. *Id.* at 175.

all sorts of decisions – including none at all – with respect to others' hate speech.[112]

A number of intermediaries have begun to consider such questions, variously motivated by concerns about the potential business, moral, and instrumental costs of digital hate. Some intermediaries see digital hate as a potential threat to profits.[113] MySpace,[114] for instance, sees its aggressive approach to hate speech – and, indeed, to a wide range of potentially offensive speech in addition to hate speech – as essential to securing online advertising for its customer base.[115] According to MySpace's former Chief Safety Officer Hemanshu Nigam, its approach stems from its sense of "what the company stood for and what would attract advertising and revenue."[116] Nigam explains that because kids and adults use MySpace, the company wanted to ensure a "family friendly" site, which could only be accomplished by taking down content that "attacked an individual or group because they are in that group and . . . made people feel bad."[117] As Nigam suggests, voluntary efforts to address hate speech may serve an intermediary's bottom line by creating market niches and contributing to consumer goodwill.[118]

[112] *See* Seltzer, *supra* note 111, at 228 (explaining that internet service providers can thus "set their own terms of service – choosing to maintain 'family-friendly' environments, attempting to build communities, or taking a hands-off, anything goes approach").

[113] Such intermediaries may explain their actions under the traditional "shareholder primacy" view that understands the corporation's primary (and perhaps exclusive) objective as maximizing shareholder wealth. *See, e.g.*, Mark J. Roe, *The Shareholder Wealth Maximization Norm and Industrial Organization*, 149 U. PA. L. REV. 2063, 2065 (2001); *For Whom Corporate Managers* Are *Trustees: A Note*, 45 HARV. L. REV. 1365, 1367-69 (1932). Along these lines, intermediaries' sense of the bottom-line benefits of addressing hate speech can be shaped by consumers' – i.e., users' – expectations.

[114] Although Facebook has now overtaken MySpace in popularity, MySpace remains popular. *Search Results for Myspace*, QUANTCAST CORP., http://www.quantcast.com/search?q=myspace (last visited Apr. 16, 2011) (noting that MySpace is the 35th most popular site in the United States).

[115] MySpace prohibits content that "promotes or otherwise incites racism, bigotry, hatred or physical harm of any kind against any group or individual . . . [or] exploits people in a sexual or violent manner." *MySpace.com Terms of Use Agreement*, MYSPACE.COM, http://www.myspace.com/help/terms (last visited Apr. 7, 2011).

[116] Interview with Nigam, *supra* note 59.

[117] *Id.*

[118] *Id.*; *see also* Paul Alan Levy, *Stanley Fish Leads the Charge Against Immunity for Internet Hosts – But Ignores the Costs*, CONSUMER L. & POL'Y BLOG (Jan. 8, 2011), http://pubcit.typepad.com/clpblog/2011/01/stanley-fish-leads-the-charge-against-immunity-for-internet-hosts-but-ignores-the-costs.html (arguing that websites that fail to provide protections against abuse will find "that the ordinary consumers whom they hope to serve will find it too uncomfortable to spend time on their sites, and their sites will lose social utility (and, perhaps more cynically, they know they will lose page views that help their ad revenue)").

Some intermediaries are motivated to address digital hate based on their sense of their own corporate social responsibility.[119] Indeed, many intermediaries explicitly invoke broad social responsibility principles when describing their services and their mission.[120] For example, Google explains that in offering the platform Blogger to users, it "want[s] to be socially responsible."[121] For this reason, it admonishes users that they can utilize "Blogger to express [their] opinions, even very controversial ones," but that they cannot "cross the line by publishing hate speech."[122]

[119] Such decisions may be justified as a matter of corporate law under the social entity theory of the corporation, which permits corporate decision-makers to consider and serve the interests of all the various constituencies affected by the corporation's operation. *See* Lisa M. Fairfax, *Doing Well While Doing Good: Reassessing the Scope of Directors' Fiduciary Obligations in For-Profit Corporations with Non-Shareholder Beneficiaries*, 59 WASH. & LEE L. REV. 409, 412 (2002).

[120] Yahoo! lists its company values as including "an infectious sense of mission to make an impact on society and empower consumers in ways never before possible. We are committed to serving both the Internet community and our own communities." *Yahoo! – What We Value,* YAHOO! INC., http://docs.yahoo.com/info/values/ (last visited Apr. 7, 2011). Yahoo! also states that it "empower[s] people through corporate social responsibility programs, products, and services to make a positive impact on their communities." *Overview*, YAHOO! INC., http://pressroom.yahoo.net/pr/ycorp/overview.aspx (last visited June 1, 2011). Google makes clear that it will pursue policies that may conflict with short-term shareholder economic gain but that it believes will benefit shareholders in the long term, and which may include benefits other than economic ones. Google Inc., Initial Public Offering Letter: *'An Owner's Manual' for Google's Shareholders* (Form S-1/A) (Aug. 18, 2004), *available at* http://investor.google.com/corporate/2004/ipo-founders-letter.html. Microsoft states:

> As a global company, we are accountable to millions of customers and stakeholders around the world. As we work to meet their needs, we are committed to creating value for our partners, employees, and wider society, and to managing our business sustainably. This commitment gives focus to our corporate citizenship work and helps us measure our performance over time.

Our Commitments, MICROSOFT, http://www.microsoft.com/about/corporatecitizenship/en-xf/our-commitments/ (last visited Apr. 7, 2011). Following the "Goals" link from that page brings a description of a link titled "Corporate Governance," which states that "considering the interests of other stakeholders – employees, customers, partners, suppliers, and the many communities around the world where we do business – is important to achieving the long-term interests of Microsoft shareholders." *Goals*, MICROSOFT, http://www.microsoft.com/about/corporatecitizenship/en-xf/our-commitments/goals/ (last visited Apr. 7, 2011).

[121] Rachel Whetstone, *Free Expression and Controversial Content on the Web*, THE OFFICIAL GOOGLE BLOG (Nov. 14, 2007, 3:58 PM), http://googleblog.blogspot.com/2007/11/free-expression-and-controversial.html.

[122] Google Blogger requires users to refrain from promoting "hate or violence towards groups based on race, ethnicity, religion, disability, gender, age, veteran status, or sexual orientation/gender identity." *Blogger Content Policy*, GOOGLE, http://www.blogger.com/content.g (last visited Apr. 7, 2011). Google further admonishes Blogger users: "don't write a blog saying that members of Race X are criminals or advocating violence against

Other intermediaries have invoked similar values in response to certain types of online hatred. After Facebook took down the *Kill a Jew Day* page in May 2010,[123] its spokesperson Andrew Noyes explained:

> Unfortunately ignorant people exist and we absolutely feel a social responsibility to silence them on Facebook if their statements turn to direct hate. That's why we have policies that prohibit hateful content and we have built a robust reporting infrastructure and an expansive team to review reports and remove content quickly.[124]

***

As this Part documents, intermediaries have the ability to decide whether and how to shape online expression. Many have elected to use that freedom to challenge online hate speech. Of course, many others have not. Indeed, some intermediaries base their business on tolerating or encouraging cyber hate. This is true, for instance, of the social network site Hate Book, which urges its users to "Post something you hate!"[125]

This Article addresses those intermediaries interested in combating their users' cyber hate. We urge them to consider the ways in which their services can be used to enrich as well as to endanger civic engagement.[126] In so doing, we recognize that a focus on the effects on civic engagement is not the only – nor necessarily the best – way of understanding the harms of hate speech. Nonetheless, we identify a commitment to digital citizenship as among the justifications for developing thoughtful approaches to hate speech, and one that could motivate interested intermediaries as well. In the remainder of this Article, we examine more specifically how intermediaries might address those challenges in developing and implementing hate speech policies.

This Article discusses intermediaries' choices in light of the freedom that they enjoy under current law. Indeed, Congress has encouraged intermediary involvement, providing immunity for intermediaries who take down "offensive material."[127] We note, however, that some thoughtful commentators challenge that status quo, arguing that select intermediaries should be treated as monopolies and thus subject to greater regulation.[128] Some of this discussion

followers of Religion Y." *Id.*

[123] *See supra* notes 1-3 and accompanying text.

[124] Lappin, *supra* note 1. Facebook is "sensitive to content that includes pornography, bullying, hate speech, and actionable threats of violence." *Id.*

[125] *Hate Book*, HATE BOOK, http://www.hatebook.com/tos.php (last visited Apr. 7, 2011).

[126] As Neil Netanel wrote of intermediaries who exclude individuals from their networks due to their race or gender, these intermediaries' actions "work a fundamental impairment not only of 'netizenship,' but also of citizenship in territorial polities." Netanel, *supra* note 27, at 457.

[127] 47 U.S.C. § 230(c)(2) (2006).

[128] *See, e.g.*, Oren Bracha & Frank Pasquale, *Federal Search Commission? Access, Fairness, and Accountability in the Law of Search*, 93 CORNELL L. REV. 1149, 1180-82 (2008) (deeming search engine Google a natural monopoly deserving of public regulation).

relates to net neutrality debates over the regulation of broadband network operators that we do not address in this Article.[129] To the extent that some urge greater regulation of social media and search engine intermediaries discussed here, their concerns do not stem from such intermediaries' attention to hate speech issues.[130]

II. IMPLEMENTING A CONCEPTION OF DIGITAL CITIZENSHIP: A TRANSPARENT COMMITMENT TO FIGHTING HATE

As explained above, significant moral and policy justifications support intermediaries who choose to engage in voluntary efforts to combat hate speech. Indeed, many intermediaries already choose to address online hatred in some way.[131] In this Part, we urge intermediaries – and others – to think and speak more carefully about the harms they hope to forestall when developing hate speech policies.

A. *The Transparency Principle*

We believe that intermediaries can valuably advance the fight against digital hate with more transparency and specificity about the harms that their hate speech policies address, as well as the consequences of policy violations. With more transparency regarding their specific reasons for choosing to address digital hate, intermediaries can make behavioral expectations more understandable.[132] Without it, intermediaries will be less effective in expressing what it means to be responsible users of their services.

In a series of articles, Frank Pasquale has argued that an Internet Intermediary Regulatory Council should oversee search engines and carriers, assisting the FCC and FTC in carrying out their present missions. Frank Pasquale, *Trusting (and Verifying) Online Intermediaries' Policing*, *in* THE NEXT DIGITAL DECADE: ESSAYS ON THE FUTURE OF THE INTERNET 347, 348 (Berin Szoka & Adam Marcus eds., 2010), *available at* http://nextdigitaldecade.com/read-book/now. Pasquale argues that such regulation of Google is warranted given "Google's dominance of the general search market," the company's indispensable role in economic, cultural, and political life, and the opacity of its practices that immobilize consumer voice options. *Id.*

[129] *See, e.g.*, BARBARA VAN SCHEWICK, INTERNET ARCHITECTURE AND INNOVATION 222-51 (2010).

[130] *See, e.g.*, DAWN C. NUNZIATO, VIRTUAL FREEDOM: NET NEUTRALITY AND FREE SPEECH IN THE INTERNET AGE, at xiv-xv (2009) (arguing that Congress should pass a law, or require the FCC, to prohibit broadband providers from blocking legal content or applications and from engaging in various forms of discrimination and prioritization of packets and that perhaps law should regulate powerful search engines such as Google as well).

[131] *See supra* notes 113-124 and accompanying text.

[132] Past calls for transparency from these entities have focused on legitimacy concerns regarding stealth marketing and undisclosed political and cultural biases. These have included Frank Pasquale, *Beyond Innovation and Competition: The Need for Qualified Transparency in Internet Intermediaries*, 104 NW. U. L. REV. 105, 155 (2010) (discussing

Indeed, those intermediaries that address hate speech in their Terms of Service (TOS) agreements or Community Guidelines rarely define key terms like "hateful" or "racist" speech with specificity.[133] The terms of service of Yahoo!, for instance, requires users of some of its services to refrain from generating "hateful, or racially, ethnically or otherwise objectionable" content without saying more.[134] Microsoft's gaming service Xbox Live warns users that they may not publish content that "incites . . . hatred [or] bigotry"[135] or that is "related to or suggestive of hate speech (including but not limited to racial, ethnic, or religious slurs)."[136] Some intermediaries attribute their reluctance to address digital hate to the difficulties in defining such speech.[137]

We do not pretend that we can make hard choices easy, nor do we advocate for a particular definition of hate speech. We recognize that intermediaries' decisions will turn on their available resources, business interests, and varied

particular institutional solutions); Frank Pasquale, *Internet Nondiscrimination Principles: Commercial Ethics for Carriers and Search Engines,* 2008 U. CHI. LEGAL F. 263, 268-69 (discussing regulation of search engines and social networks).

[133] TOS agreements typically include not only an intermediary's hate speech policy (if any), but also its privacy policies, which typically notify users that they can opt-out of the collection of personally identifiable information. Commentators have criticized TOS privacy policies on the grounds that users do not pay attention to them and thus do not make meaningful choices about their privacy, which can lead to the collection and use of personal information. Danielle Citron, *The Boucher Privacy Bill: A Little Something For Everyone Yet Nothing for All?*, CONCURRING OPINIONS (June 13, 2010, 11:37 AM), http://www.concurringopinions.com/archives/2010/06/the-boucher-privacy-bill-a-little-something-for-everyone-yet-nothing-for-all.html. Ryan Calo's scholarship thoughtfully responds to critiques of notice provisions. *See, e.g.*, M. Ryan Calo, *Against Notice Skepticism*, 87 NOTRE DAME L. REV. (forthcoming 2012), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1790144#%23.

[134] *Yahoo! Terms of Service*, YAHOO! INC., http://info.yahoo.com/legal/us/yahoo/utos/utos-173.html (last visited Apr. 7, 2011). These policies apply to some of Yahoo!'s services other than its search engine, such as its Flickr photo-sharing service. *Id.* For instance, Yahoo! explains that its Answer application is not "a soapbox to vent personal frustrations or rant about issues. We are a community of people with diverse beliefs, opinions, and backgrounds, so please be respectful and keep hateful and incendiary comments off Yahoo! Answers." *Yahoo! Answers Community Guidelines*, YAHOO! INC., http://answers.yahoo.com/info/community_guidelines (last visited Apr. 7, 2011).

[135] *Xbox LIVE Terms of Use*, MICROSOFT, http://www.xbox.com/en-US/legal/LiveTOU (last visited Apr. 7, 2011).

[136] *Xbox LIVE Code of Conduct*, MICROSOFT, http://www.xbox.com/en-US/legal/codeofconduct (last visited Apr. 7, 2011).

[137] For instance, Twitter's Director of Program Development remarked: "What counts as name calling? There are sites that do employ teams of people that do that investigation . . . but we feel that's a job we wouldn't do well." Anick Jesdanun, *On the Internet, Free Speech Is No Guarantee*, HAMPTONROADS.COM (July 21, 2008), http://hamptonroads.com/2008/07/internet-free-speech-no-guarantee.

assessments of their corporate social responsibility.[138] Instead, we hope to encourage intermediaries – and others – to think and speak more carefully about the harms they hope to forestall when developing hate speech policies. The more intermediaries and users understand *why* a particular policy prohibits a certain universe of speech, the more they will be able to execute the policy in a way that achieves those objectives.[139] This understanding will require intermediaries to engage in thoughtful conversations with stakeholders both externally and internally to identify the particular potential harms of hate speech – and the harms of its constraint – that they find most troubling.

No matter the particular definition of hate speech that intermediaries choose, an accessible and transparent policy can help users develop a better appreciation of their responsibilities as they work, debate, and connect with others online. Hard judgment calls will inevitably remain, regardless of how an intermediary chooses to define hate speech. But those decisions – however difficult – can be made in a more principled way when an intermediary grounds its policy's hate speech definition and application in terms of the specific harms it seeks to avoid. In the next section, we explore a spectrum of definitions available to intermediaries to guide them in this effort.

B. *An Illustrative Definitional Menu*

We propose that an intermediary's voluntary efforts to define and proscribe hate speech should expressly turn on the harms to be targeted and prevented. Rather than identifying new harms, here we rely on thoughtful commentary about First Amendment controversies over proposed governmental regulation of hate speech in outlining a menu of possible harm-based definitions.

[138] *See* Frank Pasquale, *Asterisk Revisited: Debating a Right of Reply on Search Results*, 3 J. BUS. & TECH. L. 61, 73 (2008) (recognizing that resource constraints will limit intermediary duties, but recommending some such duties nevertheless); Frank Pasquale, *Rankings, Reductionism, and Responsibility*, 54 CLEV. ST. L. REV. 115, 117 (2006) (cautioning against too easy acceptance of reductionist presentations of reality by intermediaries).

[139] Along these lines, Lisa Fairfax has documented the extent to which an institution's written commitments – such as a corporation's rhetoric evincing a responsibility to groups and interests beyond their shareholders – encourage and shape its actual behavior. Lisa M. Fairfax, *Easier Said Than Done? A Corporate Law Theory for Actualizing Social Responsibility Rhetoric*, 59 FLA. L. REV. 771, 776 (2007) (debunking "the notion that corporate [social responsibility] rhetoric has no connection to actual practice [and] demonstrat[ing] the manner in which such rhetoric can be used strategically" to shape behavior). Fairfax continues: "when an individual expresses a commitment to a given idea or principle, the human preference for consistency generates internal and external pressures to engage in behavior consistent with that commitment." *Id.* She also points to social psychology literature showing that the "more often someone makes a commitment, the more likely she is to engage in corresponding behavior." *Id.* at 777.

1. Speech that Threatens and Incites Violence

Intermediaries may define prohibited hate speech as that which threatens or encourages violence against individuals. In an area where consensus is exceedingly rare, most commentators seem to agree that these harms are sufficiently serious to warrant prohibiting such speech.[140] Indeed, the United States Supreme Court has held that speech that constitutes a "true threat"[141] or intentional incitement to imminent violence[142] is unprotected by the First Amendment and within the government's power to regulate.

Whether certain speech is likely to incite imminent violence or will lead reasonable people to fear violence will vary with the content and context of the expression.[143] Key factors in making such evaluations include the clarity with which the speech advocates violence and the specificity with which individuals are identified as potential targets. As courts have noted, for example, the inclusion of a target's personal identification information can contribute to a reasonable person's conclusion that the expression communicates the intent to inflict bodily harm upon the target.[144]

[140] *See, e.g.*, John T. Nockleby, *Hate Speech in Context: The Case of Verbal Threats*, 42 BUFF. L. REV. 653, 708 (1994) (urging government regulation of only that universe of hate speech perceived by the listener as threatening violence); Frederick Schauer, *Uncoupling Free Speech*, 92 COLUM. L. REV. 1321, 1349 (1992) (defining actionable hate speech as "first, utterances intended to and likely to have the effect of inducing others to commit acts of violence or acts of unlawful discrimination based on the race, religion, gender, or sexual orientation of the victim; and, second, utterances addressed to and intended to harm the listener (or viewer) because of her race, religion, gender, or sexual orientation").

[141] *See* Virginia v. Black, 538 U.S. 343, 365-66 (2003) (holding that the First Amendment permits states to prohibit individuals from burning crosses only when it is done "with the intent to intimidate."); Watts v. United States, 394 U.S. 705, 705-08 (1969) (defining a true threat as that which a reasonable person would consider an expression of the speaker's intent to inflict bodily harm).

[142] *See* United States v. White, 610 F.3d 956, 957, 962 (7th Cir. 2010) (holding that, under Supreme Court precedent, internet speech in which the poster intends to request or solicit a violent crime is not protected by the First Amendment, and declining to dismiss the government's solicitation case based on the defendant's website that "posted personal information about a juror who served on the Matthew Hale jury, along with postings calling for the use of violence on enemies of white supremacy").

[143] Of course causation remains a challenging issue even under some of the narrower definitions of hate speech. But although we may not be able to say with certainty that certain statements will actually lead to violence, we can be more confident in stating that certain speech will reasonably lead targets to fear such violence.

[144] *See* Planned Parenthood of Columbia/Willamette, Inc. v. Am. Coal. of Life Activists, 290 F.3d 1058, 1080 (9th Cir. 2002) (en banc) (holding that the Nuremberg Files' website could be characterized as an unprotected true threat, where the site listed the names, addresses, and license plate numbers of abortion providers, with the names of those who had been murdered lined through in black, and the names of those wounded highlighted in grey); *see also* United States v. Fullmer, 584 F.3d 132, 156 (3d Cir. 2009) (concluding that animal rights activists' website included expression that instilled fears in its targets and thus could

These factors can help intermediaries determine whether certain situations should be characterized as threats of, or incitement to, violence. Posters on a Yahoo! bulletin board, for instance, listed names of specific Arab-Americans alongside their home addresses, telephone numbers, and the suggestion that they are "Islamic terrorists."[145] There, the targeted individuals notified Yahoo!, which immediately took down the postings.[146] Neo-Nazi Hal Turner's blog postings offer another illustration of targeted speech that threatens or incites violence. A jury convicted Turner in a criminal case based on his postings saying that Judges Frank Easterbrook, Richard Posner, and William Bauer "deserve to be killed," along with the targets' photographs, work locations, and a picture of their courthouse modified to show the locations of "anti-truck bomb barriers."[147]

Intermediaries could also define hate speech as that which urges violence against groups as well as specific individuals. For example, Turner's website also urged readers to murder "illegal aliens": "We're going to have to start killing these people . . . . I advocate using extreme violence against illegal aliens. Clean your guns . . . . Find out where the largest gathering of illegal aliens will be near you . . . and then do what has to be done."[148] In response to similar concerns, Facebook explained that neo-Nazi and other hate groups calling for violence against gypsies,[149] Jews,[150] and even red-headed people[151] violated its hate speech policy.

To be sure, definitional challenges remain under a policy that constrains only hate speech that threatens or incites violence against specific individuals or groups. Of course, some situations present more difficult questions than others. For example, would a reasonable person understand certain online

be prosecuted as true threats unprotected by the First Amendment).

[145] Tom Spring, *Digital Hate Speech Roars*, PC WORLD, (Sept. 21, 2001, 7:00 PM), http://www.pcworld.com/article/63225/digital_hate_speech_roars.html.

[146] *Id.* Another web hosting company took down sites proclaiming that minorities should be hanged. Raphael Cohen-Almagor & Sharon Haleva-Amir, *Bloody Wednesday in Dawson College – The Story of Kimveer Gill, or Why Should We Monitor Certain Websites to Prevent Murder*, 2 STUD. IN ETHICS, L. & TECH. J. no. 3, 2008 at 1, 22-23.

[147] James Joyner, *Hal Turner and the Limits of Free Speech*, OUTSIDE THE BELTWAY (Aug. 16, 2009), http://www.outsidethebeltway.com/hal_turner_and_the_limits_of_free_speech/; *see also* Tom Hays, *NJ Blogger Convicted of Threatening Ill Judges*, ASSOCIATED PRESS, Aug. 13, 2010, *available at* http://www.boston.com/news/local/connecticut/articles/2010/08/13/nj_blogger_convicted_of_threatening_ill_judges.

[148] Susy Buchanan & David Holthouse, *Extremists Advocate Murder of Immigrants, Politicians*, S. POVERTY L. CTR. INTELLIGENCE PROJECT (Mar. 30, 2006), http://www.splcenter.org/intel/news/item.jsp?aid=49.

[149] Robin Pomeroy, *Facebook Pulls Italian Neo-Nazi Pages After Outcry*, REUTERS (Nov. 14, 2008), http://www.reuters.com/article/idUSTRE4AD3KZ20081114.

[150] Lappin, *supra* note 1, at 5.

[151] *See supra* notes 72-73 and accompanying text (discussing Facebook's decision to take down the *Kick a Ginger Day* groups).

speech – such as the use of certain cultural symbols, like nooses, burning crosses, and swastikas[152] – to communicate a true, if implied, threat? As the Supreme Court has observed with respect to cross-burning, some symbols in certain contexts – but not in all contexts – effectively express frightening threats.[153] But contextual inquiry is as inevitable as it is difficult under *any* definition of hate speech. Focusing on the specific harms to be prevented can help us sharpen and justify our inquiry in a principled way.

Some online actors specifically prohibit users from threatening or inciting violence in a manner that helpfully explains their community norms. For instance, Beliefnet, a website devoted to providing information on a wide variety of topics related to faith and spirituality, defines hate speech to mean "speech that may cause violence toward someone (even if unintentionally) because of their age, disability, gender, ethnicity, race, nationality, religion or sexual orientation."[154] The policy explains that unlike mere insults, speech "that may cause violence" includes that which advocates violence against protected class members or states that such violence is "acceptable [or] . . . deserved . . . perhaps by characterizing them as guilty of a heinous crime, perversion, or illness, such that violence may seem allowable or inconsequential."[155] Further boosting its value to users, the policy discusses the reasons underlying the rule,[156] its relationship to free speech guarantees,[157]

[152] *See, e.g.*, Timothy Zick, *Cross Burning, Cockfighting, and Symbolic Meaning: Toward a First Amendment Ethnography*, 45 WM. & MARY L. REV. 2261, 2346-49 (2004) (describing the use of context and cultural meaning to determine whether cross-burning communicates threats of violence or instead political protest).

[153] Virginia v. Black, 538 U.S. 343, 365-66 (2003) (holding that the First Amendment permits states to prohibit individuals from burning crosses "with the intent to intimidate"). Alexander Tsesis argues that cultural symbols of hate, like burning crosses or swastikas, are effective at intimidation because such symbols trigger in victims a well-grounded fear of physical violence. *See* Tsesis, *Dignity and Speech*, *supra* note 67, at 503-04 ("Destructive messages are particularly dangerous when they rely on historically established symbolism, such as burning crosses or swastikas, in order to kindle widely shared prejudices.").

[154] *Hate Speech and the Beliefnet Community*, BELIEFNET, http://www.beliefnet.com/Skipped/2004/06/Hate-Speech.aspx (last visited Apr. 8, 2011).

[155] *Id.*

[156] The website explains that it developed this policy because of its concern that certain forms of hate speech can, and have, inspired violent acts. *Id.*

[157] *Id.* ("Hate speech is legal in the United States. Americans may choose to read or engage in hate speech. Likewise, Americans may choose to gather in groups where they mutually agree upon standards of conduct that do not include hate speech. As a private website, Beliefnet is a choice for those who want civil discussion that is free of hate speech. When speech could incite harm to individuals, harm to the Beliefnet community, or harm to Beliefnet, it is appropriate for us to place limits on it. If you wish to engage in hate speech, there are numerous options available on the Internet. This is not one of them.").

its application to certain challenging contexts (e.g., discussions of homosexuality),[158] and specific practical guidelines for its use.[159]

2. Speech that Intentionally Inflicts Severe Emotional Distress

Along the same lines, intermediaries might define hate speech to include that which intentionally inflicts severe emotional distress. Although this inquiry too is inevitably context-specific, a body of tort law illuminates factors that courts use in determining if speech amounts to intentional infliction of emotional distress.[160] As Benjamin Zipursky explains, "[o]ver decades and even centuries, courts recognized clusters of cases" that constituted extreme and outrageous behavior outside the norms of decency.[161] These most often involve expression that is individually targeted, especially threatening or humiliating, repeated, or reliant on especially sensitive or outrageous material.[162]

[158] *Id.* ("We recognize that many faith groups are engaged in important debate about homosexuality and its relationship to faith. We encourage members to discuss this topic on Beliefnet and have created specific forums for this debate. . . . You may express the belief that homosexuality is wrong, or that it is sinful. . . . You may not advocate violence against anyone because of their sexual orientation.").

[159] *Id.*

[160] Scholars have cautioned that the First Amendment requires a very narrow understanding of this tort to ensure that government does not constrain offensive speech on the basis of viewpoint. *See, e.g.*, Eugene Volokh, *Freedom of Speech and the Intentional Infliction of Emotional Distress Tort*, 2010 CARDOZO L. REV. DE NOVO 300, 300-03; Christina Wells, *Regulating Offensiveness:* Snyder v. Phelps, *Emotion, and the First Amendment*, 1 CALIF. L. REV. CIRCUIT 71, 72 (2010). Along these lines, the Supreme Court has held that the First Amendment prohibits the tort's application to a defendant who "addressed matters of public import on public property, in a peaceful manner, in full compliance with the guidance of local officials." Snyder v. Phelps, 131 S. Ct. 1207, 1220 (2011).

[161] Benjamin Zipursky, Snyder v. Phelps, *Outrageousness, and the Open Texture of Tort Law*, 60 DEPAUL L. REV. (forthcoming 2011) (manuscript at 31), *available at* http://ssrn.com/abstract=1687688.

[162] *See* Citron, *Cyber Civil Rights*, *supra* note 20, 87-88; Smith, *supra* note 86, 35-48; *see also* Nadine Strossen, *The Tensions Between Regulating Workplace Harassment and the First Amendment: No Trump*, 71 CHI.-KENT L. REV. 701, 716-17 (1995) (suggesting that proscribable harassment under Title VII focus on workplace speech that directly targets a particular individual and that is so extreme that it amounts to intentional infliction of emotional distress). Along these lines, intermediaries' policies might also address defamatory hate speech. As the Supreme Court's First Amendment doctrine makes clear, the harms of defamatory speech – i.e., culpably false statements of fact that damage the target's reputation – are sufficiently great to justify its regulation by the government under certain circumstances. *See, e.g.*, N.Y. Times Co. v. Sullivan, 376 U.S. 254, 301-02 (1964); *see also* BOLLINGER, *supra* note 107, at 186 (explaining that his tolerance theory permits the regulation of libel because it targets an individual for harm and the purposes of toleration are not served by insisting that an individual – rather than the community as a whole – bear

Recall, for example, Bonnie Jouhari's experience with digital hate.[163] There, an administrative law judge determined that the website operator intentionally inflicted emotional distress on Jouhari and her daughter through "a relentless campaign of domestic terrorism."[164]

3. Speech that Harasses

Intermediaries might choose to define hate speech as that which would rise to the level of actionable harassment if it occurred at work or in school. Although harassment in the employment and education contexts does not parallel that in cyberspace in important respects,[165] internet intermediaries remain free to consider these efforts when crafting their own policies.

Longstanding anti-harassment principles permit government to regulate harassing speech at work or at school if such speech is sufficiently severe or pervasive to create a discriminatory educational or workplace environment.[166] Factors relevant to assessing whether verbal or written conduct meets this standard include "the frequency of the discriminatory conduct; its severity; whether it is physically threatening or humiliating, or a mere offensive utterance;" and whether it inflicts psychological harm.[167]

In the educational context, for example, verbal or written conduct violates Title IX's statutory prohibitions on discrimination by federally funded educational activities when the "harassment is so severe, pervasive, and objectively offensive that it can be said to deprive the victims of access to the educational opportunities or benefits provided by the school."[168] Along these

the harm of speech activity).

[163] *See supra* notes 86-90 and accompanying text.

[164] Ryan Wilson, HUDALJ 03-98-0692-8, at 19 (July 19, 2000).

[165] *See* Helen Norton, *Regulating Cyberharassment: Some Thoughts on* Sexual Harassment 2.0, 87 DENV. U. L. REV. ONLINE 11, 11-15 (2010) (identifying the difficulties in extending the First Amendment analysis applicable to governmental regulation of harassment at work and school to proposed government regulation of cyber harassment).

[166] *See* R.A.V. v. City of St. Paul, 505 U.S. 377, 389 (1992) (describing Title VII's regulation of harassing speech in the workplace as permissible under the First Amendment as proscribing "sexually derogatory 'fighting words,'" within "Title VII's general prohibition against sexual discrimination in employment practices"); Wisconsin v. Mitchell, 508 U.S. 476, 487 (1993) (explaining that the Court in *R.A.V.* "cited Title VII['s prohibition of sexual harassment] as an example of a permissible content-neutral regulation of conduct").

[167] Harris v. Forklift Sys., Inc., 510 U.S. 17, 23 (1993) (identifying factors relevant to a conclusion that workplace harassment rises to the level of a Title VII violation).

[168] Davis v. Monroe Cnty. Bd. of Educ., 526 U.S. 629, 633 (1999) (interpreting Title VI's prohibition on sex discrimination by federally funded educational activities); *see also* Racial Incidents and Harassment Against Students at Educational Institutions; Investigative Guidance, 59 Fed. Reg. 11448, 11449 (Mar. 10, 1994) (interpreting Title VI's prohibition on race and national origin discrimination by federally funded activities to include "harassing conduct (e.g., physical, verbal, graphic, or written) that is sufficiently severe,

lines, Bryn Mawr College defines harassment to include "verbal behavior such as unwanted sexual comments, suggestions, jokes or pressure for sexual favor; [and] nonverbal behavior such as suggestive looks or leering," and offers as examples "[c]ontinuous and repeated sexual slurs or sexual innuendoes; offensive and repeated risqué jokes or kidding about sex or gender-specific traits; [and] repeated unsolicited propositions for dates and/or sexual relations."[169] The College of William and Mary prohibits "conduct that is sufficiently severe, persistent or pervasive enough so as to threaten an individual or limit the ability of an individual to work, study, or participate in the activities of the College" and defines such conduct to include "making unwanted obscene, abusive or repetitive telephone calls, electronic mail, instant messages, or similar communications with intent to harass."[170]

***

In selecting an appropriate definition of hate speech, intermediaries may draw insight from longstanding First Amendment principles. Indeed, much of the speech described above in Subsections 1 through 3 can be regulated by the government under the First Amendment in certain contexts.[171] That the courts have held that such expression has limited constitutional value suggests that voluntary regulation by private intermediaries may impose comparatively few costs.[172] But as private actors, intermediaries remain unconstrained by the Constitution and are thus legally free to choose to respond to a wider universe of hate speech. The remainder of this Section briefly explores some additional possibilities.

pervasive or persistent so as to interfere with or limit the ability of an individual to participate in or benefit from the services, activities or privileges provided by a recipient").

[169] Letter from Jane McAuliffe, President, Bryn Mawr Coll., to William Creeley, Dir. of Legal and Pub. Advocacy, Found. for Individual Rights in Educ. (July 17, 2010), *available at* http://www.thefire.org/index.php/article/12035.html; *see also* Emory University Residence Life & Housing Standards & Policies 4-5, http://www.emory.edu/HOUSING/FORMS/form_ugrad.html (follow "download" hyperlink beside "Residence Life & Housing Policies") (last visited Apr. 8, 2011) (defining prohibited harassment to include "objectionable epithets, demeaning depiction or treatment, and threatening or actual abuse or harm").

[170] COLLEGE OF WILLIAM AND MARY, STUDENT HANDBOOK 20 n.3 (2010), *available at* http://www.wm.edu/offices/deanofstudents/services/studentconduct/documents/studenthandbook.pdf.

[171] *See supra* notes 141-142 and accompanying text (explaining that true threats and incitement can be prosecuted without running afoul of the First Amendment); *supra* note 160 (explaining that certain speech that intentionally inflicts severe emotional distress can trigger civil liability without running afoul of the First Amendment); *supra* notes 166-167 and accompanying text (explaining that verbal harassment in the workplace that is sufficiently severe or pervasive to alter the terms and conditions of employment can trigger civil liability without running afoul of the First Amendment).

[172] *See supra* notes 110-112 and accompanying text.

4. Speech that Silences Counter-Speech

Intermediaries may define hate speech as including that which silences or devalues its targets' counter-speech. Examples include slurs, insults, and epithets that shut down reasoned discourse, rather than facilitate it. In so doing, intermediaries might draw from private universities' extensive experience in regulating speech of this type, since they – like internet intermediaries – are unconstrained by the First Amendment yet for institutional reasons generally remain deeply attentive to free speech as well as antidiscrimination concerns.

Some private universities, for example, go beyond the anti-harassment requirements of Titles VI and IX in identifying a certain set of community norms to be protected from disruptive speech.[173] Such policies often emphasize a spirit of academic freedom that requires not only a commitment to free discourse, but also an understanding that certain expression can actually undermine that discourse.[174]

Colgate University, for example, articulates its commitment to intellectual inquiry and debate by prohibiting "acts of bigotry" because they "are not part of legitimate academic inquiry."[175] The University emphasizes the contextual nature of this inquiry, noting that prohibited bigotry "has occurred if a reasonable person would have found the behavior offensive and his or her living, learning, or working environment would be impaired," while reserving the right to "discipline offensive conduct that is inconsistent with community standards even if it does not rise to the level of harassment as defined by federal or state law."[176]

[173] As is true with virtually any proposed definition of hate speech, these efforts are not without controversy, as some argue that even private universities' efforts to address hate speech too often unwisely interfere with the unfettered flow of expression. *See, e.g.*, Azhar Majeed, *Defying the Constitution: The Rise, Persistence, and Prevalence of Campus Speech Codes*, 7 GEO. J.L. & PUB. POL'Y 481, 483-84 (2009) (criticizing public and private university efforts to regulate hate speech on campus); Nadine Strossen, *Regulating Racist Speech on Campus: A Modest Proposal*, 1990 DUKE L.J. 484, 488-89.

[174] *See, e.g.*, J. Peter Byrne, *Racial Insults and Free Speech Within the University*, 79 GEO. L.J. 399, 416 (1991) (arguing that the university is "a distinct social entity, whose commitment to enhancing the quality of speech justifies setting minimum standards for the manner of speech among its members").

[175] COLGATE UNIVERSITY STUDENT HANDBOOK 2010-2011, at 112-13 (2010), *available at* http://www.colgate.edu/portaldata/imagegallerywww/939d3f45-4876-4ef5-b567-1082dd4c58e4/ImageGallery/Student_handbook_2010.pdf.

[176] *Id.* at 115. Colgate offers the following as potential examples of impermissible harassment: "using ethnic, racial, religious or other slurs to refer to a person, or jokes or comments that demean a person" on protected bases; "creating or displaying racially, ethnically, religiously offensive pictures, symbols, cartoons, or graffiti;" and "phone calls, emails, text messages, chats or blogs that offend, demean, or intimidate another" on protected bases. *Id.*

Other proposals would similarly permit private universities to punish slurs, insults, and epithets (normally protected by the First Amendment from regulation by public actors), but would otherwise allow speech that invites a response and rational discourse. For example, Peter Byrne argues that access to free speech on campus "should be qualified by the intellectual values of academic discourse," permitting universities to bar racial insults but not "rational but offensive propositions that can be disputed by argument and evidence."[177] He argues that "[r]acial insults have no status among discourse committed to truth. They do not aim to establish, improve, or criticize any proposition."[178] Instead, racial insults simply communicate irrational hatred designed to make the target feel less worthy.[179] Along these lines, intermediaries remain free to define prohibited hate speech as that which shuts down, rather than facilitates, reasoned discourse – e.g., slurs, insults, and epithets.

5. Speech that Exacerbates Hatred or Prejudice by Defaming an Entire Group

An intermediary might choose to focus on speech that more broadly contributes to bigotry and prejudice by defaming an entire group.[180] Jeremy Waldron, for example, seeks to return to an understanding of group defamation's harms as including visible signs that group members may "be subject to abuse, defamation, humiliation, discrimination, and violence."[181] Mari Matsuda similarly characterizes Holocaust denial as a false statement of fact that defames the dead.[182] MySpace apparently adopts a definition along

[177] Byrne, *supra* note 174, at 400; *see also id.* at 415 ("[U]niversities do believe that racial insults are a meritless form of speech that poisons the atmosphere on campus for learning and discussion.").

[178] *Id.* at 419.

[179] *Id.*

[180] For additional proposals along these lines, see, for example, YAMAN AKDENIZ, RACISM ON THE INTERNET 7 (2009) (emphasizing virulent, inflammatory language that is likely to inspire hatred in defining hate speech); Kretzmer, *supra* note 68, at 454 (urging that responsive hate speech policy focus on "threatening, abusive or insulting" speech that "is likely in the circumstances to stir up hatred against a racial, ethnic, or national group"); Matsuda, *supra* note 67, at 2357 (defining hate speech as constituting any message of inferiority, "directed at a historically oppressed group," that is "persecutorial, hateful, and degrading").

[181] Waldron, *supra* note 91, at 1599 (arguing for hate speech regulations that promise that groups will not suffer these injuries).

[182] Matsuda, *supra* note 67, at 2366-67. Jeremy Waldron also urges that we abandon our limited understanding of actionable defamation as concerning false facts about specific individuals, and would instead include the defamation of an entire group through falsehoods. Waldron, *supra* note 91, at 1607-09. For these reasons, he reminds us of the Anti-Defamation League's founding to stop the defamation of the Jewish people because of "the danger that anti-Semitic signage would become an established feature of the landscape

these lines, prohibiting content that targets a group in a way that would make group members "feel bad."[183] Hemanshu Nigam thus describes MySpace's decision to remove Holocaust denial sites as an "easy" call under this conception of hate speech.[184] Other intermediaries take similar approaches.[185]

***

As the discussion above demonstrates, private intermediaries unconstrained by the First Amendment have a wide range of choices when defining hate speech. An intermediary's choice among them depends on a variety of unique institutional values: its assessment of the relative costs of hate speech and its constraint; empirical predictions about what sort of speech is indeed likely to lead to what sorts of harms; its business interests (which, in turn, may be shaped by users' demands and expectations); and the breadth of its sense of corporate social responsibility to address digital hate.

By identifying a spectrum of possible approaches, this Part has sought to provide a framework within which to have these conversations and to make these choices. As the next Part explores, intermediaries also have a wide range of available options when responding to hate speech that violates their chosen policy.

III. RESPONDING TO HATE SPEECH

Many intermediaries have already identified and deployed a number of responses to hate speech. In this Part, we identify promising efforts, critique others, and offer recommendations.

A. *Removing Hateful Content*

The removal of hateful content is the most powerful tool at intermediaries' disposal. Some intermediaries aggressively enforce their hate speech policies by removing offending language, blocking access to sites, or terminating user accounts. For instance, MySpace actively looks for and then deletes pages

and that Jews would have to lead their lives in a community whose public aspect was permanently disfigured in this way." *Id.* at 1610.

[183] Interview with Nigam, *supra* note 59.

[184] *Id.*

[185] Under the title "Don't be sexist, racist, or a hater," Digg describes its hate speech policy as: "Would you talk to your mom or neighbor like that? Digg defines hate speech as speech intended to degrade, intimidate, or incite violence or prejudicial action against members of a protected group. For instance, racist or sexist content may be considered hate speech." *Community Guidelines*, DIGG, http://about.digg.com/guidelines (last visited Apr. 8, 2011). YouTube appears to take a similar definitional approach. *YouTube Community Guidelines*, YOUTUBE, http://www.youtube.com/t/community_guidelines?gl=GB&hl=en-GB (last visited Apr. 8, 2011) ("We encourage free speech and defend everyone's right to express unpopular points of view. But we don't permit hate speech (speech which attacks or demeans a group based on race or ethnic origin, religion, disability, gender, age, veteran status, and sexual orientation/gender identity).").

and/or bans users who "promote[] or otherwise incite[] racism, bigotry, hatred or physical harm of any kind against a group or individual" or who "exploit[] people in a sexual or violent manner."[186] Other intermediaries apparently define removable hate speech more narrowly – for example, only where threats of violence are involved.[187]

Removal's enormous power counsels against its overuse, as speakers' access to certain communities can depend upon the cooperation of intermediaries. While intermediaries can prominently display websites and blogs, they can also prevent people from accessing them.[188] Thoughtful and effective responses thus do not, and should not, always require removal.

In our view, intermediaries should consider blocking forms of hate speech that satisfy certain of the narrower definitions described in Part II.B – that is, expression that is more directly related to threats of, or incitement to, violence and intentional infliction of emotional distress, and for these reasons may be

[186] *Terms & Conditions*, MYSPACE, http://www.myspace.com/help/terms (last visited Apr. 8, 2011); *see also* Nora Flanagan, *Social Networking: A Place for Hate?*, IMAGINE 2050 (May 19, 2009), http://imagine2050.newcomm.org/2009/05/19/social-networking-a-place-for-hate (describing MySpace's strict enforcement of its hate speech policy); Interview with Nigam, *supra* note 59. MySpace employs forum moderators who "keep an eye out for anti-semitism and derogatory comments." Michael Arrington, *MySpace Wants to Avoid this Whole Holocaust Denial Thing*, TECHCRUNCH BLOG (May 12, 2009), http://techcrunch.com/2009/05/12/myspace-wants-to-avoid-this-whole-holocaust-denial-thing/. Its terms of service explains that it "expressly reserves the right to remove your profile and/or deny, restrict, suspend, or terminate your access to all or any part of the MySpace Services if MySpace determines, in its sole discretion, that you have violated this Agreement." *Terms & Conditions*, MYSPACE, http://www.myspace.com/help/terms (last visited Apr. 8, 2011).

[187] *See supra* notes 149-151 and accompanying text (discussing Facebook's removal of pages threatening violence like *Kick a Ginger*). Assuming that Facebook understands removable hate speech to mean only that which threatens violence, it should say so more clearly in its actual hate speech policy, which instructs users not to "post content that: is hateful, threatening, or pornographic; incites violence; or contains nudity or graphic or gratuitous violence." *Statement of Rights and Responsibilites*, FACEBOOK, http://www.facebook.com/terms.php (last visited June 1, 2011).

[188] Video-sharing services and social network sites can remove content, precluding users from seeing them. Social media services can ban users by blocking their IP addresses. *Cf. Google and Internet Control in China: A Nexus Between Human Rights and Trade?: Hearing Before the Cong.-Exec. Comm'n on China*, 111th Cong. 68-76 (2010) (prepared testimony of Rebecca MacKinnon) (exploring the Chinese government's efforts to censor its citizens' online activities, including through the use of IP address blocking). Search engines can refuse to sell advertising to companies and thus limit their visibility to customers engaging in relevant searches. *See* Floyd Norris, *France Calls Google a Monopoly*, N.Y. TIMES, July 2, 2010, at B1 (describing how Google refused to sell online advertising to French company Navx, which lets French drivers know where the police operate radar traps, because "Google found Navx's business distasteful" – thus search terms like "radar trap" no longer yielded advertisements for the company's product, whose sales "plunged").

subject even to government regulation under the First Amendment.[189] Removal may be especially appropriate where counter-speech is unlikely to eliminate the harms posed by the hateful expression.

Calls for violence strike at the very heart of digital citizenship. They can inspire actual physical attacks.[190] Threats of violence also violate principles of digital citizenship even if they do not directly lead to actual violence, such as YouTube's *How to Kill a Beaner* or *Execute the Gays* videos,[191] because they deny group members the opportunity to engage in activities free from fear. In our view, Facebook and YouTube appropriately removed these and similar postings as soon as they received notice of them.[192] Threats and encouragement of violence undermine their targets' security and peace of mind, without facilitating discourse. Moreover, intermediaries generally can surgically remove threats of violence with little risk to other speech.[193]

Online hate that inflicts severe emotional distress accomplishes a similar denial of digital citizenship. For instance, recall that persistent and menacing online harassment coerced a California teenager into closing his website and leaving his school.[194] Similar results followed the attacks on Kathy Sierra: she shut down her well-known blog after anonymous posters uploaded doctored photographs, revealed her home address and Social Security number, and threatened violence.[195] This type of online hate has little chance of generating

[189] *See supra* notes 141-142 and accompanying text.

[190] *See, e.g.*, Nordlinger, *supra* note 11, at 8; Moore, *supra* note 11.

[191] Howard, *supra* note 7, at 4D.

[192] *Id.* (reviewing Facebook and Google/YouTube policies on removal of hate speech); Lappin, *supra* note 1 (chronicling Facebook's removal of "Kill a Jew" pages).

[193] Intermediaries also employ other strategies in addition to removal. For instance, they might accompany the removal of speech that violates their TOS with other sanctions, including warnings followed by temporary or permanent banning of individual users found in violation. *Cf.* David A. Hoffman & Salil K. Mehra, *Wikitruth Through Wikiorder*, 59 EMORY L.J. 151, 182 & n.162 (2009) (discussing Wikipedia's warning of users to stop certain behaviors and placement on probation as well as banning of users).

[194] *See supra* notes 74-79 and accompanying text.

[195] *See supra* notes 80-85 and accompanying text. Chris Locke operated the blog where the threatening comments and doctored photographs were posted. Chris Locke, *Re Kathy Sierra's Allegations*, THE EGR WEBLOG (Mar. 27, 2007, 3:16 AM), http://www.rageboy.com/2007/03/re-kathy-sierras-allegations.html. He summarized his reaction to the posts in this way:

> [T]here were a couple posts – the ones Kathy mentions in her post – that were over the top. I didn't think for a minute that they were "threatening" – and again, they were not my doing – but when I saw mail from her objecting to them, I nuked the entire site rather than censor any individual.
>
> I was a conference host on the Well 15 years ago where the core ethos was acronymized to YOYOW – You Own Your Own Words. This has remained a guiding principle for me ever since. I will not take responsibility for what someone else said, nor will I censor what another individual wrote. However, it was clear that Sierra was upset, so it seemed the best course to make the whole site go away.

counter-speech – it seems designed to cause deep distress, not to generate dialogue.

Even when content is appropriately removed, however, acknowledging its deletion can support a commitment to transparent and accountable enforcement.[196] For example, Facebook readers could see Facebook's acknowledgment that it took down *Kick a Ginger Day* and *Kill a Jew Day*.[197] And although Google's search engine does not take down hateful content in the United States as a matter of policy, it alerts web users of content removal when the law of another country requires it to block in-country users from certain sites otherwise available on the internet.[198]

B. *Countering Hate Speech with Speech*

Rather than – or in addition to – the removal of online hatred, intermediaries can counter digital hate with speech of their own. Google offers an instructive – if rare – example. In 2004, the number-one Google result for a search of "jew" was the URL jewwatch.com, a site featuring anti-Semitic content.[199] In response, a Jewish activist asked people around the Web to link the word "jew" to a Wikipedia article so that search engine users would more likely see that article at the top of search results rather than the Jew Watch site, a practice known as a "Googlebomb."[200] Neo-Nazi sites, in turn, launched a counter-

Id. In our view, Locke, as the blog operator and relevant intermediary, should have taken down the posts as soon as he saw them in light of their clear potential to threaten and inflict fear and distress without offering any genuine opportunity for dialogue. Locke's "You Own Your Own Words" philosophy, applied there, seems ironic given that the posters wrote anonymously and thus avoided owning their own words to avoid bearing responsibility for the threats and doctored photographs. *See id.*

[196] *See State AG Questions Research on Child-to-Child Online Bullying*, WASHINGTON INTERNET DAILY (Warren Comm'ns News, Inc., Washington, D.C.), Dec. 12, 2008.

[197] *See 'Kill a Jew Day': Spike in Virulent Anti-Jewish Facebook Pages*, THE NEW YORK BLUEPRINT (Oct. 3, 2010), http://nyblueprint.com/articles/view.aspx?id=796. Unfortunately, however, "Kill a Jew" groups continue to appear on Facebook. *See* Anomaly100, *Two Dozen 'Kill a Jew Day' Pages Found in the Last Seven Days*, FREAKOUTNATION (Oct. 3, 2010, 6:09 PM), http://freakoutnation.com/2010/10/03/two-dozen-facebook-kill-a-jew-day-pages-found-in-the-last-seven-days.

[198] Seltzer, *supra* note 23, at 46-47 ("[W]hen sites are blocked at a search-engine level, it is up to the search providers to notify their end-users. If they do not, the page disappears invisibly. In most engines, pages simply disappear from listings, leaving searchers unaware that a site they never saw is gone. . . . Among the major search engines, only Google gives indication when it removes results from a search page because of legal demands."). Of course, Google's current policy might stem from its objection to those countries' restrictive laws. Whatever the reasons underlying its policy, we still laud the company for the transparency of those decisions to remove content.

[199] *See* JEW WATCH, http:/jewwatch.com (last visited Apr. 8, 2011). *See generally Jew Watch*, WIKIPEDIA, http://en.wikipedia.org/wiki/Jew_Watch (last visited Apr. 8, 2011).

[200] John Brandon, *Dropping the Bomb on Google*, WIRED (May 11, 2004), http://www.

Googlebomb, leading the results back to Jew Watch.[201] Individuals asked Google to remove Jew Watch entirely from its search results.[202]

After the story drew significant media and interest-group attention, Google announced that it would not change its software to eliminate Jew Watch in its results pages.[203] It explained that it chose not to change its algorithms because it "views the comprehensiveness of [its] search results as an extremely important priority," and it does not "remove a page from [its] search results simply because its content is unpopular or because we receive complaints concerning it."[204]

Instead, Google inserted its own advertisement entitled "Offensive Search Results" on top of its page where the link to Jew Watch appeared among other search results.[205] Google explained the company's understanding that the Jew Watch site may be offensive and "apologize[d] for the upsetting nature of the experience you had using Google."[206] Google assured readers that it did not

wired.com/culture/lifestyle/news/2004/05/63380. "Googlebombing" refers to a practice in which users can artificially inflate a page's search ranking by linking to a page in as many other pages as possible. James Grimmelmann, *The Google Dilemma*, 53 N.Y.U. SCH. L. REV. 939, 942-43 (2008-09); *see* Bracha & Pasquale, *supra* note 128, at 1167-88 (describing search engines' capacity to manipulate their results).

[201] Grimmelmann, *supra* note 200, at 943.

[202] *Id.*

[203] *An Explanation of Our Search Results*, GOOGLE, http://www.google.com/explanation.html (last visited Apr. 8, 2011).

[204] *Id.* Apparently, however, Google does change search results for at least some purposes. Consider the example of a merchant who deliberately engaged in bad behavior because the sheer volume of negative mentions that then appeared on consumer advocacy websites improved his ranking in search results. David Segal, *A Bully Finds a Pulpit on the Web*, N.Y. TIMES, Nov. 28, 2010, at BU1. In response, Google changed its algorithm to penalize sites that others link because it provided "extremely poor user experience." Amit Singhal, *Being Bad to Your Customers Is Bad for Business*, THE OFFICIAL GOOGLE BLOG (Dec. 1, 2010, 12:06 PM), http://googleblog.blogspot.com/2010/12/being-bad-to-your-customers-is-bad-for.html. In a blog posting, Google explained that it developed an algorithmic solution to ensure that "being bad is, and hopefully will always be, bad for business in Google's search results." *Id.* For another example, see David Segal, *The Dirty Little Secrets of Search*, N.Y. TIMES, Feb. 13, 2011, at BU1 (discussing Google's changes in search results to counter the effects of manipulative efforts to maximize J.C. Penney's search result rankings).

[205] Google Search for "Jew", GOOGLE, http://www.google.com/search?q=jew (last visited Mar. 26, 2011).

[206] *An Explanation of Our Search Results*, *supra* note 203. If, however, you type "jew" into Google's German version, google.de, Jew Watch does not appear at all. Grimmelmann, *supra* note 201, at 948. At the bottom of the results page, a notice explains that Google has removed three results from the page. *Id.* Google changed its results because German law criminalizes incitement of hatred against minorities. *Id.* at 947. For a discussion of whether and how countries that have experienced genocide may take more aggressive approaches to hate speech, see Jennifer M. Allen & George H. Norris, *Is Genocide Different? Dealing*

endorse the views expressed by Jew Watch.[207] Google's explanation added that readers "may be interested in some additional information the Anti-Defamation League has posted about this issue."[208] To date, however, Jew Watch continues to appear prominently in a Google search of "jew."[209]

Google similarly inserted an explanatory advertisement after images of the First Lady, altered to resemble a monkey, prominently appeared among the results of Google image searches for "Michelle Obama."[210] After Google posted its advertisement, a Chinese blog that had recently featured the image took it down, saying, "I am very sorry for this article."[211]

This kind of intermediary counter-speech is, however, far from routine. For instance, although Google has a "Report Offensive Image" function, it rarely responds to such reports, and Google has to date bought "Offensive Search Results" advertisements in only the cases discussed here.[212] Such counter-speech by intermediaries thus remains extremely rare.[213]

with Hate Speech in a Post-Genocide Society, 7 J. INT'L L. & INT'L REL. (forthcoming 2011), *available at* http://ssrn.com/abstract=1640812.

[207] *An Explanation of Our Search Results*, *supra* note 203.

[208] *Id.*

[209] Google Search for "Jew", *supra* note 205 (showing Jew Watch as the second result).

[210] Saeed Ahmed, *Google Apologizes for Results of 'Michelle Obama' Image Search*, CNN (Nov. 25, 2009, 12:05 PM), http://www.cnn.com/2009/TECH/11/25/google.michelle.obama.controversy-2/index.html. A Google forum user flagged the picture. *Id.* Initially, Google de-indexed the website that posted the photograph on the grounds that "it could spread a malware virus." *Id.*

[211] Bianca Bosker, *Michelle Obama Pictures UPDATE: Offensive Image REMOVED, Google 'SORRY'*, HUFFINGTON POST (updated Mar. 18, 2010, 5:12 AM), http://www.huffingtonpost.com/2009/11/24/michelle-obama-photo-goog_n_368760.html.

[212] *See* Barry Schwartz, *Report Offensive Images on Google Does Not Work*, SEARCH ENGINE ROUNDTABLE (Apr. 13, 2010, 7:54 AM), http://www.seroundtable.com/archives/022010.html. Google's inaction on other cases has sparked much criticism. *See, e.g.*, Esra'a Al Shafei, *Google Apologizes for Offending Jews Through Search Results*, MIDEAST YOUTH (Mar. 21, 2007), http://www.mideastyouth.com/ 2007/03/21/google-apologizes-for-offending-jews-through-search-results.

[213] Google has bought ads in at least one other instance outside of the context of hate speech: a Google user searching for "suicide" will encounter Google ads featuring suicide prevention resources. Noam Cohen, *'Suicide' Query Prompts Google to Offer Hotline*, N.Y. TIMES, Apr. 5, 2010, at B6. The "icon of a red phone and the toll-free number for the National Suicide Prevention Hotline" appear over the linked results in a way that is "different and more prominent than an advertisement." *Id.* Google has also provided the telephone number for national poison control in searches like "poison emergency." *Id.* MySpace has gone further than putting up advertisements when users write about suicide. Interview with Nigam, *supra* note 59. As Hemanshu Nigam explained, when MySpace identified, or received notice of, users noting a desire to commit suicide, it would contact the National Suicide Prevention hotline and local police to recruit help for the users. *Id.* According to Nigam, MySpace's intervention helped prevent ninety-three suicides in 2009. *Id.*

To be sure, we recognize – and remain concerned by – the possibility that counter-speech may shine a spotlight on, and thus bring more attention to, digital hate. But silence in response to digital hate carries significant expressive costs as well. When powerful intermediaries rebut demeaning stereotypes (like the Michelle Obama image) and invidious falsehoods (such as Holocaust denial), they send a powerful message to readers. Because intermediaries often enjoy respect and a sense of legitimacy, users may be inclined to pay attention to their views.[214] With counter-speech, intermediaries can demonstrate by example what it means to treat others with respect and dignity.

Moreover, such counter-speech can expose digital citizens to diverse views, piercing the insularity of hateful messages that may lead to more extreme views. This is just the sort of strategy Cass Sunstein alludes to in his book, *Republic.com 2.0*, where he calls for "self-conscious efforts by private institutions" to expose citizens to diverging views.[215] He urges intermediaries to adopt best practices that expose citizens to different perspectives on public issues, such as through "creative use of links to draw people's attention to multiple views."[216]

By challenging hate speech with counter-speech, intermediaries can help transform online dialogue by documenting the continuing existence of racism and other forms of hatred while concomitantly rebutting it. In this way, intermediary action may help develop the qualities of tolerance advocated by Lee Bollinger,[217] while repairing the public discourse by speaking for silenced or devalued targets. Intermediaries could play a valuable role in challenging hate without defusing the safety valve and other attributes of permitting the

[214] For additional arguments of the value of counter-speech by powerful speakers in response to hate, see Corey Brettschneider, *When the State Speaks, What Should It Say? The Dilemmas of Freedom of Expression and Democratic Persuasion*, 8 PERSP. ON POL. 1005, 1005 (2010) (urging that "a proper theory of the freedom of expression obligates the legitimate state" to respond to hateful but protected speech by emphasizing the importance of respect for equality and dignity); Helen Norton, *Campaign Speech Law with a Twist: When Government Is the Speaker, Not the Regulator*, 61 EMORY L.J. (forthcoming 2011) (urging government to engage in political speech on contested ballot measures that counters that of powerful private speakers); Charlotte H. Taylor, *Hate Speech and Government Speech*, 12 U. PA. J. CONST. L. 1115, 1188 (2010) (urging government – generally prohibited by the First Amendment from banning hate speech – to engage in counter-speech to "help forge consensus about the nature of social practices" and "shift the ground under the hateful speaker's feet, robbing her of her confidence that she can invoke an entire system of subordination by using a few cheap words").

[215] SUNSTEIN, *supra* note 100, at 191.

[216] *Id.* at 192, 200-01, 208 (calling for radio stations, television stations, and newspapers to provide links to diverse views on their online sites).

[217] BOLLINGER, *supra* note 107, at 172-73 (arguing that tolerating the expression of hatred may actually enhance our intellectual capacities and embolden civic courage).

expression of hateful views.[218] Importantly, intermediaries that respond to hate speech through forceful counter-speech or in some other way short of removal appear to trigger few, if any, of the expressive concerns about intermediaries' voluntary measures identified above.[219]

In some respects, Facebook's response to Holocaust denial groups illustrates a missed opportunity for meaningful counter-speech. Facebook vigorously defended its refusal to take down the sites on the grounds that such refusal allows people to see that the sites' proponents are "stupid."[220] Facebook, however, could have explained to its users through counter-speech *why* it views those sites as "stupid." Many other instances of hate – from demeaning characterizations of groups[221] and individuals[222] to falsehoods meant to inspire hate[223] – similarly invite intermediaries' counter-speech.

To be sure, an intermediary's ability to respond to cyber hate will inevitably depend on available resources. Indeed, cyber hate's exponential growth could overwhelm intermediaries interested in engaging in counter-speech. Nonetheless, the ability to automate functions like searching for key terms and inserting prepared responses may help cut down on costs.[224]

Given limited resources, intermediaries might attend carefully to hate speech targeted at children given electronic media's profound impact on children's behavior and views.[225] Indeed, some hate sites are designed specifically to influence youths.[226] *MartinLutherKing.org*, a hate site, is

[218] *See, e.g.*, Blasi, *supra* note 107, at 408 (highlighting Bollinger's recognition "of the safety-valve function of letting discontent surface"); Hughes, *supra* note 107, at 365 (suggesting that hate speech regulation creates martyrs and converts to the cause of hatred); Lidsky, *supra* note 107, at 1099-1100 (concluding that punishing Holocaust denial will paradoxically entrench that view and inspire stronger belief in conspiracy theories).

[219] *See supra* notes 110-124 and accompanying text.

[220] Michael Arrington, *Facebook Remains Stubbornly Proud of Position on Holocaust Denial*, TECHCRUNCH (May 12, 2009), http://techcrunch.com/2009/05/12/facebook-remains-stubbornly-proud-of-position-on-holocaust-denial.

[221] *See, e.g.*, *Common Pro-N----- Arguments*, N-----MANIA, http://niggermania.com/tom/niggerarguments/niggerargumentstextpagetwo.htm (last visited Mar. 27, 2011) ("We hate n[-----]s" because they are a "failed ape species.").

[222] *See, e.g.*, *The Beast as Saint: The Truth About "Martin Luther King, Jr."*, http://www.martinlutherking.org/thebeast.html (last visited Apr. 8, 2011) (arguing that Dr. Martin Luther King, Jr. was an academic cheat, communist, and sex addict).

[223] *See, e.g.*, JEW WATCH, *supra* note 199.

[224] Those costs would be comparatively minor in instances where an intermediary can automate counter-speech. Although an intermediary would need to incur the fixed cost of designing, or purchasing, responsive software, it would incur virtually no expenditures for the software's implementation in future cases. *Cf.* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1284-85 (2008).

[225] *See* Ybarra et al., *supra* note 96, at 929.

[226] LORRAINE TIVEN & PARTNERS AGAINST HATE, HATE ON THE INTERNET: A RESPONSE GUIDE FOR EDUCATORS AND FAMILIES 20, 21 (2003) ("[J]ust as fashion editors and e-book

directed at students researching the civil rights movement.[227] Neo-Nazi adherent Vincent Breeding (credited by Partners Against Hate as creating and maintaining the site) writes: "If you are a teacher or student, I hope you will take a stand for right and wrong and use this information to enlighten your peers."[228] The Klan has Youth News video games available online; some are ethnic cleansing games.[229] Hateful games aimed at young people have seeped into the mainstream where they are either hosted or reviewed on regular gaming sites.[230] Because children are particularly vulnerable to influence, intermediaries might thus be quicker to challenge hate speech that targets them.[231] Hate speech can teach children that prejudice is socially acceptable. Hate speech that condones violence against group members, notably ethnic cleansing games,[232] is especially troubling because video games that engage in fantasies about killing group members can desensitize children to violence and promote violent behavior.[233] Counter-speech – and, indeed, sometimes the removal of such speech altogether – is thus especially important with respect to hate speech that targets children.

C. *Educating and Empowering Community Users*

Just as we see with other mediating institutions like schools, workplaces, and churches, intermediaries can help develop an understanding that citizenship – here, digital citizenship – should include attention to the dignity and safety of other users. Educators, supervisors, and pastors have long played this sort of role with regard to bullying – they endeavor to teach children and adults alike how to treat others with respect.[234] Intermediaries can play a similar role with regard to online hatred.

publishers have started reaching out to elementary school children and teens . . . so have hate groups." (quoting Tara McKelvey, *Father and Son Target Kids in a Confederacy of Hate*, USA TODAY, Jul. 16, 2001, at 3D)).

[227] *Id.* at 20.

[228] *Id.*

[229] SIMON WIESENTHAL CENTER, *supra* note 9 (displaying screenshots of video games based in hate speech)

[230] *Id.*

[231] *See* Shiffrin, *supra* note 74, at 89 ("As an African American father once said to me when I spoke about the contribution of racist speech to the democratic dialogue, 'Tell that to my seven-year-old daughter.'").

[232] *See supra* notes 9-10 and accompanying text (discussing video games posted on YouTube and neo-Nazi social network sites).

[233] Social science research demonstrates the significant harm caused by exposing children to violence. *See, e.g.*, Amitai Etzioni, *On Protecting Children from Speech*, 79 CHI.-KENT L. REV. 3, 36-37 (2004).

[234] *See, e.g.*, Susan Engel & Marlene Sandstrom, *There's Only One Way to Stop a Bully*, N.Y. TIMES, July 23, 2010, at A23.

1. Education

Intermediaries' educational efforts can take a variety of forms. For example, intermediaries can valuably educate their users about digital citizenship norms by more transparently explaining their enforcement decisions. They can offer examples of instances when they did, and did not, remove contested content, along with their reasoning. Intermediaries with similar priorities could join forces in drafting a set of principles and explanatory examples.[235] Just as the preceding Part urged greater transparency and specificity when identifying the harms to be targeted – and thus the objectives to be achieved – by a particular hate speech policy and definition, this Part highlights the value of greater transparency when explaining the reasons behind certain decisions enforcing these policies.

For example, Facebook, MySpace, YouTube, AOL, and other intermediaries currently devote significant staff and energy addressing abuse complaints.[236] Yet their actual practices – that is, what decisions they actually make and how – remain unclear.[237] As part of a commitment to transparent policy implementation, they could explain the grounds of certain decisions, including the definition of hate speech that they employed and specific examples of the harms that they sought to forestall in rendering those decisions. The more clearly and specifically that intermediaries identify and explain their approach to hate speech, the more informed users' choices will then be about the sort of online community with which they choose to interact. The Beliefnet policy discussed in Part II provides a helpful illustration.[238]

Intermediaries can also engage in efforts to educate the public more broadly about hate. For instance, YouTube's Safety & Security Center features information and links to resources developed by the Anti-Defamation League (ADL) to help internet users respond to and report offensive material and

[235] This recalls the international standards organization for the World Wide Web – the W3C group – that identifies voluntary standards. *See W3C Mission*, WORLD WIDE WEB CONSORTIUM, http://www.w3.org/Consortium/mission (last visited Apr. 8, 2011). We thank Neil Richards, Berin Szoka, and Chris Wolf for their helpful thoughts on this notion.

[236] *See, e.g.*, *Fourth Law and Information Society Symposium: Hate Versus Democracy on the Internet*, FORDHAM LAW EVENT CALENDAR (Mar. 26, 2010), http://law2.fordham.edu/ihtml/cal-2uwcp-calendar_viewitem.ihtml?idc=10320.

[237] Indeed, clearer and more transparent policies might have averted the situation where Facebook pulled Sarah Palin's controversial-but-not-hateful posting about proposals to build a mosque near Ground Zero after a number of users responded to a campaign encouraging them to click the "Report Note" hyperlink indicating the posting as hate speech. When Palin questioned the action, Facebook put it back up, apologized for pulling the comment, and promised to modify their process for taking down postings. *See* Brian Ries, *My Facebook War with Palin*, THE DAILY BEAST (JULY 23, 2010, 11:10 AM), http://www.thedailybeast.com/blogs-and-stories/2010-07-23/palins-facebook-ground-zero-mosque-post-how-it-disappeared/full/. Transparent policies might thus have additional salutary effects: the prevention of user manipulation of intermediaries' reporting tools.

[238] *See supra* notes 154-159 and accompanying text.

extremist content that violates YouTube's Community Guidelines on hate speech.[239] It includes tips from the ADL on how to confront hate speech, including flagging offensive videos for review by the YouTube team, posting videos or comments that oppose the offensive point of view, and talking to friends, family, and teachers about what they have seen.[240] As one more of the many ways that intermediaries might help educate users about the impact and treatment of cyber hatred, intermediaries might also consider funding cyber literacy campaigns to teach students about digital citizenship.

2. Empowerment

Empowering users to respond to hate speech on their own sites and to report Terms of Service violations can help communicate and enforce community norms and expectations of digital citizenship.[241] As Clay Shirky observes:

> Any group trying to create real value must police itself to ensure it isn't losing sight of its higher purpose . . . . Governance in such groups is not just a set of principles and goals, but of principles and goals that have been internalized by the participants. Such self-governance helps us behave according to our better natures.[242]

Note, however, that such efforts are most likely to be effective when intermediaries have educated their users and enforcement personnel about the specific harms to be addressed by their specific hate speech policy.

How can an intermediary help its users internalize norms of digital citizenship? As Shirky explains, communities that permit "mutually visible action among the participants, credible commitment to shared goals, and group members' ability to punish infractions" create contexts in which users "can do a better job both in managing the resource and in policing infractions than can markets or government systems designed to accomplish the same goals."[243]

[239] *Safety Center: Hateful Content*, YOUTUBE, http://www.google.com/support/youtube/bin/answer.py?hl=en&answer=126264 (last visited Apr. 8, 2011).

[240] *Id.*

[241] *See, e.g.*, Jon M. Garon, *Wiki Authorship, Social Media, and the Curatorial Advantage*, 1 HARV. J. SPORTS & ENT. L. 95, 99 (2010) ("By expanding opportunity for interaction and fostering behavioral norms of trust among users, these communications tools can expand the reach of social networks for mutual advantage.").

[242] CLAY SHIRKY, COGNITIVE SURPLUS: CREATIVITY AND GENEROSITY IN A CONNECTED AGE 165 (2010); *see also id.* at 177 ("Unlike personal or communal value, public value requires not just new opportunities for old motivations; it requires governance, which is to say ways of discouraging or preventing people from wrecking either the process or the product of the group.").

[243] *Id.* at 113; *see also* ELINOR OSTROM, GOVERNING THE COMMONS: THE EVOLUTION OF INSTITUTIONS FOR COLLECTIVE ACTION 88-102 (1990) (identifying the factors key to community regulation of common resources to include institutions well-equipped to gather information about the resource, forums to discuss its management, community participation in developing and enforcing the rules, and appropriate and graduated sanctions to discipline

Intermediaries will likely have greater success setting norms if they contain code designed to foster social governance, such as reputation scoring systems.[244]

The Wikipedia experience provides a powerful example of such dynamics in action to foster effective online norms of good behavior. As Jonathan Zittrain explains, Wikipedia's key distinguishing attributes – and one that may explain much of its success – included its initial core of editors who shared a "common ethos" and then shared those behavioral norms with new users "through informal apprenticeships as they edited articles together."[245] These norms include administrators' power to create locks to prevent misbehaving users from editing and to ensure that articles prone to vandalism are not subject to changes by unregistered or recently registered users.[246] Users acquire such administrative powers "by making lots of edits and then applying for an administratorship" – that is, by demonstrating their compliance with community norms.[247]

Moreover, Wikipedia enlists volunteer editors called "Third Opinion Wikipedians" who resolve disputes between editors.[248] As David Hoffman and Salil Mehra document, Wikipedia's guidelines urge Third Opinion Wikipedians to "read the arguments, avoid reckless opinions, be civil and nonjudgmental, offer neutral opinions, and monitor the page after offering an opinion."[249] Wikipedia also permits users to report impolite, uncivil, or other

abuse).

[244] Amazon, eBay, Craigslist, and other commercial sites permit users to rate other users or to flag potential misbehavior. Kahn, *supra* note 25, at 198-201 (describing Wikipedia and eBay's use of community norms to police users' behavior); Kim, *supra* note 18, at 1016. Daniel Kahn similarly observes that sites like Wikipedia and eBay that successfully rely on community norms to encourage good behavior share a few key characteristics: "the sites provide easy methods for users to view each others' reputational information"; "reputational information is reciprocal: those who wish to comment on others' behavior must also open themselves to being rated"; "the sites do not merely expect norms to emerge in a vacuum, but instead contain code designed to help foster social governance"; and "they give users incentives to opt into the norm system and to take it seriously." Kahn, *supra* note 25, at 202-03. In response to Neil Netanel's assertion that the internet is not the sort of environment in which norms can generally shape behavior, Netanel, *supra* note 27, at 432, Kahn replies that "the Web is no longer simply too big to handle norms" because social intermediaries enable the formation of smaller communities of manageable size. Kahn, *supra* note 25, at 235.

[245] JONATHAN ZITTRAIN, THE FUTURE OF THE INTERNET – AND HOW TO STOP IT 134-35 (2008). These norms include "the three-revert rule," in which "an editor should not undo someone else's edits to an article more than three times in one day." *Id.* at 135.

[246] *Id.*

[247] *Id.* at 135-36.

[248] Hoffman & Mehra, *supra* note 193, at 172-73.

[249] *Id.* (explaining that Third Opinions are provided under separate headings from the original disputes). Wikipedia has an Arbitration Committee, whose elected members adjudicate disputes between users. *Id.* at 154. Hoffman and Mehra explain that while the

difficult communications with editors in its Wikiquette alerts notice board.[250] On the non-binding Wikiquette alerts page, users seek advice, informal mediation, or referrals to a more appropriate forum.[251] The Wikiquette alerts page also explicitly asks those who have benefited from the process to contribute to other alerts.[252] The Wikipedia model may prove helpful to intermediaries when devising systems for responding to user's abuse reports about cyber hate.

Intermediaries could rely on users to help them identify and respond to hateful content. Currently, many intermediaries depend upon users to report prohibited content, which their employees then address. YouTube's global communications director Scott Rubin has explained that the company cannot "prescreen" content because "[t]here are 20 hours of video uploaded to our site every minute." [253] According to Rubin, YouTube counts on its community to "know the guidelines and to flag videos that they believe violate guidelines."[254] YouTube also offers to its users a Safety Mode tool that blocks videos with objectionable material and encourages users to address hate speech appearing on their own profiles.[255] It reminds users that they can remove others hateful comments from their videos and moderate comments on their channels.[256]

A few intermediaries even allow users to make initial decisions about whether material ought to appear online. For example, Mozilla, the developer of the web browser Firefox, allows users to personalize their browser with

Arbitration Committee generates norms, its task is to rule on specific cases and set forth concrete rules on how users should behave. *Id.* The Arbitration Committee has sanctioned users who make "homophobic, ethnic, racial or gendered attacks" or who are stalkers and harassers. *Id.* at 180. The Arbitration Committee can ban individuals from participation on all or part of the site or place them on probation. *Id.* at 182. Generally speaking, there is a 63% chance that the arbitrators caution the parties or impose probations, and a 16% chance that they will ban a party from the site. *Id.* at 184. In cases when either impersonation or anti-social conduct like hate speech occurs, the Administrative Committee will ban the user in 21% of cases. *Id.* at 189. Wikipedia's more than 1500 administrators, in turn, enforce those rules. *Id.* at 174. The Arbitration Committee publishes its final decisions. *Id.* at 177.

[250] *Id.* at 173.

[251] *Id.*

[252] *Id.*

[253] Howard, *supra* note 7, at 4D. Facebook similarly urges users to provide the company's team of professional reviewers with "accurate and detailed information" so that "you can help us locate and remove abuse on the site as quickly and efficiently as possible." Jessica Ghastin, *Responding to Abuse Reports More Effectively*, THE FACEBOOK BLOG (Oct. 14, 2009, 10:43 AM), http://blog.facebook. com/blog.php?post=144628037130.

[254] Howard, *supra* note 7, at 4D.

[255] Dan Raywood, *YouTube Safety Mode Introduced to Block Inappropriate Content*, SC MAG. (U.K.) (Feb. 15, 2010), http://www.scmagazineuk.com/youtube-safety-mode-introduced-to-block-inappropriate-content-but-claims-made-that-it-will-only-have-a-minor-impact/article/163784/; *Safety Center: Hateful Content*, *supra* note 239.

[256] *Safety Center: Hateful Content*, *supra* note 239.

artwork using an application called Personas.[257] Mozilla lets community members review users' Persona requests; once approved, the user's artwork is available for others to adopt.[258] Mozilla provides guidelines on artwork's potentially offensive and hateful content to its community members to assist them in their review of applications.[259] Mozilla, however, retains the ability to oversee the community members' decisions, especially when users contest those decisions.[260]

Intermediaries might also empower users in other ways: those who dispute hateful distortions might be provided a space to present their case and discuss it.[261] Google's news service, for example, has taken steps in this direction by permitting the subjects of news articles to reply to stories that include their name.[262] Along similar lines, search engines could offer discounted advertisement rates for counter-speakers targeted by digital hate, who could use that advertising space to directly respond to hate speech generated by a search engine's results. Just as Google itself placed "Offensive Search Results" ads,[263] it could provide discounted rates for other groups to do the same. A group like the NAACP could inexpensively purchase ads providing links to counter-speech about Dr. Martin Luther King in searches of his name to ensure that readers see their link alongside links to the neo-Nazi website *MartinLutherKing.org*.[264] Google could also award free online advertising to targeted groups as it does for certain charitable organizations.[265]

[257] *How to Create Your Own Persona*, MOZILLA, http://www.getpersonas.com/en-US/demo_create (last visited Apr. 8, 2011). Personas is a feature in the Firefox browser that allows a user to select simple-to-use themes, known as Personas, to personalize their browser and status bar. *Personas for Firefox*, WIKIPEDIA, http://en.wikipedia.org/wiki/Personas_for_Firefox (last visited Apr. 8, 2011). Over 220,000 Personas are available for users to choose from on the *GetPersonas.com* website. *Id.*

[258] E-mail from Julie Martin, Assoc. Gen. Counsel, Mozilla, to Danielle Citron, Professor of Law, Univ. of Maryland School of Law (Aug. 11, 2010) (on file with author).

[259] *Id.*

[260] *Id.*

[261] *See* Pasquale, *supra* note 138, at 62.

[262] ALEXANDER HALAVAIS, SEARCH ENGINE SOCIETY 136 (2009). Google News Service provides news to users, a service that is separate from its work as a search engine. With respect to its search engine services, Google has also considered "expos[ing] user reviews and ratings for various merchants alongside their results" to address the problem of high rankings for merchants with "extremely poor user experience." Singhai, *supra* note 204. It ultimately rejected that course of action because it "would not demote poor quality merchants in our results and could still lead users to their websites." *Id.*

[263] *See supra* notes 205-208 and accompanying text.

[264] *See supra* notes 227-228 and accompanying text (discussing racist website *MartinLutherKing.org* aimed at children researching the civil rights leader).

[265] *In Kind Advertising for Non-profit Organizations*, GOOGLE GRANTS, http://www.google.com/grants/ (last visited Apr. 8, 2011) (explaining its "unique in-kind donation program awarding free AdWords advertising to select charitable organizations" that share

3. Architectural Choices

Intermediaries can also help encourage the development of digital citizenship norms through architectural choices.[266] As Jaron Lanier reminds us, the web's anonymity – often extolled as an irreplaceable virtue – was neither an inevitable feature of net design,[267] nor necessarily a salutary one. Indeed, the internet's great communicative strengths – e.g., its ability to aggregate large numbers of speakers as well as disaggregate speakers' offline identities from their online voices – also magnify its capacity to empower certain socially destructive behaviors.[268] Anonymity is thus valuable when it enables speakers to avoid retaliation,[269] but not when it simply enables speakers to avoid responsibility for destructive behavior. For this reason, Lanier urges users: "Don't post anonymously unless you really might be in danger."[270]

Private intermediaries can play an important role in shaping these norms by discouraging anonymity in appropriate circumstances. For example, intermediaries might permit anonymity as a default matter, revoking it when users violate TOS agreements or Community Guidelines.[271] Or they might instead follow Facebook's lead.[272] Facebook requires every user to register under his or her real name and to provide an email address to assist Facebook in verifying his or her identity.[273] On Facebook, "there would be no pseudonymous role-playing, as in so many online social networks."[274] Facebook's philosophy is one of "radical transparency," which its founder

its "philosophy of community service to help the world in areas such as science and technology").

[266] *See* Lessig, *The New Chicago School*, *supra* note 27, at 662-63 (explaining that institutions can shape others' behavior through the development of social norms, as well as through law, markets, and architecture).

[267] JARON LANIER, YOU ARE NOT A GADGET 6 (2010) (as originally introduced, the web "emphasized responsibility, because only the owner of a website was able to make sure that their site was available to be visited").

[268] Citron, *Cyber Civil Rights*, *supra* note 20, at 63-65; *see also* Smith, *supra* note 86, at 59-60 (explaining how unique features of the internet exacerbate its power to spread hate).

[269] *See, e.g.*, Amy J. Schmitz, *"Drive-Thru" Arbitration in the Digital Age: Empowering Consumers Through Binding ODR*, 62 BAYLOR L. REV. 178, 202-04 (2010) (discussing how consumers may be more likely to challenge corporate misbehavior through online vehicles that offer some measure of anonymity and thus protection from retaliation).

[270] LANIER, *supra* note 267, at 21.

[271] Intermediaries might accomplish this strategy by requiring users to register with intermediaries, e.g., requiring credit card information or email address. We thank Julie Cohen for this insightful suggestion.

[272] DAVID KIRKPATRICK, THE FACEBOOK EFFECT: THE INSIDER STORY OF THE COMPANY THAT IS CONNECTING THE WORLD 13 (2010).

[273] Richard A. Posner, *Just Friends*, NEW REPUBLIC, Aug. 12, 2010, at 27 (reviewing KIRKPATRICK, *supra* note 272).

[274] *Id.*

Mark Zuckerberg believes will help make people more tolerant of each other's eccentricities.[275]

Facebook justifies its comparatively hands-off approach to hate speech[276] partly because it does not permit truly anonymous speech.[277] A Facebook employee asked: "Would we rather Holocaust denial was discussed behind closed doors or quietly propagated by anonymous sources? Or would we rather it was discussed in the open on Facebook where people's real names and their photo is associated with it for their friends, peers, and colleagues to see?"[278] Although, as discussed above, we think Facebook and other intermediaries in this context have missed valuable opportunities to engage in counter-speech, we urge more intermediaries to make architectural choices that discourage speakers from refusing to take responsibility themselves for their own hateful expression.

Although focusing on website operators rather than on intermediaries, Nancy Kim has similarly urged architectural designs that default to identified rather than anonymous postings, thus challenging the assumption that all postings should be afforded equal weight.[279] Along these lines, some newspapers and games have moved away from anonymous comments on their online versions.[280] Kim also offers thoughtful suggestions on how website

[275] *Id.*

[276] *See supra* notes 149-151, 220 and accompanying text (explaining that Facebook deems threats of violence to groups as prohibited hate speech worthy of removal and refuses to recognize Holocaust denial as prohibited hate speech).

[277] As Richard Posner explains, Facebook requires every user to register under his real name and to provide an email address to assist Facebook in verifying his identity. Posner, *supra* note 273, at 27.

[278] Chris Matyszczyk, *Facebook: Holocaust Denial Repulsive and Ignorant*, CNET NEWS BLOG (May 6, 2009, 1:04 PM), http://news.cnet.com/8301-17852_3-10234760-71.html.

[279] Kim, *supra* note 18, at 1016-17; *see also id.* at 1017 ("The point is not to make identified postings mandatory, but to make identified postings easier than slightly more burdensome anonymous postings.").

[280] *See, e.g.*, Levy, *supra* note 118 (explaining that because the New York Times, but not the Washington Post, devotes staff time to moderating comments before they appear on their blogs, "comments at the Times tend to be much more thoughtful – and hence worth reading – while comments on the Post's political blogs tend to be much more partisan and much more full of rant"); Roy Greenslade, *Paper Puts Up a Paywall for Comments*, GREENSLADE BLOG (July 13, 2010, 17:23), http://www.guardian.co.uk/media/greenslade/2010/jul/13/paywalls-us-press-publishing; Stephanie Goldberg, *News Sites Reining in Nasty User Comments*, CNN (July 19, 2000), http://articles.cnn.com/2010-07-19/tech/commenting.on.news.sites_1_comments-news-sites-credit-card?_s=PM:TECH (discussing news websites like the Huffington Post that require registration or real names as a condition of commenting); Richard Pérez-Peña, *News Sites Rethink Anonymous Online Comments*, N.Y. TIMES, Apr. 11, 2010, http://www.nytimes.com/2010/04/12/technology/12comments.html (discussing news websites that are considering real names or otherwise regulating comments to their online news content).

sponsors might design their systems to "slow down the posting process, encouraging posters to more carefully consider what they say before they press 'Send'" – for example, by requiring a waiting or cooling-off period before the post is published, during which the poster may choose to edit or remove the message.[281] These are just a few examples: the possibilities for community education, empowerment, and encouragement are substantial, especially as emerging technologies facilitate even more interactivity online. [282]

CONCLUSION

Troubled by the considerable harms posed by digital hate to civic engagement and thus digital citizenship, we nonetheless recognize the considerable legal and political barriers to governmental responses. For this reason here we leverage the interest and commitment to addressing digital hate already expressed by a number of intermediaries to explore promising alternatives, while noting users' potential role in shaping that interest and commitment through consumer demand.

To this end, we suggest that interested intermediaries can valuably advance the fight against digital hate with increased transparency – e.g., by ensuring that their voluntary efforts to define and proscribe hate speech explicitly turn on the harms to be targeted and prevented. We also urge them to consider the range of available responses to hateful speech that include not only removal, but also engaging in or facilitating counter-speech, as well as educating and empowering users with respect to digital citizenship. We remain optimistic that a thoughtful intermediary-based approach to hate speech can significantly contribute to norms of equality and respect within online discourse without sacrificing expression.

[281] Kim, *supra* note 18, at 1017.

[282] KLEIN, *supra* note 66, at 191-92 ("[T]he net generation must equip themselves with the new awareness that many websites *are not* what they appear to be . . . . In addition to asking questions and promoting awareness about the nature of media and information in cyberspace, a socially responsible net generation must acquire a mature understanding about the sinister elements that purvey that world, and where they lead.").

THE EFFECTS OF RACIAL ANIMUS ON A BLACK PRESIDENTIAL CANDIDATE: USING GOOGLE SEARCH DATA TO FIND WHAT SURVEYS MISS*

Seth Stephens-Davidowitz
sstephen@fas.harvard.edu

June 9, 2012

Abstract

How can we know how much racial animus costs black candidates if few voters will admit such socially unacceptable attitudes to surveys? I use a new, non-survey proxy for an area's racial animus: Google search queries that include racially charged language. I compare the proxy to an area's votes for Barack Obama, the 2008 black Democratic presidential candidate, controlling for its votes for John Kerry, the 2004 white Democratic presidential candidate. Previous research using a similar specification but survey proxies for racial attitudes yielded little evidence that racial attitudes affected Obama. Racially charged search, in contrast, is a robust negative predictor of Obama's vote share. My estimates imply that continuing racial animus in the United States cost Obama 3 to 5 percentage points of the national popular vote in 2008, yielding his opponent the equivalent of a home-state advantage country-wide.

JEL Codes: C80/D72/J15/P16.
Keywords: Discrimination, Voting, Google
First Version: November 2011

*I thank Alberto Alesina, David Cutler, Ed Glaeser, and Lawrence Katz for immensely helpful conversations. I also benefited from discussions with James Alt, Joe Altonji, Joshua Angrist, Sam Asher, Susan Athey, Thomas Barrios, Lorenzo Casaburi, Gary Chamberlain, Raj Chetty, Ryan Enos, John Friedman, Roland Fryer, Joshua Gottlieb, Nathaniel Hilger, Jennifer Hochschild, Guido Imbens, Gary King, David Laibson, Jacob Leshno, Brendan Nyhan, Aurélie Ouss, Ben Ranish, Matt Resseger, Andrei Shleifer, Greg Sobolski, and Danny Yagan. Pan-Pan Jiang and Ramon Roullard provided outstanding research assistance. I thank three anonymous referees for comments. All remaining errors are my own.

I Introduction

Does racial animus cost a black candidate a substantial number of votes in contemporary America? The most recent review of the literature is inconclusive: "Despite considerable effort by numerous researchers over several decades, there is still no widely accepted answer as to whether or not prejudice against blacks remains a potent factor within American politics" (Huddy and Feldman, 2009).[1]

There are two major reasons this question has been of such enduring interest to scholars: first, it helps us understand the extent of contemporary prejudice;[2] second, it informs us on the determinants of voting.[3] There is one major reason the question has proven so difficult: individuals' tendency to withhold socially unacceptable attitudes, such as negative feelings towards blacks, from surveys (Tourangeau and Ting, 2007; Berinsky, 1999; Berinsky, 2002; Gilens et al., 1998; Kuklinski et al., 1997).

This paper uses non-survey-based methodology. I use a novel data source to proxy an area's racial animus: Google search queries that include racially charged language. I compare the proxy to an area's change in Democratic vote shares from the 2004 all-white presidential election to the 2008 biracial presidential election. This empirical specification is most similar to that of Mas and Moretti (2009). They use a survey measure of support for a law banning interracial marriage from the General Social Survey (GSS) as their state-level proxy for racial attitudes. They do not find evidence that racial attitudes affected Barack Obama's 2008 vote share.

Google data, evidence suggests, are unlikely to suffer from major social censoring: Google searchers are online and likely alone, both of which make it easier to express socially taboo thoughts (Kreuter et al., 2009). Furthermore, individuals say they are forthcoming with Google (Conti and Sobiesk, 2007). The large number of searches for pornography and sensitive health information adds additional evidence that Google searchers express interests not easily elicited by other means. Relative to measures from the GSS, Google-based mea-

[1] The authors are referring to the effects of racial attitudes on both voting for black candidates and policy opinions. This paper will focus on the former. The data source could potentially also add evidence on the latter.

[2] Charles and Guryan (2011) surveys some of the voluminous literature studying modern discrimination. Creative field environments used to study discrimination include NBA referees (Price and Wolfers, 2010); baseball umpires (Parsons et al., 2011); baseball card sales (List, 2004); motor vehicle searches (Knowles et al., 2001); and employers receiving manipulated resumes (Bertrand and Mullainathan, 2004).

[3] Rational choice theory says that calculation of economic effects of outcomes fully determine voting. A number of scholars have previously found important deviations from an extreme interpretation of this model (Benjamin and Shapiro, 2009; Alesina and Rosenthal, 1995; Berggren et al., 2010; Wolfers, 2002).

sures are also meaningfully available at a finer geographic level, use more recent data, and aggregate information from much larger samples.[4]

The baseline proxy that I use is the percentage of an area's total Google searches from 2004-2007 that included the word "nigger" or "niggers." I choose the most salient word to constrain data-mining.[5] I do not include data after 2007 to avoid capturing reverse causation, with dislike for Obama causing individuals to use racially charged language on Google.[6] My regression analysis includes 196 of 210 media markets, encompassing more than 99 percent of American voters.

The epithet is a common term used on Google. During the period 2004-2007, there were roughly the same number of Google searches that included the word "nigger(s)" as there were Google searches that included words and phrases such as "migraine(s)," "economist," "sweater," "Daily Show," and "Lakers." (Google data are case-insensitive.) The most common searches including the epithet (such as "nigger jokes" and "I hate niggers") return websites with derogatory material about African-Americans. The top hits for the top racially charged searches are nearly all textbook examples of antilocution, a majority group's sharing stereotype-based jokes using coarse language outside a minority group's presence. This was determined as the first and crucial stage of prejudice in Allport's (1979) classic treatise. From 2004-2007, the searches were most popular in West Virginia; upstate New York; rural Illinois; eastern Ohio; southern Mississippi; western Pennsylvania; and southern Oklahoma.

I find that racially charged search is a large and robust negative predictor of Obama's vote share. A one standard deviation increase in an area's racially charged search is associated with a 1.5 percentage point decrease in Obama's vote share, controlling for John Kerry's vote share.[7] The statistical significance and large magnitude are robust to controls for changes

[4] The measure used by Mas and Moretti (2009) has data available for 45 states. Aggregating data since 1990, four states have 20 observations or fewer; 10 states have 50 or fewer; and 19 have 100 or fewer.

[5] Kennedy (2003, p.22) says this is "the best known of the American language's many racial insults ... the paradigmatic slur." Using just one word or phrase, even one that can be used for different reasons, to proxy an underlying attitude builds on the work of scholars who have conducted text analysis of newspapers. For example, Saiz and Simonsohn (2008) argue that news stories about a city that include the word "corruption" can proxy a city's corruption. And Gentzkow et al. (2011) show that, historically, Republican (Democratic) newspapers include significantly more mentions of Republican (Democratic) presidential candidates.

[6] About five percent of searches including "nigger" in 2008 also included the word "Obama," suggesting feelings towards Obama were a factor in racially charged search in 2008. Search volume including both the epithet and "Obama" was not a large factor in 2007. It is also worth noting that search volume for the racial epithet is highly correlated through time, and, though effects would be a bit stronger including 2008 data, any choice of dates will yield roughly similar results. For example, the correlation between 2004-2007 and 2008-present state-level racially charged search is 0.94.

[7] My baseline measures are based on the two-party vote share. The vote share is thus the total votes for the Democratic candidate divided by the total votes for the Democratic and Republican candidates. This is

in unemployment rates; home-state candidate preference; Census division fixed effects; prior trends in presidential voting; changes in Democratic House vote shares; swing state status; and demographic controls. The estimated effect is somewhat larger when adding controls for an area's Google search volume for other terms that are moderately correlated with search volume for "nigger" but are not evidence for racial animus. In particular, I control for searches including other terms for African-Americans ("African American" and "nigga," the alternate spelling used in nearly all rap songs that include the word) and profane language.

The results imply that, relative to the most racially tolerant areas in the United States, prejudice cost Obama between 3.1 percentage points and 5.0 percentage points of the national popular vote. This implies racial animus gave Obama's opponent roughly the equivalent of a home-state advantage country-wide. The cost of racial prejudice was not decisive in the 2008 election. But a four percentage point loss by the winning candidate would have changed the popular vote winner in the majority of post-war presidential elections.[8]

I argue that any votes Obama gained due to his race in the general election were not nearly enough to outweigh the cost of racial animus, meaning race was a large net negative for Obama. Back-of-the-envelope calculations suggest Obama gained at most only about one percentage point of the popular vote from increased African-American support. The effect was limited by African-Americans constituting less than 13 percent of the population and overwhelmingly supporting every Democratic candidate. Evidence from other research, as well as some new analysis in this paper, suggest that few white voters swung in Obama's favor in the *general* election due to his race.[9] A large cost of race in the general election is consistent with some scholars' estimates that, in light of the immensely unpopular incumbent Republican president, Obama substantially underperformed in the 2008 general election (Lewis-Beck et al., 2010; Tesler and Sears, 2010). It also can explain why white male Democratic candidates consistently outperformed Obama in hypothetical general election polls (Jackman and Vavreck, 2011). And it can explain why House Democrats' vote gains from 2004 to 2008 were significantly larger than Obama's gain relative to Kerry.

The main contributions of this paper are threefold: First, I offer new evidence that

a standard measure of changing partisan support (e.g. Gentzkow et al., 2011; DellaVigna and Kaplan, 2007). I obtain similar results using Democratic share of total votes instead.

[8] Racially charged search is not significantly correlated with any measure of 2004 or 2008 swing state status of which I am aware. Thus, the effect on the popular vote would be expected to translate similarly to the electoral college. States consistently defined as swing states in recent elections that rank among the top 15 highest states on racially charged search include Pennsylvania, Michigan, Ohio, and Florida.

[9] The effect of race on the overall probability of being elected president would also have to consider the effects of race on primary voting and on fundraising. These questions are beyond the scope of this paper.

racial attitudes remain a potent factor against African-Americans, nationwide, in modern American politics. This suggests the null result in Mas and Moretti (2009) was due to limitations in the GSS proxies for racial attitudes. The results are larger than those of most studies using individual-level survey data (e.g., Piston, 2010; Schaffner, 2011; Pasek et al., 2010). In addition, my main results rely on administrative, rather than reported, vote data; some scholars argue misreporting is a significant concern with reported vote data (Atkeson, 1999; Wright, 1993; Ansolabehere and Hersh, 2011).

Second, the new data source for area-level proxies of racial attitudes may be useful to other researchers.[10] Researchers studying the causes or consequences of an area's racial attitudes previously have used decades of aggregated GSS data on views towards interracial marriage or similar issues to obtain such proxies (e.g., Alesina et al., 2001; Alesina and La Ferrara, 2002; Charles and Guryan, 2008; Cutler et al., 1999; Card et al., 2008). The Google data add to greatly improved individual-level proxies from list experiments and implicit attitude tests as tools for scholars studying racial attitudes.[11]

The third, and probably most important, contribution is methodological: I show that Google search data can yield new evidence on a question complicated by social desirability bias. This builds on a nascent literature finding promise in Google data. Previous papers using Google search data have tended to focus on its timing advantage. Since Google makes its data available the next day, while many agencies take weeks, Google can yield quicker information on health (Ginsberg et al., 2009; Seifter et al., 2010); demand (Varian and Choi, 2010); and jobs (Askitas and Zimmermann, 2009).[12] In addition, previous work has tended to report correlations between Google data and existing proxies from alternative data sources rather than find new evidence on an empirical question. Scheitle (2011), for example, notes correlations between Google searches on a variety of topics, though not racial animus, and existing measures. This paper shows clearly that Google search query data can do more than correlate with existing proxies; on socially sensitive topics, they can give better

[10] I am not aware of any previous academic paper that uses this data source to proxy racial attitudes.

[11] Researchers are in the process of developing area-level proxies of racial attitudes using implicit association tests. I interpret the results in this paper as the effects of racial animus. An alternative explanation is that this reflects racial attitudes more broadly, with perhaps the Google search proxy correlating with other types of prejudice, such as implicit prejudice. My interpretation is based on: how common the searches are; the clear interpretation of searches as animus; the fact that it is not clear how correlated an area's implicit prejudice and animus are; and some research using individual data that do not find implicit prejudice an important factor when controlling for admitted explicit prejudice (Compare, for example, Piston (2010) to Pasek et al. (2010)). When area-level averages for implicit prejudice are available, this interpretation can be further tested.

[12] Some papers have used the ultra-high-frequency nature of the data to proxy information otherwise unattainable by day, such as investor interest in companies (Drake et al., 2012).

data and open new research on old questions. Researchers might also use Google data to understand the causes and consequences of animus towards other groups.[13] In addition, the Conclusion lists topics of interest across the social sciences, from Bound et al. (2001), on socially sensitive topics in which research has been similarly hampered and may be open to a similar methodology as that of this paper.

The remainder of this paper is organized as follows. Section II discusses the new Google proxy for an area's racial animus. Section III introduces the empirical specification and results on the effects of racial animus in an election with a black candidate. Section IV interprets the magnitude of the effects, comparing them both to other research on the 2008 election and to research on other factors found to influence voting. Section V concludes.

II Google-Search Proxy For an Area's Racial Animus

II.A. Motivation

Before discussing the proxy for racial animus, I motivate using Google data to proxy a socially sensitive attitude. In 2007, nearly 70 percent of Americans had access to the internet at home (CPS, 2007). More than half of searches in 2007 were performed on Google (Burns, 2007). Google searchers are somewhat more likely to be affluent, though large numbers of all demographics use the service (Hopkins, 2008).

Aggregating millions of searches, Google search data consistently correlate strongly with demographics of those one might most expect to perform the searches. The percent of a state's residents believing in God explains 65 percent of the variation in search volume for the word "God." A state's gun ownership rate explains 62 percent of the variation in a state's search volume for "gun." (See Table I). These high signal-to-noise ratios hold despite some searchers typing the words for reasons unrelated to religion or firearms and not all religious individuals or gun owners actually including the term in a Google search.[14] If a certain group is more likely to use a term on Google, aggregating millions of searches will give a good proxy for that group's area-level population.

Furthermore, evidence strongly suggests that Google elicits socially sensitive attitudes. As mentioned in the Introduction, the conditions under which people search – online and likely alone – limit concern of social censoring. Table II documents substantial search volume

[13] I am not aware of any previous academic paper that suggests this data source to proxy animus towards any group.

[14] The 'top search' for "God" is "God of War," a video game. The 'top search' for "gun" is "Smoking Gun," a website that reveals sensational, crime-related documents.

for various terms that academics suspect may be underreported to surveyors, including sexual topics and sensitive health conditions. The strong signal-to-noise ratio in Google data combined with high search volume for socially sensitive attitudes motivate using Google data to proxy racial animus.

II.B. Proxy

The baseline proxy of racial animus is the percentage of an area's searches, from 2004-2007, that included the word "nigger" or its plural.[15] The racial epithet is *not* a fringe, rare search: It is now included in more than 7 million searches annually.[16] Figure I shows terms included in a similar number of searches, from 2004-2007, as the racial epithet.[17] The word "migraine" was included in about 30 percent fewer searches. The word "Lakers" and the phrase "Daily Show" were each included in about five percent more searches than the racial epithet. While these words and phrases were chosen rather arbitrarily as benchmarks, the number of searches including the racial epithet can also be compared to the number of searches including one of the most common terms, "weather." Search volume including the racial epithet, from 2004-2007, was within two orders of magnitude of search volume including "weather."

For this to be a strong proxy of an area's racial prejudice does *not* require that every individual using the term harbors racial animus, nor that every individual harboring racial animus will use this term on Google. The only assumption necessary is racial animus makes one more likely to use the term. Aggregating millions of searches, areas with more such individuals will search the term more often than areas with fewer such individuals. Returns for common searches including the term strongly support this assumption.

About one quarter of the searches including the epithet, from 2004-2007, also included the word "jokes," searches that yield derogatory entertainment based on harsh African-American stereotypes. These same joke sites, with derogatory depictions of African-Americans, are also

[15] The actual measure is slightly different, as Google counts numerous searches over a small period of time including a word as just one search. The singular of the epithet is searched about 3.6 times as often as the plural. The state-level correlation between the singular and plural is 0.96.

[16] These are approximations calculated using AdWords. It combines searches on 'Desktop and Laptops' and 'Mobile devices.'

[17] The percentage of Google searches including the racial epithet was roughly constant from 2004 through late 2008. There were, though, notable spikes in the days following Hurricane Katrina and in early November 2008, particularly on Election Day. The percentage of Google searches including the racial epithet dropped after the 2008 election and has consistently been about 20 percent *lower* during Obama's presidency than prior to his presidency. A nascent literature is examining how Obama's presidency has affected racial attitudes (DellaVigna, 2010; Valentino and Brader, 2011).

among the top returns for a Google search of just the epithet or its plural, representing about 10 percent of total searches that included the epithet.[18] More information on the searches can also be gleaned from the 'top searches,' the most common searches before or after searches including the word (See Table III). Searchers are consistently looking for entertainment featuring derogatory depictions of African-Americans.

All data are obtained using Google Insights. Google Insights reports the percentage of an area's searches including a word, taken from a random sample of total Google searches, divided by a common factor such that the top area has a value of 100. In particular, an area j's measure of racially charged search is approximately equivalent to

$$\text{Racially Charged Search}_{\text{j}} = 100 \cdot \frac{\left[\frac{\text{Google searches including the word ``nigger(s)''}}{\text{Total Google searches}}\right]_{\text{j}}}{\left[\frac{\text{Google Searches including the word ``nigger(s)''}}{\text{Total Google searches}}\right]_{\text{max}}} \tag{1}$$

The data are readily obtained for all states as well as the District of Columbia. (See Table A.1.) Using the 51 state-level observations would yield a large and robust negative relationship between racially charged search and Obama vote shares. However, much of my analysis will instead use data from roughly 200 media markets in the United States.[19]

The media market data are not as readily obtained. Google Insights does not report data for a local area's search volume for a word if its absolute search volume for the word is below an unreported threshold. The threshold is fairly high, such that data for only the most popular search words are available over this time period for a significant number of the 210 media markets in the United States. Search volume data for the epithet, from 2004-2007, is only available for about 40 media markets.[20] Google Insights does allow researchers to obtain search volume for a union of words; the data will be shown if the total searches that include any of the words rises above the threshold. To obtain data for almost all media markets, I do the following: First, I obtain an area's number of searches that include the word "weather," from 2004-2007. This is available for 200 media markets in the United States. Then I obtain an area's number of searches that include either the word "weather" or the word "nigger(s)," from 2004-2007. Since this, by definition, includes a larger number of searches

[18] I do not know the order of sites prior to my beginning this project, in June 2011. The ordering of sites for searches of just the epithet has changed slightly, from June 2011-April 2012 . For example, while joke sites were the second, third, and fourth returns for a search for "niggers" in June 2011, these sites were passed by an Urban Dictionary discussion of the word by April 2012.

[19] Google Insights says that the media market data corresponds to measures of Arbitron. I have confirmed that they actually correspond to designated media markets, as defined by Nielsen. I match other data to the media markets using Gentzkow and Shapiro (2008).

[20] Search volume data for the epithet, from 2004-present, is available for about 100.

than searches that include “weather,” it rises above the threshold for the same 200 media markets in the United States. Subtracting the difference will yield the desired proxy for these 200 media markets: searches that include the word “nigger(s),” from 2004-2007. Some complications arise from sampling, rounding, normalizing, and the small number of searches that include both “weather” and the racial epithet. I discuss these in Appendix B.[21] Racially charged search volume at the media market level is shown in Figure II. The searches were most popular in West Virginia; upstate New York; rural Illinois; eastern Ohio; southern Mississippi; western Pennsylvania; and southern Oklahoma. They were least popular in Laredo, TX – a largely Hispanic media market; Hawaii; parts of California; Utah; and urban Colorado.[22]

II.C. Correlates with Racially Charged Search

II.C.1. Google Compared to GSS

Figure III compares the Google-based proxy to the GSS measure of Mas and Moretti (2009). Since the GSS only includes data for 44 states plus the District of Columbia, the figures and regressions only include 45 observations. The Google measure has a correlation of 0.6 with the measure of Mas and Moretti (2009), support for a law banning interracial marriage from 1990 to 2004.[23] Some of the outliers are likely due to small samples for some states using GSS data. For example, Wyoming ranks as significantly more racially prejudiced using the Mas and Moretti (2009) proxy than the Google proxy. However, only 8 white individuals living in Wyoming were asked this question by the GSS. (Two, or twenty-five percent, said they supported a law banning interracial marriage.)

II.C.2. Demographics and Use by African-Americans

Table IV shows the demographic predictors of racially charged search at the media market level. The demographic factor correlating strongest with racially charged search is the percentage of the population with a bachelor’s degree. A 10 percentage point increase in college

[21] The process requires thousands of downloads. Since Google’s Terms of Service do not allow the use of an application programming interface, I downloaded these by hand.

[22] Utah’s relatively low volume is, in small part, explained by lower search volume for profane words. However, Utah generally scores low on most measures of racial prejudice. For example, Utah residents were 8th least likely to support a law banning interracial marriage (Mas and Moretti, 2009). And, according to the measure of Charles and Guryan (2008), Utah is the 5th least prejudiced state.

[23] The Google measure has a correlation of 0.66 with the measure of Charles and Guryan (2008), average prejudice from 1972 to 2004. I thank the authors for providing their data.

graduates is correlated with almost a one standard deviation decrease in racially charged search.[24] Younger and more Hispanic areas are less likely to search the term.

There is a small positive correlation between racially charged search and percent black. Readers may be concerned that this is due to African-Americans searching the term, limiting the value of the proxy. This is unlikely to be a major factor: the common term used in African-American culture is "nigga(s)," which Google considers a separate search from the term ending in "er." (Rahman, 2011).[25] Table VI shows the top searches for "nigga(s)." In contrast to the top searches for the term ending in "er," the top searches for "nigga(s)" are references to rap songs. Table VI also shows that, even among the five percent of searches that include the epithet ending in "er" and also include the word "lyrics," the 'top searches' are for racially charged country music songs.

The positive correlation between racially charged search and percent black is better explained by racial threat, the theory that the presence of an out-group can threaten an in-group and create racial animosity (Key Jr., 1949; Glaser, 1994; Glaser and Gilens, 1997). Racial threat predicts a quadratic relationship between the percentage of the population that is black and racial animus (Blalock, 1967; Taylor, 1998; Huffman and Cohen, 2004; Enos, 2010). Zero African-Americans means race is not salient and racial animus may not form. Near 100 percent African-American communities have few white people; white individuals with racial animus are unlikely to choose such a community. Figure IV and columns (3) and (4) of Table IV offer support for this theory. Indeed, the preferred fit between racially charged search and percent black is quadratic. The numbers imply that racial animus is highest when African-Americans make up between 20 and 30 percent of the population.[26] Three of the ten media markets with the highest racially charged search – Hattiesburg-Laurel, Biloxi-Gulfport, and Florence-Myrtle Beach – are between 20 and 30 percent black. Therefore, the relationship between racially charged search and percent black is consistent with racially charged search being a good proxy for racial animus.

[24] Scholars have long debated whether the negative correlation between education and expressed racial prejudice is because education decreases prejudice or increases social censoring. These results fit with recent research from list experiments suggesting that the correlation is due to prejudice differing by education level, rather than social desirability bias differing by education level (Heerwig and McCabe, 2009).

[25] Rap songs including the version ending in 'a' are roughly 45 times as common as rap songs including the version ending in 'er.' – Author's calculations based on searches at http://www.rapartists.com/lyrics/. A famous comedy routine by Chris Rock used the word "nigga." One by Dave Chappelle used the term ending in 'er.'

[26] Interestingly, the max point is similar to the max point of segregation by school. Echenique and Fryer (2007) find a quadratic relationship between a school's proportion of black students and segregation, with the maximum segregation occurring when a school is about 25 percent black.

It is also worth noting that there is not a statistically significant correlation between a media market's racially charged search and its support for John Kerry in 2004, a proxy for an area's liberalism.[27] This fact (along with the results in the rest of the paper) offer evidence against some popular wisdom that racial animus is now predominantly a factor among Republicans.

III The Effects of Racial Animus on a Black Presidential Candidate

Section II argues that Google searches that include the word "nigger(s)" – about as common as searches that include terms such as "Daily Show" and "Lakers," with most of them returning derogatory material about African-Americans – give a strong proxy for an area's racial animus. This section uses the proxy to test the effects of racial animus on an election with a black candidate. The section focuses on the significance and robustness of the results. I hold off until Section IV in fully interpreting the magnitude of the effects.

III.A. The Effects of Racial Animus on Black Vote Share

To test the effects of racial animus on a black candidate's vote share, I compare the proxy to the difference between an area's support for Barack Obama in 2008 and John Kerry in 2004. In particular, define Obama_{j} as the percent of total two-party votes received by Obama in 2008 and Kerry_{j} as the percent of total two-party votes received by Kerry in 2004. In other words, Obama_{j} is an area's total votes for Obama divided by its total votes for Obama or John McCain. Kerry_{j} is an area's total votes for Kerry divided by its total votes for Kerry or George W. Bush. Then $(\text{Obama} - \text{Kerry})_{\text{j}}$ is meant to capture an area's relative preference for a black compared to a white candidate.

The idea is that the different races of the Democratic candidates was a major difference between the 2004 and 2008 presidential races. The 2004 and 2008 presidential elections were relatively similar in terms of perceived candidate ideology. In 2004, about 44 percent of Americans viewed John Kerry as liberal or extremely liberal. In 2008, about 43 percent viewed Barack Obama as liberal or extremely liberal.[28] There were slightly larger differences in perceived ideology of the Republican candidates. Roughly 59 percent viewed George W.

[27] There is not a statistically significant correlation between an area's racially charged search and support for John Kerry, controlling for percent black, either.

[28] Calculations on perceived ideology are author's calculations using ANES data.

Bush as conservative or very conservative in 2004; 46 percent viewed John McCain as conservative or very conservative in 2008. Neither Kerry nor Obama came from a Southern state, important as Southern states have been shown to prefer Southern Democratic candidates (Campbell, 1992). One major difference between the 2004 and 2008 elections was the popularity of the incumbent Republican president. George W. Bush had an historically low approval rating at the time of the 2008 election, which we would expect to give a countrywide positive shock to Obama relative to Kerry.[29]

Before adding a full set of controls, I plot the correlation between Racially Charged Search_j and $(\text{Obama} - \text{Kerry})_\text{j}$. Figure V shows the relationship at the media market level.[30]. Likely due to Bush's low 2008 approval rating, Obama does indeed perform better than Kerry country-wide. (See Table V for a set of summary statistics, including Obama and Kerry support.) However, Obama loses votes in media markets with higher racially charged search. The relationship is highly statistically significant ($t = -7.36$), with the Google proxy explaining a substantial percentage of the variation in change in Democratic presidential support ($R^2 = 0.24$).

One non-racial explanation for the correlation between Racially Charged Search_j and $(\text{Obama} - \text{Kerry})_\text{j}$ might be that areas with high racially charged search were trending Republican, from 2004 to 2008, for reasons other than the race of the candidates. Figures VI, VII, and VIII offer graphical evidence against this interpretation.

Panel (a) of Figure VI shows no relationship between states' racially charged search and changes in states' liberalism, as measured by Berry et al. (1998). Figure VI, panel (b), shows a small, and not significant, negative correlation between media markets' racially charged search and change in Democratic support in House races from 2004 to 2008. (In results shown later, I find that racial animus affected turnout, likely explaining the small relationship with

[29] Bush's approval rating from October 17-20, 2008 was the lowest for any president in the history of the NBC News-Wall Street Journal tracking poll (Hart/McInturff, 2012). He was nearly twice as popular in the run-up to the 2004 election as in the run-up to the 2008 election (Gallup, 2012). A candidate is hurt significantly by running on the party of an unpopular incumbent president, and Obama relentlessly linked McCain to Bush (Jacobson, 2009).

[30] There are 210 media markets in the United States. Ten of the smallest media markets do not have large enough search volume for "weather" and thus are not included. Two additional small media markets (Juneau and Twin Falls) search "weather" much more frequently than other media markets. Since they often score 100 on both "weather" and "weather" or the racial epithet, I cannot pick up their racial animus from the algorithm. Alaska changed its vote reporting boundaries from 2004 to 2008. I was unable to match the media market data with the boundaries for Alaskan media markets. I do not include data from Alaska. Overall, the 196 media markets included represent 99.3 percent of voters in the 2004 election. All of the summary statistics in Table V are virtually identical to summary statistics over the entire population of the United States.

House voting.)

As further evidence that changing preferences for a Democratic candidate are not driving the results, I use data from SurveyUSA, first used by Donovan (2010). In February 2008, SurveyUSA asked voters in 50 states the candidate they would support in two then-hypothetical match-ups: Obama versus McCain and Hillary Clinton versus McCain. If the correlation between racially charged search and changes in Democratic presidential support from 2004 to 2008 were due to areas with higher racially charged search generally losing preference for Democrats, areas with high racially charged search should punish Clinton as much as Obama. Figure VII compares the difference, among white voters, of Obama and Clinton support in the hypothetical McCain match-ups to a state's racially charged search. Obama loses substantial support relative to Clinton in states with higher racially charged search ($t = -9.05$; $R^2 = 0.49$).

Figure VIII uses 2004 and 2008 exit poll data to examine relative preferences for Obama and Kerry, by race. If states with high racially charged search were more likely to support Obama, independent of whites' racial attitudes, the effect would likely show up for both black and white voters in these states. Instead, there is no relationship between racially charged search and change in black support for Obama relative to Kerry; the relationship is driven entirely by white voters.[31]

Reported voting data are never ideal. However, the results in Figures VII and VIII – combined with those in Figure VI – strongly suggest that decreased support for Obama in areas with high racially charged search is caused by white voters supporting Obama less than they would a white Democrat.

I now return to administrative vote data at the media market level and examine the relationship more systematically using econometric analysis. I add a number of controls for other potential factors influencing voting. I find no evidence for an omitted variable driving the negative correlation between a media market's racially charged search and its preference for Obama compared to Kerry. The empirical specification is

$$(\mathrm{Obama} - \mathrm{Kerry})_{\mathrm{j}} = \beta_0 + \beta_1 \cdot \mathrm{Racially\ Charged\ Search}_{\mathrm{j}} + \mathrm{X}_{\mathrm{j}}\phi^1 + \mu_{\mathrm{j}} \tag{2}$$

where X_j are area-level controls that might otherwise influence change in support for the Democratic presidential candidate from 2004 to 2008; β_0 is a country-wide shock to Democratic popularity in 2008; and μ_j is noise.

[31] Throughout this paper I refer to non-blacks, including Hispanics and Asians, rather imprecisely, as "whites."

$\text{Racially Charged Search}_j$ is as described in Equation 1, normalized to its z-score.[32] Thus, the coefficient β_1 measures the effect of a one standard deviation increase in $\text{Racially Charged Search}_j$ on Obama's vote share. All regressions predicting voting behavior, unless otherwise noted, are weighted by 2004 total two-party votes. Unweighted regressions and regressions using alternative weighting schemes are run as robustness checks.

The results are shown in Table VII. All columns include two controls known to consistently influence Presidential vote choice (Campbell, 1992). I include Home State_j, a variable that takes the value 1 for states Illinois and Texas; -1 for states Massachusetts and Arizona; 0 otherwise.[33] I also include proxies for economic performance in the run-up to both the 2004 and 2008 elections: the unemployment rates in 2003, 2004, 2007, and 2008.[34]

Column (1), including just the standard set of controls, shows that a one standard deviation increase in a media market's racially charged search is associated with 1.5 percentage points fewer Obama votes. Column (2) adds controls for nine Census divisions. Any omitted variable is likely to be correlated with Census division. Thus, if omitted variable bias were driving the results, the coefficient should drop substantially upon adding these controls. The coefficient, instead, remains the same. Column (3) adds a set of demographic controls: percent Hispanic; black; with Bachelor's degree; aged 18-34; 65 or older; veteran; and gun magazine subscriber; as well as changes in percent black and percent Hispanic. Since there is some measurement error in the Google-based proxy of racial animus, one would expect the coefficient to move towards zero as these controls are added. It does. However, the change is not particularly large (less than a 10 percent decline in magnitude) considering the number of controls. The stability of the coefficient to a rich set of observable variables offers strong evidence for a causal interpretation (Altonji et al., 2005).

III.A.1. Adding Google Controls to Reduce Measurement Error

There is not a one-to-one correspondence between an individual's propensity to type the racial epithet into Google and his or her racial animus. Individuals may type the epithet for a variety of reasons other than animus.[35] Individuals harboring racial animus may express

[32] If I use ln(Racially Charged Search) as the proxy for racial animus, I find larger estimates for the effect of racial animus.

[33] Since I run the regressions at the media market level and some media markets overlap states, I aggregate Home State_j from the county level, weighting by 2004 turnout. For the Chicago media market, as an example, Home State = 0.92, as some counties in the media market are in Indiana.

[34] Since Campbell (1992) and similar papers forecast the election at the state-level, they use growth in GDP. This is not available at the county or media market level.

[35] One large spike in racially charged search occurred when the epithet was used repeatedly by *Seinfeld* actor Michael Richards in a comedy club. Importantly, non-animus uses of the word made at the same

it in different ways – either on different search engines or offline.

Any motivations of searches of the word unrelated to animus that do not differ at the area level will not create any bias in the area-level proxy. However, alternative motivations that differ at the area level will lead to measurement error in the area-level proxy. Classical area-level measurement error will cause attenuation bias in the estimates in Columns (1)-(3) of Table VII. In Columns (4)-(6), I reproduce the results from Columns (1)-(3) but add controls for an area's search volume for other words correlated with the search term unlikely to express racial animus. This should reduce measurement error in the proxy.

Row (8) of Table III shows that some searchers are looking for information on the word.[36] I add a control for "African American(s)" search volume to proxy an area's interest in information related to African-Americans. Since a small percentage of searches for the word ending in "er" are looking for particular cultural references, I add a control for "nigga(s)" search volume. Finally, as some areas may be more prone to use profane language on Google, I add a control for an area's search volume for profane language.[37] Columns (4)-(6) show that the coefficient is more negative in each specification when adding the Google controls.[38]

III.A.2. Robustness Checks and Placebo Tests

Table VIII presents a number of robustness checks. The coefficient is more negative with unweighted regressions or with alternative weighting schemes: 2008 turnout and voting age population. Obama received about 20 percentage points more of the two-party vote share in Hawaii than Kerry did. Obama was born in Hawaii. Excluding Hawaii, though, changes the coefficient towards zero by less than 5 percent. Including trends in presidential voting and Census division effects does not meaningfully change the coefficient. The coefficient is of a similar magnitude including changes in House Democratic support from 2004 to 2008 and swing state status.[39]

The results are also of a similar magnitude defining Democratic presidential support in 2004 and 2008 as the Democratic candidate's share of all votes rather than the share of two-party votes. They are of a similar magnitude instead using $\mathrm{Obama}_\mathrm{j}$ as the dependent variable

frequency across area do not create noise in the area-level proxy.

[36] About 2 percent of searches that include the singular also include "word." Fewer than 1 percent that include the epithet also include "definition" or "define."

[37] Following my general strategy of selecting the most salient word if possible, I use the word "fuck."

[38] The word most correlated with racially charged search, of the three chosen, is the profane word. Including this word has the biggest effect on the coefficient, suggesting an area's tendency to use profanity on Google is the biggest source of measurement error in the baseline proxy for an area's racial animus.

[39] I do not include these controls in the main specifications as they could be affected by Obama support and thus not exogenous.

and including Kerry_j as an independent variable. And they are of a similar magnitude using Obama_j as the dependent variable and including a 4th-order polynomial for Kerry_j as independent variables. Including this polynomial allows for liberal areas to differ from conservative areas in their relative support for Obama and Kerry. The fact that the coefficient on racially charged search is unchanged (perhaps not surprising since racially charged search is not significantly correlated with liberalness and voters perceived the candidates as having similar ideologies) offers additional evidence that racial attitudes, not ideology, explains the results.

Figure IX shows the results of 13 placebo tests. In particular, it examines whether racially charged search predicts changed support for Democratic candidates in other post-war presidential elections.[40] If racially charged search were correlated with positions or characteristics other than race that might systematically differ among candidates, we would expect significant relationships in many of these elections, not just the 2008 election. It is important, in this exercise, to control for Census division effects. It is well-known that Southern states give greater support to Southern Democrats than non-Southern Democrats (Campbell, 1992). And Southern states do have, on average, higher racial animus. While neither the 2004 nor 2008 election included a Southern Democratic candidate, many post-war elections did. Including Census division fixed effects, only one of the other 13 elections (the change from 1988 to 1992) is significant at the 10 % level. The 2008 election is the only one significant at the 1 % level and stands out as a large outlier.

III.B. The Effects of Racial Animus on Turnout in Biracial Election

The robust cost of racial animus on Obama's vote share are the main results of the paper. I can also use the area-level proxy for racial animus to test the effects of racial attitues on turnout and thus better understand the mechanism by which racial prejudice hurt Obama.

The effect is theoretically ambiguous. The effect of racial prejudice on Obama's vote share could be driven by any of three reasons, each with different implications for turnout: Individuals who would have voted for a white Democrat instead stayed home (decreasing turnout); individuals who would have voted for a white Democrat instead voted for McCain (not affecting turnout); individuals who would have stayed home instead voted for McCain (increasing turnout).

I first use the area-level proxy of racial animus to test the average effect of prejudice on

[40] I thank an anonymous referee for suggesting this exercise.

turnout. I regress

$$\Delta\ln(\mathrm{Turnout})_{\mathrm{j}} = \delta_0 + \delta_1 \cdot \mathrm{Racially\ Charged\ Search}_{\mathrm{j}} + \mathrm{Z}_{\mathrm{j}}\phi^2 + \psi_{\mathrm{j}} \tag{3}$$

where $\Delta\ln(\mathrm{Turnout})_{\mathrm{j}}$ is the change in the natural log of the total Democratic and Republican votes from 2004 to 2008; Z_j is a set of controls for other factors that might have changed turnout and Racially Charged $\mathrm{Search}_{\mathrm{j}}$ is as described in Equation 1, normalized to its z-score.

The results are shown in Columns (1) through (3) of Table X. In all specifications, I include percent black and change in the natural log of an area's population from 2000 to 2010. Column (2) adds Census fixed effects. Column (3) adds the same demographic controls used in the vote share regressions in Table VII. In none of the specifications is there a significant relationship between racially charged search and turnout. I can always reject that a one standard deviation increase in racially charged search – which lowers Obama's vote share by 1.5 to 2 percentage points – changes turnout by 1 percent in either direction.[41]

The null effect of racial attitudes on turnout is consistent with animus not affecting any individuals' decision to turnout (but convincing many who would have supported a Democrat to instead vote for McCain). It is also consistent with racial prejudice driving an equal number of individuals who would have voted for a white Democrat to stay home as it convinced individuals who would have stayed home to vote for McCain.

To better distinguish these two stories, I add to the independent variables in Columns (1) to (3) of Table X the interaction between an area's percent Democrats and racially charged search. If racial attitudes affect some individuals' decisions of whether or not to vote, I expect the following: it should increase turnout when there are few Democrats in an area. (There are few Democrats available to stay home due to racial prejudice.) The effect of racial prejudice on turnout should be decreasing as the percentage of the population that supports Democrats increases.

More formally, the regression is:

[41] Washington (2006) finds a 2-3 percentage point increase in turnout in biracial Senate, House, and Gubernatorial elections. Perhaps these results can be reconciled as follows: Obama won a close primary. An average black candidate would be expected to have won his or her primary by a bigger margin. We would thus expect that the average black candidate would have faced lower racial animus in his or her primary than Obama did in a country-wide Democratic primary. Thus, relatively few voters would stay home in the general election rather than support the black candidate in the average election in Washington's (2006) sample.

$$\begin{aligned}\Delta\ln(\mathrm{Turnout})_{\mathrm{j}} = \alpha_0 + \alpha_1 \cdot \mathrm{Kerry}_{\mathrm{j}} + \alpha_2 \cdot \text{Racially Charged Search}_{\mathrm{j}} + \\ \alpha_3 \cdot \text{Racially Charged Search}_{\mathrm{j}} \times \mathrm{Kerry}_{\mathrm{j}} + \mathrm{Z}_{\mathrm{j}}\phi^3 + \epsilon_{\mathrm{j}}\end{aligned} \tag{4}$$

where $\mathrm{Kerry}_{\mathrm{j}}$ is a proxy for an area's percent Democrats.

If racial animus affected Obama vote shares, in part, through changes in turnout, I expect $\alpha_2 > 0$ and $\alpha_3 < 0$.

The coefficients on α_2 and α_3 are shown in Columns (4)-(6) of Table X. In all three specifications, corresponding to the same specifications in Columns (1)-(3), $\alpha_2 > 0$ and $\alpha_3 < 0$. In areas that supported Kerry in 2004, an increase in racial animus decreased 2008 turnout. In areas that supported Bush in 2004, an increase in racial animus increased 2008 turnout. The coefficients tend to be marginally significant, and the standard errors are always too large to say anything precise.

In sum, the evidence on the effects of racial animus on turnout is as follows: Some Democrats stayed home rather than vote for Obama due to his race; a similar number of individuals who would not have otherwise voted turned out for McCain due to Obama's race. However, there is not enough statistical power to determine this number.[42]

IV Interpretation

Section III compares Google racially charged search to changing voting patterns from the 2004 all-white presidential election to the 2008 biracial presidential election and finds that racial animus played a significant role in the 2008 election. Section III.A. finds that racially charged search is a robust negative predictor of Obama's vote share. Section III.B. finds that higher racially charged search predicts increased turnout in Republican parts of the country; decreased turnout in Democratic parts of the country; and, on average, no change in turnout. This section aims to give some intuition to the magnitude of the effects of racial attitudes on presidential voting. I consider three effects, each slightly different: the total votes lost from racial animus; the net effect of race on presidential vote totals; and the percent of whites motivated against voting for a black candidate. I compare the estimated magnitudes both to other factors academics have shown to influence voting and other scholars' estimates of the effects of racial prejudice.

[42] Mas and Moretti (2009) also find that racial prejudice was not, on average, correlated with changes in turnout, as part of their findings that racial prejudice was not a factor in the 2008 election. However, they do not examine whether the correlation differed among Democratic and Republican areas.

IV.A. Lost Votes from Racial Animus

How many additional votes would Obama have received if the whole country had the racial attitudes of the most tolerant areas? Media markets' mean racially charged search is 2.34 standard deviations higher than the minimum racially charged search. Using the least negative number of Table VII, from the specification including standard and demographic controls, and assuming that there is no racial animus in the media market scoring the lowest gives a point estimate for the country-wide effect of 3.1 percentage points.[43] Using the most negative number of Table VII, the specification including standard and Google controls, gives a point estimate for the country-wide effect of 5.0 percentage points.[44]

The effects of racial animus on a black compared to a white Democratic candidate can be compared to voters' well-established comparative preference for a home state compared to a non-home-state candidate. Studies show, on average, voters will reward a candidate from their own home-state with about four percentage points of the two-party vote (Lewis-Beck and Rice, 1983; Mixon and Tyrone, 2004). This is roughly consistent with the home-state advantage found in the regressions in Table VII. Racial animus gave Obama's opponent the equivalent of a home-state advantage country-wide.

While racial animus obviously did not cost Mr. Obama the 2008 election, examining more elections shows that effects of the magnitude found are often decisive. Figure X shows that a two percentage point vote loss would have switched the popular vote winner in 30 percent of post-War presidential elections. A four percentage point loss would have changed more than 50 percent of such elections.

The effect of racial prejudice found by the methodology of this paper can also be compared to estimates obtained using different data sources and methodology. I find that the effects using Google data are larger than effects found using other methodologies. The specification used in this paper is slightly different from the one used in Mas and Moretti (2009). Mas and Moretti (2009) predict a county's Democratic vote share in 2004 and 2008 House and presidential elections from a set of dummy variables (Year=2008; Election Type=presidential;

[43] All these estimates would be higher, of course, if Obama lost some votes from racial animus in the area with the lowest racially charged search.

[44] The Google controls are presumably reducing measurement error. Multiplying the coefficient by 2.34 yields an approximation of the true effect. This would be biased upwards if measurement error substantially lowered the measured minimum racial animus. I do not find this is the case. I calculate a new measure of racial animus as the difference in racially charged search relative to predictions from all the controls in Column (4) of Table VII. This still leaves Loredo, TX as having the minimum value. Regressing $(\text{Obama} - \text{Kerry})_j$ on this measure of racial animus and multiplying the coefficient on the regression by the difference between the mean and the minimum of the measure yields roughly the same result.

Election Type=presidential & Year=2008) and an interaction between a state's GSS racial attitudes and the dummy variables. This specification makes it difficult to pick up the effects of racial attitudes on voting for Obama since House elections are high-variance (sometimes, one of the two major parties does not field a candidate, dramatically shifting the Democratic proportion of vote share). A large swing in House voting can falsely suggest a large trend in Democratic voting.[45]

Nonetheless, I do confirm that effects using the GSS measures and the specification of this paper are not robust. Table IX compares the estimates obtained using the Google measure and the specification of this paper to estimates using GSS measures and the specification of this paper. Using either the measure from Mas and Moretti (2009) or Charles and Guryan (2008) yields smaller estimates of the country-wide effect. (The graphical analysis using GSS measures is shown in Figure XI). In addition, the effect picked up using the GSS data is largely due to a few Southern states which measure high on racial prejudice and also voted for Obama significantly less than they voted for Kerry. In contrast to regressions using the Google measure, where the effect is robust to including Census division fixed effects, regressions using the GSS measures lose significance when including the Census division fixed effects.[46] Furthermore, the preferred fit with the GSS measures (as seen in Figure XI) is quadratic. The fit suggests no effect in just about all parts of the country but an effect in a few southern states. The GSS is ineffective at capturing racial prejudice in all but a few Southern states.

[45] For example, in Mas and Moretti's (2009) Figure 4, the authors compare the difference between the change in Obama and Kerry's vote shares and the change in House voting to their measure of racial prejudice. The difficulty with this comparison is that House elections in which one party does not field a candidate will create enormous noise in the voting metric, swamping any other changes. In 2004 in Vermont, Bernie Sanders won as a highly popular left-wing independent. In 2008 in Vermont, Democrat Peter Welch won with no Republican challenger. Thus, there was a huge gain in Vermont Democratic House support from 2004 to 2008. And the difference between the change in Democratic presidential support and change in Democratic House support, from 2004 to 2008 in Vermont, is -70 percentage points. Adding this kind of noise to the Obama and Kerry difference, and having only 45 state-level GSS observations, it is unlikely that, even if the GSS measure of racial attitudes did predict opposition to Obama, this methodology could pick it up.

[46] Highton (2011) located an alternative data source for racial attitudes from The Pew Research Center Values Study. Pew has asked for 20 years individuals whether they approve of blacks dating whites. Aggregating 20 years of data among whites, Highton (2011) constructs a measure available for 51 states and tests the effects of racial animus on voting in the Obama election. While standard errors are still large and the point estimate is always smaller than using Google data, the Pew data source does lead to more robust estimates than the GSS data source, in part due to the six additional observations. This data source is definitely worth considering in the future. However, there still remain numerous advantages to the Google data. In other work using state-level data, Silver (2009) finds that Obama underperformance at the state level is correlated with voters who admitted race was a factor in their vote.

The final row of Table IX includes the estimates from Piston (2010), Schaffner (2011), and Pasek et al. (2010).[47] Each uses individual data and obtains a smaller preferred point estimate. This suggests individual surveys underestimate the true effect of racial attitudes. There are additional advantages to the empirical specification of this paper relative to the empirical specifications using individual data, in testing for causality, besides just the likely improved measure of racial animus. First, area-level data allow the use of administrative voting data; vote misreporting may be a substantial issue with survey data (Atkeson, 1999; Wright, 1993; Ansolabehere and Hersh, 2011).[48] Second, limiting the data to 2004-2007 limits issues of reverse causation. This is not possible with surveys conducted near election time; racial prejudice may be activated by a negative response to Obama. Third, the controls for ideology (actual voting in previous elections, such as the 2004 presidential election) may be better than the controls – reported ideology – necessary using individual, non-panel survey data.[49]

IV.B. Net Effect of Race

I find that, relative to the attitudes of the most racially tolerant area, racial animus cost Obama between 3 to 5 percentage points of the national popular vote. Obama, though, also gained some votes due to his race. Was this factor comparatively large?

A ballpark estimate from increased support from African-Americans can be obtained from exit poll data. In 2004, 60.0 percent of African-Americans reported turning out, 89.0 percent of whom reported voting for John Kerry. In 2008, 64.7 percent of African-Americans reported turning out, 96.2 percent of whom reported supporting Barack Obama. Assuming these estimates are correct and, with a white Democrat, black support would have been the

[47] A recent paper by Kam and Kinder (2012) finds that ethnocentrism was a factor against Obama. Tesler and Sears (2010) also finds an important role of anti-Muslim sentiment in evaluating Obama. Using Google data (such as searches for "Obama Muslim" or "Obama birth certificate") to further investigate this phenomenon is a promising area for future research.

[48] It is not clear the direction of the bias. One might expect individuals hiding racial prejudice would also feel compelled to claim they were voting for Obama. Alternatively, one might think that individuals scoring high on racial prejudice would incorrectly claim they were voting for Obama nonetheless. Malter (2010) finds unusually large deviations between final state polling data and state votes in 2008 but no difference on average. In results not reported, I find that Obama over(under) performance relative to final polling data is negatively (positively) correlated with racially charged search; however, standard errors are large, and the result is not statistically significant at even the 10 percent level.

[49] An earlier version of this paper models these issues and the relative advantages of the different approaches. Many of the papers using individual data compare the results obtained from the 2008 election to results obtained using the same data source for different elections. Since none of the major individual election surveys are panel data, the exercise does require the surveyed population not changing in a way correlated with the coefficient of interest.

same as in 2004, increased African-American support added about 1.2 percentage points to Obama's national popular vote total.[50] The pro-black effect was limited by African-Americans constituting only 12.6 percent of Americans and overwhelmingly supporting any Democratic candidate.

A variety of evidence suggest that few white voters swung, in the general election, for Obama due to his race. Only one percent of whites said that race made them much more likely to support Obama (Fretland, 2008). In exit polls, 3.4 percent of whites did report both voting for Obama and that race was an important factor in their decision. But the overwhelming majority of these voters were liberal, repeat voters likely to have voted for a comparable white Democratic presidential candidate.[51] Furthermore, Piston (2010) finds no statistically significant relationship, among white voters, between pro-black sentiment and Obama support, when controlling for ideology. Although social scientists strongly suspect that individuals may underreport racial animus, there is little reason to suspect underreporting of pro-black sentiment. Finally, in unreported results, I add an area's normalized search volume for "civil rights" to the regressions in Table VII. The coefficient on Racially Charged Search is never meaningfully changed, and the coefficient on Civil Rights is never statistically significant.

Overall, then, I estimate race cost Obama about two to four percentage points of votes. A large vote loss for Obama due to his race is consistent with research from some other scholars suggesting Obama underperformed in 2008. Jackman and Vavreck (2011), using polling data with hypothetical 2008 match-ups, find an "average" white Democrat would have received about 3 percentage points more votes than Obama did.[52] Lewis-Beck et al. (2010) estimate Obama underperformed by 5 percentage points based on economic fundamentals. Furthermore, Table V shows that House Democratic candidates received a 2.3 percentage

[50] Assume 65 percent of whites turned out in 2008 and 47.6 percent of white voters supported Obama. If African-Americans had voted as they did in 2004, Obama would have instead received $\frac{0.126\times0.6\times0.89+0.874\times0.65\times0.476}{0.126\times0.65+0.874\times0.65} = 52.5$ percent of the two-party vote. This is likely an upper-bound, as any Democrat likely would have seen some improvement in black support due to Bush's high disapproval rating among African-Americans.

[51] Among the 3.4 percent, 87 percent both reported voting for the Democratic candidate in the House race and disapproving of Bush. Among this subset, only 25 percent reported voting for the first time. And, among such first-time voters, 60 percent were 18-24, possibly ineligible to vote in any prior elections.

[52] The authors estimate that John Edwards would have received 3.0 percentage points more votes than Obama in an election against McCain. Their model suggests an "average" Democrat would have received about 3.3 percentage points more votes. Interestingly, the authors find that Hillary Clinton performs similarly to Obama. This fits with the SurveyUSA data used in Figure VII, in which Obama and Clinton receive similar national vote shares. Whether animus towards women help explain Clinton's shortfall relative to white male Democratic candidates is another area of research in which Google data may prove useful.

point larger gain in 2008 relative to 2004 than Obama received relative to Kerry; the results in Section III.B. suggest the House Democratic swing would have been even larger absent turnout effects due to Obama's race.

IV.C. White Voters Swung by Racial Animus

As another way of giving intuition for the magnitude of the effect, I combine the vote share results in Section III.A. with the turnout results in Section III.B.. I can then estimate the percent of white voters who would have voted for a white Democrat in 2008 but did not support a black one. This number is higher than the 3 to 5 percentage point number for two reasons: First, any white American who would have voted for a Republican could not have been moved by prejudice against Obama. Second, the negative effect was presumably limited to whites. African-Americans, while only 13 percent of the population, are disproportionately represented among Democrats.

The percent motivated by animus is the number of votes lost due to animus divided by the total number of whites who would have supported a Democrat absent prejudice. Section III.B. finds that turnout was unaffected, on average, by prejudice. Thus, the denominator (the percent of whites who would have supported a Democrat, absent prejudice) is the number of whites who supported Obama plus the number of votes lost due to prejudice.[53] Exit polls suggest 41.7 percent of 2008 voters were white Obama supporters. The percent motivated by animus is estimated between $\frac{3}{44.7} = 6.7$ and $\frac{5}{46.7} = 10.7$ percent. Between 6.7 and 10.7 percent of white Democrats did not support Obama because he was black.

How do these numbers compare to what whites tell surveys? Among whites who told the GSS in 2008 and 2010 that they voted for Kerry in 2004, 2.6 percent said they would not vote for a black president.[54] Three percent of whites told Gallup Obama's race made them much less likely to support him (Fretland, 2008). Approximately 4.8 percent of whites told exit pollsters they voted for McCain and race was an important factor in their vote.[55] Evidence strongly suggests that many whites voted against Obama due to his race but did

[53] If turnout actually increased from racial prejudice, in contrast to the results of Section III.B., fewer white Democrats would have been motivated by racial prejudice against Obama than suggested by this section. If turnout actually decreased from racial prejudice, in contrast to the results of Section III.B., more white Democrats would have been motivated by racial prejudice against Obama than suggested by this section.

[54] This is a stronger condition than voting against a particular black candidate due to his race. Among Bush supporters, the number was 4.3 percent.

[55] These numbers are perhaps better compared to lost votes from racial animus divided by percentage of whites in the electorate. As with whites reporting Obama support due to race, it seems unlikely that a substantial percentage of these voters actually decided their vote on race. Nearly 90 percent also reported voting for the Republican House candidate. Nearly 90 percent said they had also voted in previous elections.

not admit that to surveys.

For additional intuition on the size of the effect, these numbers can be compared to persuasion rates as calculated by media scholars. The field experiment of Gerber et al. (2009) implies that 20 percent of individuals reading *The Washington Post* were persuaded to vote for a Democrat. Gentzkow et al. (2011) find that, historically, partisan newspapers persuade fewer than 3.4 percent of readers. The results of DellaVigna and Kaplan (2007) imply that Fox News persuades 11.6 percent of viewers to vote Republican. Thus, the proportion of white Democrats who will not vote for a black Democratic Presidential candidate is roughly equivalent to the proportion of Democrats who can be persuaded by Fox News to not vote for a white Democratic Presidential candidate.

V Conclusion

Whether many white Americans will not vote for a black presidential candidate is perhaps the most renowned problem complicated by social desirability bias. Scholars have long doubted the accuracy of survey results on this sensitive question. Google search query data, this paper shows, offer clear evidence that continuing racial animus in the United States costs a black candidate substantial votes.

There are important additional questions on sensitive topics that may similarly yield to Google-based methodology. In a study of measurement error in surveys, Bound et al. (2001) include the following sensitive behaviors as difficult to measure for surveyors due to social censoring: “the use of pugnacious terms with respect to racial or ethnic groups”; voting; use of illicit drugs; sexual practices; income; and embarrassing health conditions. Words related to all these topics are searched often on Google. Scholars conducting research in these sensitive areas should consider the data source as containing otherwise inaccessible information on the human psyche.

Author Affilications: Harvard University.

A Racially Charged Search, State

Table A.1
Racially Charged Search, State

Rank	*State*	*Racially Charged Search*
1.	West Virginia	100
2.	Louisiana	86
3.	Pennsylvania	85
4.	Mississippi	83
5.	Kentucky	82
6.	Michigan	78
7.	Ohio	78
8.	South Carolina	76
9.	Alabama	76
10.	New Jersey	74
11.	Tennessee	73
12.	Florida	71
13.	New York	71
14.	Rhode Island	70
15.	Arkansas	70
16.	North Carolina	69
17.	Georgia	69
18.	Connecticut	68
19.	Missouri	68
20.	Nevada	67
21.	Illinois	65
22.	Delaware	65
23.	Oklahoma	65
24.	Maryland	64
25.	Indiana	63
26.	Wisconsin	63
27.	Kansas	62
28.	Texas	62
29.	Virginia	59
30.	Vermont	59
31.	California	57
32.	Maine	56
33.	Nebraska	55
34.	New Hampshire	54
35.	North Dakota	54
36.	Iowa	53
37.	Massachusetts	52
38.	Arizona	51
39.	Washington	50
40.	South Dakota	50
41.	Alaska	50
42.	Wyoming	48
43.	Montana	48
44.	Oregon	47
45.	Minnesota	46
46.	District of Columbia	44
47.	Idaho	39
48.	New Mexico	39
49.	Colorado	39
50.	Hawaii	34
51.	Utah	30

Notes: *Racially Charged Search* is Web Search, from January 2004-December 2007, for either “nigger” or “niggers.” This data can be found here: `http://www.google.com/insights/search/#q=nigger%2Bniggers%2C%2C%2C%2C&geo=US&date=1%2F2004%2048m&cmpt=q`.

B Algorithm to Determine Search Volume at Media Market Level

Take a sample s in Google.
Let X be a set of possible searches. Denote $X_{j,s}$ as the value that Google Insights gives. This is $X_{j,s} = x_{j,s}/x_{max,s}$ where $x_{j,s}$ is the fraction of Google searches in area j in sample s that are in X. (See Equation 1).
Take two words N and W. And let $C = N \cup W$ and $B = N \cap W$. Then $n_{j,s} = c_{j,s} - w_{j,s} + b_{j,s}$. Denoting x_j as the expected value of x in area j, then $n_j = c_j - w_j + b_j$ Assume we have an area for which, for $x \in \{c, w, n, b\}$, $x_{j,s}$ is independent of $x_{max,s}$. Then $X_j = x_j/x_{max}$. Then

$$N_j = \frac{c_{max}}{n_{max}} C_j - \frac{w_{max}}{n_{max}} W_j + \frac{b_{max}}{n_{max}} B_j \tag{5}$$

Assume B_j is negligible, a reasonable assumption for words used in this paper. The issue is that N_j, the word of interest, is only reported for about 30 media markets, whereas C_j and W_j are reported for about 200 media markets. Since N_j depends linearly on W_j and C_j I can find $\frac{c_{max}}{n_{max}}$ and $\frac{w_{max}}{n_{max}}$ using data for any media market that reports all 3 values. I can then use these numbers to find N_j for all 200 that report W_j and C_j. If C_j, W_j, and N_j were reported with no error for media markets, I could find exact numbers. Even with 5,000 downloads, I do not get perfect estimates of C_j, W_j, and N_j. I thus back out the coefficients by regressing the averages for 30 media markets that have all data available. The R^2 on this regression is 0.86, meaning there is minor remaining error. After 5,000 downloads, regressing halves of the samples suggest this strategy has captured about 80 percent of the variation in the actual number. To deal with the minor remaining error, I use the first half sample estimate as an instrument for the second half sample when racially charged search is an independent variable in regressions.

Algorithm in Practice:
1. Download 5,000 samples for "weather," from 2004-2007.
2. Download 5,000 samples for "nigger+niggers," from 2004-2007. (A " + " signifies an "or.")
3. Download 5,000 samples for "nigger+niggers+weather," from 2004-2007.
4. Eliminate any media market that ever scores 0 or 100 for "weather." (A 0 means absolute search volume is too small. A 100 means it scores the maximum.)
(12 of the smallest media markets in the country are eliminated, 10 that never show up and 2 that compete for the top "weather" search spot.)
5. Calculate a media market's average score for "weather," "nigger+niggers," and "nigger+niggers+weather."
6. Regress "nigger+niggers" average score on "weather" average score and "weather+nigger+niggers" average score for the markets that never score a 0 or 100 on "nigger+niggers."
7. Use coefficients from regression to back out "nigger+niggers" for remaining markets, using their average search volume for "weather" and "nigger+niggers+weather."

References

Alesina, Alberto and Eliana La Ferrara, "Who Trusts Others?," *Journal of Public Economics*, 2002, *85* (2), 207 – 234.

___ **and Howard Rosenthal**, *Partisan Politics, Divided Government, and the Economy*, New York: Cambridge University Press, 1995.

___ **, Ed Glaeser, and Bruce Sacerdote**, "Why Doesn't the US Have a European-Style Welfare System?," *Brookings Papers on Economic Activity*, August 2001, *2001* (2), 187–254.

Allport, Gordon Willard, *The Nature of Prejudice*, New York: Addison-Wesley, 1979.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151 – 184.

Ansolabehere, Stephen and Eitan Hersh, "Pants on Fire: Misreporting, Sample Selection, and Participation," 2011.

Askitas, Nikolaos and Klaus F. Zimmermann, "Google Econometrics and Unemployment Forecasting," *Applied Economics Quarterly*, 2009, *55* (2), 107 – 120.

Atkeson, Lonna Rae, "'Sure, I Voted for the Winner!' Overreport of the Primary Vote for the Party Nominee in the National Election Studies," *Political Behavior*, 1999, *21* (3), 19.

Benjamin, Daniel J. and Jesse M. Shapiro, "Thin-Slice Forecasts of Gubernatorial Elections," *The Review of Economics and Statistics*, August 2009, *91* (33), 523–536.

Berggren, Niclas, Henrik Jordahl, and Panu Poutvaara, "The Looks of a Winner: Beauty and Electoral Success," *Journal of Public Economics*, 2010, *94* (1-2), 8–15.

Berinsky, Adam J., "The Two Faces of Public Opinion," *American Journal of Political Science*, 1999, *43* (5), 1209 – 1230.

___ , "Political Context and the Survey Response: The Dynamics of Racial Policy Opinion," *The Journal of Politics*, July 2002, *64* (02).

Berry, William D., Evan J. Ringquist, Richard C. Fording, and Russell L. Hanson, "Measuring Citizen and Government Ideology in the American States, 1960-93," *American Journal of Political Science*, 1998, *42* (1), 327–348.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 2004, *94* (4), 991 – 1013.

Blalock, Hubert, *Toward a Theory of Minority-Group Relations*, New York: Wiley, 1967.

Bound, John, Charles Brown, and Nancy Mathiowetz, "Measurement Error in Survey Data," *Handbook of Econometrics*, 2001, *5*, 3705 – 3843.

Burns, Enid, "U.S. Search Engine Rankings, April 2007," *searchenginewatch.com*, May 2007.

Campbell, James E., "Forecasting the Presidential Vote in the States," *American Journal of Political Science*, 1992, *36* (2), 386–407.

Card, David, Alexandre Mas, and Jesse Rothstein, "Tipping and the Dynamics of Segregation," *The Quarterly Journal of Economics*, 2008, *123* (1), 177 – 218.

Charles, Kerwin Kofi and Jonathan Guryan, "Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination," *Journal of Political Economy*, 2008, *116* (5), 773–809.

___ **and** ___ , "Studying Discrimination: Fundamental Challenges and Recent Progress," 2011.

Conti, Gregory and Edward Sobiesk, "An Honest Man Has Nothing to Fear," in "Proceedings of the 3rd Symposium on Usable Privacy and Security - SOUPS '07" ACM Press New York, New York, USA July 2007, p. 112.

CPS, "Computer and Internet Use in the United States: October 2007," 2007.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor, "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 1999, *107* (3), 455 – 506.

DellaVigna, Stefano, "The Obama Effect on Economic Outcomes: Evidence from Event Studies," 2010.

___ **and Ethan Kaplan**, "The Fox News Effect: Media Bias and Voting," *The Quarterly Journal of Economics*, August 2007, *122* (3), 1187–1234.

Donovan, Todd, "Obama and the White Vote," *Political Research Quarterly*, August 2010, *63* (4), 863–874.

Drake, Michael S., Darren T. Roulstone, and Thornock Jacob R., "Investor Information Demand: Evidence from Google Searches Around Earnings Announcements," *Journal of Accounting Research*, January 2012.

Duggan, Mark, "More Guns, More Crime," *Journal of Political Economy*, 2001, *109* (5), 1086 – 1114.

Echenique, Federico and Roland G. Fryer, "A Measure of Segregation Based on Social Interactions," *The Quarterly Journal of Economics*, May 2007, *122* (2), 441–485.

Enos, Ryan D., "The Persistence of Racial Threat: Evidence from the 2008 Election," *American Political Science Association Annual Meeting*, 2010.

Fretland, Katie, "Gallup: Race Not Important to Voters," *The Chicago Tribune's The Swamp*, June 2008.

Gallup, "Presidential Approval Ratings: George W. Bush," 2012.

Gentzkow, Matthew and Jesse Shapiro, *Introduction of Television to the United States Media Market, 1946-1960; Dataset 22720*, Ann Arbor, MI: ICPSR, 2008.

___, ___, and Michael Sinkinson, "The Effect of Newspaper Entry and Exit on Electoral Politics," *American Economic Review*, 2011, *101* (7), 2980–3018.

Gerber, Alan S., Dean Karlan, and Daniel Bergan, "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions," *American Economic Journal: Applied Economics*, 2009, *1* (2), 18.

Gilens, Martin, Paul M. Sniderman, and James H. Kuklinski, "Affirmative Action and the Politics of Realignment," *British Journal of Political Science*, January 1998, *28* (01), 159–183.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant, "Detecting Influenza Epidemics Using Search Engine Query Data.," *Nature*, February 2009, *457* (7232), 1012–4.

Glaser, James, "Back to the Black Belt: Racial Environment and White Racial Attitudes in the South," *The Journal of Politics*, 1994, *56*, 21 – 41.

___ and Martin Gilens, "Interregional Migration and Political Resocialization: A Study of Racial Attitudes Under Pressure," *Public Opinion Quarterly*, 1997, *61* (1), 72–86.

Hart/McInturff, "Study 12062," *NBC News/Wall Street Journal*, January 2012.

Heerwig, Jennifer A. and Brian J. McCabe, "Education and Social Desirability Bias: The Case of a Black Presidential Candidate," *Social Science Quarterly*, September 2009, *90* (3), 674–686.

Highton, Benjamin, "Prejudice Rivals Partisanship and Ideology When Explaining the 2008 Presidential Vote Across the States," *PS: Political Science & Politics*, July 2011, *44* (03), 530–535.

Hopkins, Heather, "Yahoo! Draws Younger Audience; Google Users Big Spenders Online," *Hitwise Intelligence*, February 2008.

Huddy, Leonie and Stanley Feldman, "On Assessing the Political Effects of Racial Prejudice," *Annual Review of Political Science*, June 2009, *12* (1), 423–447.

Huffman, Matt L. and Philip N. Cohen, "Racial Wage Inequality: Job Segregation and Devaluation Across U.S. Labor Markets," *American Journal of Sociology*, 2004, *109* (4), 902–936.

Jackman, Simon and Lynn Vavreck, "How Does Obama Match-Up? Counterfactuals & the Role of Obama's Race in 2008," 2011.

Jacobson, Gary C., "The 2008 Presidential and Congressional Elections: Anti-Bush Referendum and Prospects for the Democratic Majority," *Political Science Quarterly*, 2009, *124* (1), 30.

Kam, Cindy D. and Donald R. Kinder, "Ethnocentrism as a Short-Term Force in the 2008 American Presidential Election," *American Journal of Political Science*, February 2012, *56* (2), 326–340.

Kennedy, Randall, *Nigger: The Strange Career of a Troublesome Word*, New York: Vintage Books, 2003.

Key Jr., Valdimer O., *Southern Politics in State and Nation*, New York: A.A. Knopf, 1949.

Knowles, John, Nicola Persico, and Petra Todd, "Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy*, 2001, *109* (1), 203 – 232.

Kreuter, Frauke, Stanley Presser, and Roger Tourangeau, "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity," *Public Opinion Quarterly*, January 2009, *72* (5), 847–865.

Kuklinski, James H., Michael D. Cobb, and Martin Gilens, "Racial Attitudes and the 'New South'," *The Journal of Politics*, May 1997, *59* (02), 323–349.

Lewis-Beck, Michael S. and Tom W. Rice, "Localism in Presidential Elections: The Home State Advantage," *American Journal of Political Science*, 1983, *27* (3), 548–556.

____ , Charles Tien, and Richard Nadeau, "Obama's Missed Landslide: A Racial Cost?," *PS: Political Science & Politics*, January 2010, *43* (01), 69–76.

List, John A., "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field," *The Quarterly Journal of Economics*, 2004, *119* (1), 49 – 89.

Malter, Daniel, "Analyzing Obama's Out-and Under Performance in the 2008 Presidential Elections: Social Desirability Bias, Sample Selection, and Momentum Neglect in the Polls," *SSRN Electronic Journal*, June 2010.

Mas, Alexandre and Enrico Moretti, "Racial Bias in the 2008 Presidential Election," *American Economic Review*, 2009, *99* (2), 323 – 29.

Mixon, Franklin and J. Matthew Tyrone, "The 'Home Grown' Presidency: Empirical Evidence on Localism in Presidential Voting, 1972-2000," *Applied Economics*, 2004, *36* (16), 1745–1749.

Parsons, Christopher A., Johan Sulaeman, Michael C. Yates, and Daniel Hamermesh, "Strike Three: Discrimination, Incentives, and Evaluation," *American Economic Review*, 2011, *101* (4), 1410–35.

Pasek, Josh, Alexander Tahk, Yphtach Lelkes, Jon A. Krosnick, B. Keith Payne, Omair Akhtar, and T. Tompson, "Determinants of Turnout and Candidate Choice in the 2008 U.S. Presidential Election: Illuminating the Impact of Racial Prejudice and Other Considerations," *Public Opinion Quarterly*, January 2010, *73* (5), 943–994.

Piston, Spencer, "How Explicit Racial Prejudice Hurt Obama in the 2008 Election," *Political Behavior*, 2010, *32* (4), 431–451.

Price, Joseph and Justin Wolfers, "Racial Discrimination Among NBA Referees," *The Quarterly Journal of Economics*, 2010, *125* (4), 1859–1887.

Rahman, Jacquelyn, "The N Word: Its History and Use in the African American Community," *Journal of English Linguistics*, July 2011.

Saiz, Albert and Uri Simonsohn, "Downloading Wisdom from Online Crowds," October 2008.

Schaffner, Brian F., "Racial Salience and the Obama Vote," 2011.

Scheitle, Christopher P., "Google's Insights for Search: A Note Evaluating the Use of Search Engine Data in Social Research," *Social Science Quarterly*, 2011, *92* (1), 285 – 295.

Seifter, Ari, Alison Schwarzwalder, Kate Geis, and John Aucott, "The Utility of 'Google Trends' for Epidemiological Research: Lyme Disease as an Example," *Geospatial Health*, May 2010, *4* (2), 135–137.

Silver, Nate, "Does Race Affect Votes?," *TED Talk*, 2009.

Taylor, Marylee C., "How White Attitudes Vary with the Racial Composition of Local Populations: Numbers Count," *American Sociological Review*, 1998, *63* (4), 512–535.

Tesler, Michael and David O. Sears, *Obama's Race: The 2008 Election and the Dream of a Post-Racial America*, Vol. 2010, Chicago: University of Chicago Press, 2010.

Tourangeau, Roger and Yan Ting, "Sensitive Questions in Surveys," *Psychological Bulletin*, 2007, *133* (5), 859–883.

Valentino, Nicholas A. and Ted Brader, "The Sword's Other Edge: Perceptions of Discrimination and Racial Policy Opinion after Obama," *Public Opinion Quarterly*, May 2011, *75* (2), 201–226.

Varian, Hal R. and Hyunyoung Choi, "Predicting the Present with Google Trends," *SSRN Electronic Journal*, August 2010.

Washington, Ebonya, "How Black Candidates Affect Voter Turnout," *Quarterly Journal of Economics*, August 2006, *121* (3), 973–998.

Wolfers, Justin, "Are Voters Rational? Evidence from Gubernatorial Elections," Technical Report March 2002.

Wright, Gerald C., "Errors in Measuring Vote Choice in the National Election Studies, 1952-88," *American Journal of Political Science*, 1993, *37* (1), 291–316.

Table I
Signal-to-Noise Ratio in Google Search Terms

Term	*Underlying Variable*	*t-stat*	R^2
God	Percent Believe in God	8.45	0.65
Gun	Percent Own Gun	8.94	0.62
African American(s)	Percent Black	13.15	0.78
Hispanic	Percent Hispanic	8.71	0.61
Jewish	Percent Jewish	17.08	0.86

Notes: The t-stat and R^2 are from a regression with the normalized search volume of the word(s) in the first column as the independent variable and measures of the value in the second column as the dependent variable. The normalized search volume for all terms are from 2004-2007. All data are at the state level. Percent Black are Percent Hispanic are from the American Community Survey, for 2008; the Jewish population is from 2002, gun ownership from 2001, and belief in God from 2007. Jewish data are missing one observation (South Dakota); belief in God data are missing for 10 states. The data for belief in God, percent Jewish, and percent owning guns can be found at `http://pewforum.org/how-religious-is-your-state-.aspx`, `http://www.jewishvirtuallibrary.org/jsource/US-Israel/usjewpop.html`, and `http://www.washingtonpost.com/wp-srv/health/interactives/guns/ownership.html`, respectively.

Table II
Google Search Volume on Topics Thought to be Underreported in Surveys

Term	*Annual Google Searches*
Porn	996 million
Weather	*816 million*
Sex	667 million
Marijuana	49 million
Suicide	40 million
Escort	27 million
Herpes	21 million
Viagra	18 million

Notes: Data downloaded from Google AdWords on 8/26/11. At `http://adwords.google.com`, I click on Tools & Analysis - Keyword Tool. I then type in the word and click on Match Types (Phrase). This tells me the average monthly searches including the word, on Desktops and Laptops, in the United States. I multiply by 12 to obtain the annual estimate.

Table III
Top Searches for “nigger(s)”

Rank	*’04-’07 Search* DATA USED	*’08-’11 Search* DATA NOT USED
1.	jokes	jokes
2.	nigger jokes	nigger jokes
3.	white nigger	obama nigger
4.	nigga	nigga
5.	hate niggers	black nigger
6.	i hate niggers	funny nigger
7.	black jokes	nigger song
8.	the word nigger	the word nigger
9.	racist jokes	nas nigger
10.	kkk	i hate niggers

Notes: This table shows the ‘top searches’ for “nigger(s).” 2004-2007 is the time period for the search volume used in the regressions and figures to limit reverse causation. Results would be similar regardless of time period selected, as the state-level correlation between the two periods is 0.94. Depending on the draw, the ‘top searches’ might be slightly different. Top searches, according to Google, ‘are related to the term,’ as determined ‘by examining searches that have been conducted by a large group of users preceding the search term you’ve entered, as well as after,’ as well as by automatic categorization.

Table IV
Predictors of an Area's Racially Charged Search

	Dependent Variable: Racially Charged Search			
	(1)	(2)	(3)	(4)
Percent Age 65 or Older	6.884**	3.341	6.492**	3.757
	(3.142)	(3.567)	(3.133)	(3.543)
Percent w/ Bachelor's Degree	-9.309***	-8.532***	-10.104***	-9.459***
	(1.535)	(1.770)	(1.470)	(1.755)
Percent Hispanic	-2.620***	-2.298***	-2.659***	-2.297***
	(0.386)	(0.506)	(0.373)	(0.474)
Percent Black	2.556***	0.283	11.245***	6.734**
	(0.715)	(1.134)	(2.111)	(2.786)
(Percent Black)-squared			-24.731***	-16.517***
			(5.477)	(5.710)
Observations	196	196	196	196
R-squared	0.36	0.49	0.41	0.50
Census Div. FE		X		X

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Notes: Robust standard errors in parentheses. The dependent variable is Racially Charged Search, as defined in Equation 1, obtained by the algorithm described in Appendix B, normalized to its z-score. The demographic variables are individuals in the group divided by total individuals; thus a one-unit change represents a change from 0 to 100 percent.

Table V
Summary Statistics

	mean	sd	min	max
Racially Charged Search	39.78	9.21	16.62	100.00
Obama	53.71	10.20	22.16	75.05
Kerry	48.79	9.59	19.89	70.06
Obama-Kerry	4.93	3.18	-10.98	18.60
Δ House Dems	7.26	8.75	-39.16	72.59
Δ ln(Turnout)	0.07	0.06	-0.10	0.25

Notes: All summary statistics are reported for the 196 media markets for which data on Racially Charged Search and voting data are available. All summary statistics reported are weighted by 2004 two-party turnout, the weighting used in Tables VII and X. Racially Charged Search is as defined in Equation 1, obtained by the algorithm described in Appendix B. Obama is Barack Obama's share of 2008 two-party Presidential votes. Kerry is John Kerry's share of 2004 two-party Presidential votes. Δ House Dems is the change in the Democratic House two-party vote share from 2004 to 2008. Δ ln(Turnout) is the change in the natural log of total two-party votes from 2004 to 2008.

Table VI
Music, “nigger,” and “nigga,” 2004-2007

Rank	*Top searches for ‘nigger lyrics’*	*Top searches for ‘nigga(s)’*
1.	nigger song	nigga lyrics
2.	nigger song lyrics	my nigga
3.	nigger jokes	niggas lyrics
4.	white nigger	hood nigga
5.	nigger hatin me	my niggas
6.	white nigger lyrics	lyrics hood nigga
7.	johnny rebel lyrics	nigga stole
8.	johnny rebel	nigga stole my
9.	david allen coe	my nigga lyrics
10.	lyrics alabama nigger	nigga what

Notes: The second column shows the ‘top searches’ reported for searches including both “nigger” and “lyrics.” The third column shows the ‘top searches’ reported for searches including either “nigga” or “niggas.” The method for calculating ‘top searches’ is discussed in Table III. Also noted there, depending on the particular draw, the ranks and terms might differ somewhat.

Table VII
Obama-Kerry and Racially Charged Search

	Dependent Variable: Obama - Kerry					
	(1)	(2)	(3)	(4)	(5)	(6)
Racially Charged Search	-1.490***	-1.486***	-1.341***	-2.124***	-2.002***	-1.776***
	(0.261)	(0.259)	(0.294)	(0.343)	(0.324)	(0.358)
Home State	2.616***	4.234***	3.556***	2.481***	4.070***	3.636***
	(0.579)	(0.859)	(0.807)	(0.694)	(0.992)	(0.791)
Observations	196	196	196	196	196	196
R-squared	0.26	0.51	0.62	0.30	0.52	0.62
Standard Controls	X	X	X	X	X	X
Census Div. FE		X	X		X	X
Demographic Controls			X			X
Google Controls				X	X	X

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Notes: Robust standard errors in parentheses. OLS regressions are weighted by total two-party presidential votes in the 2004 election. Racially Charged Search is as defined in Equation 1, obtained by the algorithm described in Appendix B, normalized to its z-score. Home State takes the value 1 for Illinois and Texas; -1 for Massachusetts and Arizona; 0 otherwise. Standard controls are Home State and unemployment rates in years 2003, 2004, 2007, and 2008 (from Local Area Unemployment Statistics). Demographic controls are percent African-American, percent Hispanic, percent with bachelor's degree, percent 18-34, percent 65+, and percent veteran (from American Community Survey '05-'09); change from 2000 to 2010 in percent African-American and percent Hispanic (from the Census); and gun magazine subscriptions per capita (from Duggan (2001)). Google controls are normalized search volume for "African-American(s);" "nigga(s);" and "fuck," also obtained by the algorithm described in Appendix B.

Table VIII
Sensitivity of Coefficient on Racially Charged Search to Alternative Specifications

	Coefficient on Racially Charged Search	
Specification	*Standard Controls*	*Standard Controls and Google Controls*
Baseline	−1.490	−2.124
	(0.261)	(0.261)
Unweighted	−1.547	−1.840
	(0.218)	(0.303)
Weight by '08 Turnout	−1.496	−2.149
	(0.263)	(0.347)
Weight by Voting Age Population	−1.577	−2.256
	(0.263)	(0.348)
Exclude Hawaii	−1.415	−2.027
	(0.248)	(0.331)
Include Control for Kerry-Gore	−1.259	−1.875
	(0.287)	(0.336)
Include Controls for Kerry-Gore and Census Div. FE	−1.483	−1.983
	(0.268)	(0.327)
Include Control for Change in House Voting	−1.457	−2.094
	(0.254)	(0.338)
Include Control for Swing State	−1.507	−2.116
	(0.267)	(0.342)
Use All Votes Instead of Two-Party Votes	−1.432	−2.065
	(0.270)	(0.339)
Include Control for Percent Kerry	−1.492	−2.152
	(0.261)	(0.349)
Include 4th-Order Polynomial Percent Kerry	−1.531	−2.162
	(0.260)	(0.346)

Notes: Robust standard errors in parentheses. Results in this table are variations on Columns (1) and (4) reported in Table VII. The dependent variable for all but the final two rows is the same as in Table VII. Swing State status are Battleground States, as defined by *The Washington Post*, available at `http://www.washingtonpost.com/wp-dyn/content/graphic/2008/06/08/GR2008060800566.html`. The dependent variable for the final two rows is Obama's share of the 2008 two-party Presidential vote. The penultimate row includes Kerry's share of the 2004 two-party Presidential vote as an independent variable. The final row includes a 4th-order polynomial of Kerry's vote share.

Table IX
Country-Wide Effect: Google Compared to Other Measures

				Percentage Points:	
Source	*Obs*	*Measure*	*Controls*	*Point Estimate*	*Lower-end of 95 Percent Confidence Interval*
Google	Media Market	Racially Charged Search, '04-'07	Standard	-3.5	-2.3
		Racially Charged Search, '04-'07	Standard+Census Div. FE	-3.5	-2.3
		Racially Charged Search, '04-'07	Standard+Google	-5.0	-3.4
		Racially Charged Search, '04-'07	Standard+Google+Census Div. FE	-4.7	-3.2
GSS	State	Oppose Interracial Marriage, '90-'04	Standard	-2.0	-0.5
		Oppose Interracial Marriage, '90-'04	Standard+Census Div. FE	-0.5	1.9
		Average Prejudice, '72-'04	Standard	-2.8	-0.7
		Average Prejudice, '72-'04	Standard+Census Div. FE	-0.5	2.6
ANES	Individual	Explicit Prejudice	Piston (2010)	-2.3	-0.4
APYN		Explicit and Implicit Prejudice	Pasek et al. (2010)	-2.7	
CCES		Racial Salience	Schaffner (2011)	-2.0	

Notes: This table compares the results obtained using the Google data to those using the same specification but measures from the GSS and the estimate obtained by other scholars using individual proxies for racial attitudes and individual reported votes. For all regressions used to calculate the estimated percentage points using Google or GSS, the regressions are weighted by total two-party presidential votes in 2004. The point estimate is then the country-wide effect of moving from the area with the lowest value. Standard controls are Home State and unemployment rates in 2003, 2004, 2007, and 2008. The first GSS measure is from Mas and Moretti (2009). The second GSS measure is from Charles and Guryan (2008). The lower-end of the 95 percent confident interval is calculated using robust standard errors. Piston (2010) finds that overall prejudice cost Obama 2.66 percent of the white vote. Assuming whites accounted for 87% of the electorate yields the number of -2.3.

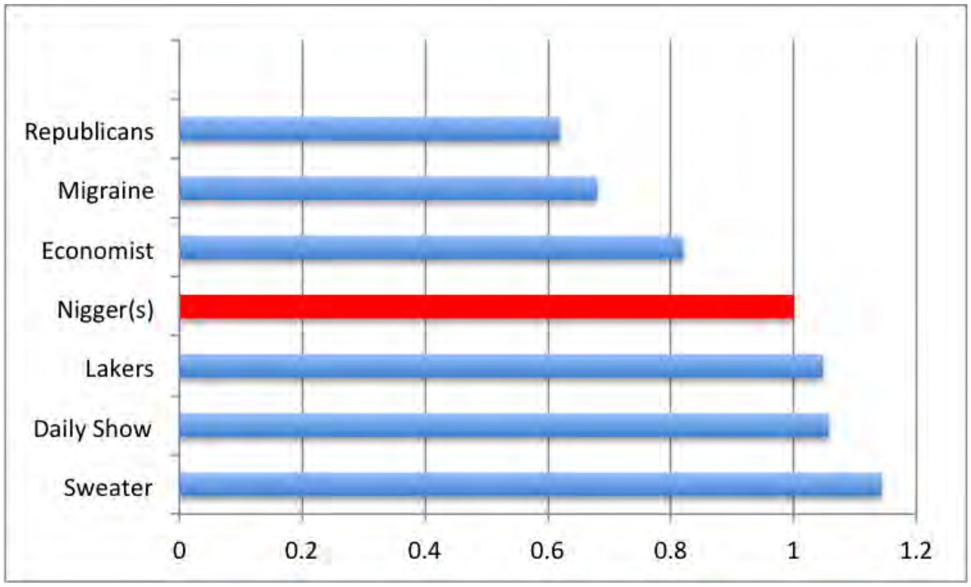
Table X
Change in Turnout (2004-2008) and Racially Charged Search

		Dependent Variable: Δ ln(Turnout)				
	(1)	(2)	(3)	(4)	(5)	(6)
Racially Charged Search	-0.001	-0.001	0.004	0.025*	0.032*	0.033**
	(0.004)	(0.004)	(0.004)	(0.013)	(0.019)	(0.017)
Racially Charged Search · Kerry				-0.056*	-0.071	-0.064*
				(0.031)	(0.045)	(0.038)
Observations	196	196	196	196	196	196
R-squared	0.67	0.73	0.80	0.67	0.74	0.80
Census Div. FE		X	X		X	X
Demographic Controls			X			X

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

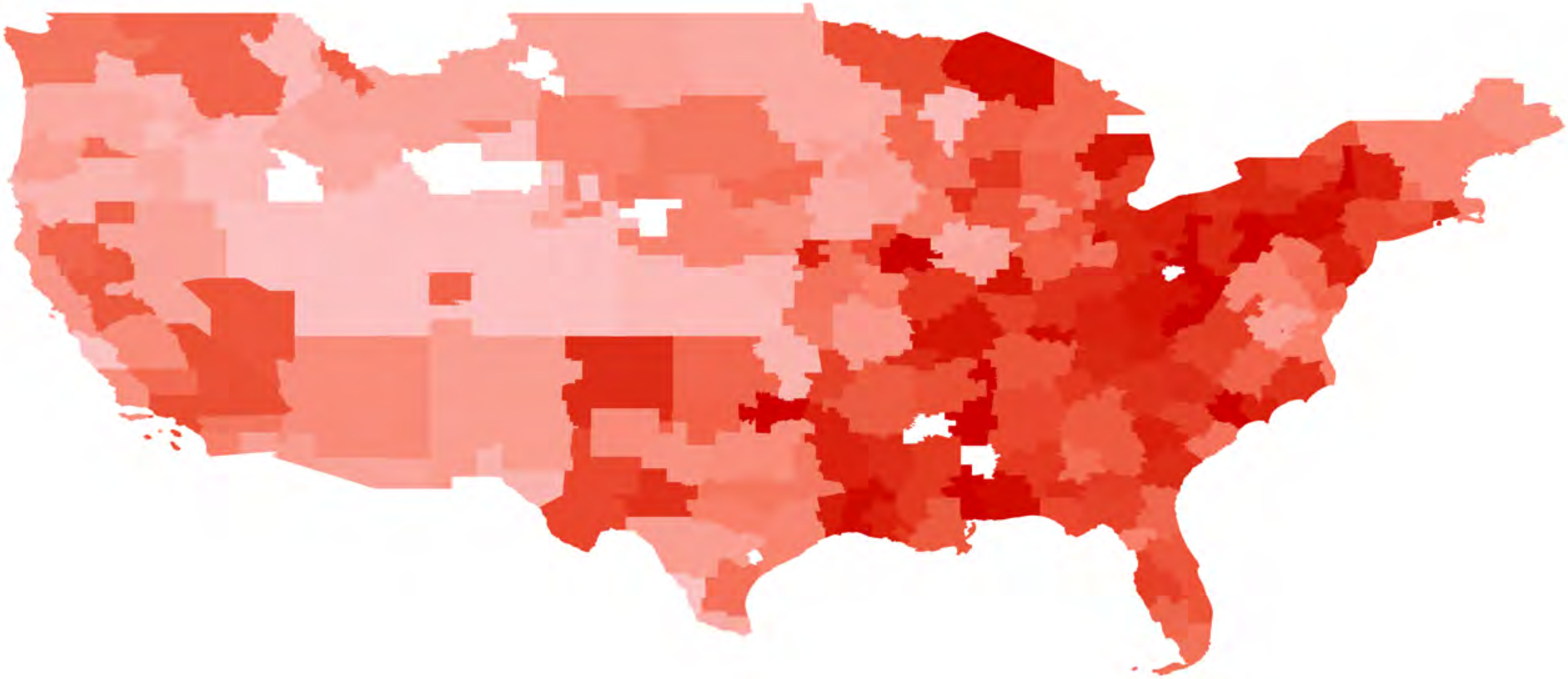
Notes: Robust standard errors in parentheses. OLS regressions are weighted by total two-party presidential votes in the 2004 election. Racially Charged Search is as defined in Equation 1, obtained by the algorithm described in Appendix B, normalized to its z-score. Dependent variable in all specifications is the natural log of two-party presidential votes in 2008 minus the natural log of two-party presidential votes in 2004. All regressions include change in log population from 2000 to 2010 (from the Census); percent African-American (from American Community Survey '05-'09); and Kerry's share of the two-party vote. Columns (3) and (6) add percent African-American, percent Hispanic, percent with bachelor's degree, percent 18-34, percent 65+, and percent veteran (from American Community Survey '05-'09); change from 2000 to 2010 in percent African-American and percent Hispanic (from the Census); and gun magazine subscriptions per capita (from Duggan (2001)).

Figure I
Selected Words and Phrases Included in Google Searches Roughly as Frequently as “nigger(s),” 2004-2007



Notes: This figure shows selected words and phrases included in a similar number of searches, from 2004-2007, as “nigger(s).” The number corresponds to the ratio of total Google searches that include that word to total Google searches that include the racial epithet. “Daily Show,” for example, was included in about 6 % more searches than the racial epithet. “Economist” was included in about 20 % fewer searches. It is worth emphasizing again that this counts any searches including the word or phrase. So searches such as “The Daily Show” and “Daily Show clips” will be counted in the search total for “Daily Show.” And Google considers searches case-insensitive. So “daily show” and “daily show clips” would also count. While the words included were rather arbitrarily selected, another benchmark to use is “weather.” “Weather” was included in only about 81 times more searches than “nigger(s)” during this time period. All numbers presented were estimated using Google Insights.

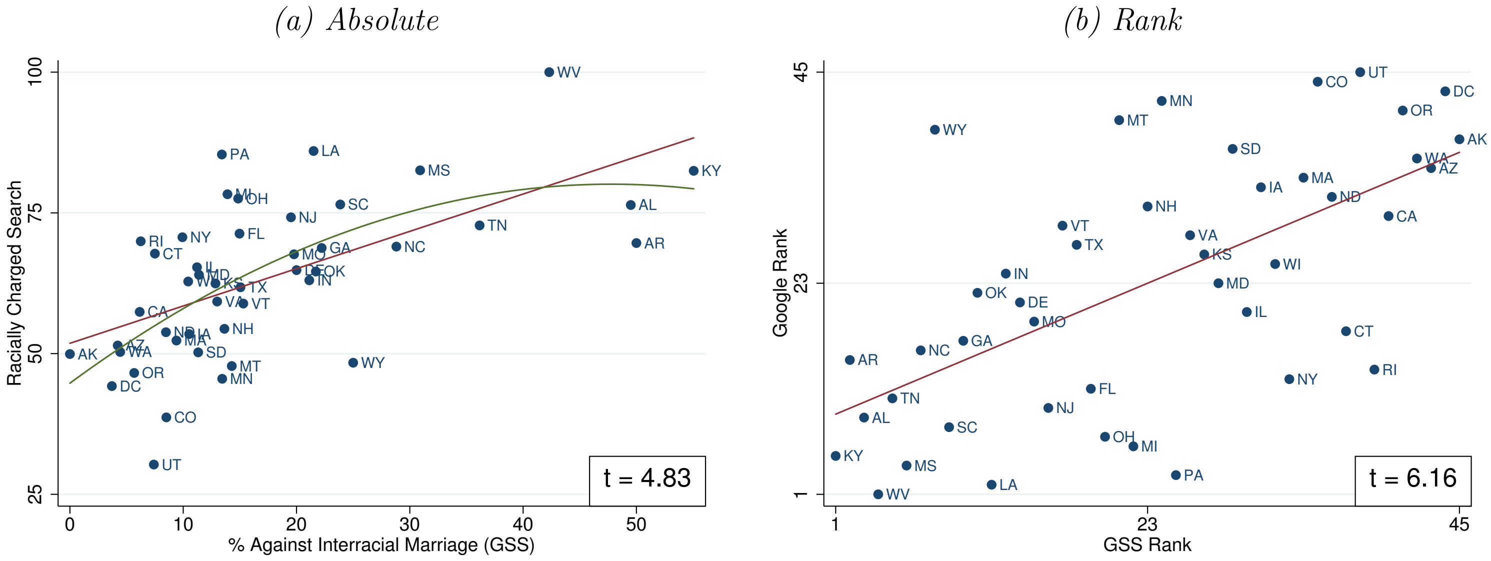
Figure II
Racially Charged Search, Media Market



Notes: This maps search volume for "nigger(s)," from 2004-2007, at the media market level. The numbers were obtained by an algorithm explained in Appendix B. Darker areas signify higher search volume. White areas signify media markets without data. Alaska and Hawaii, for which data are available, are not shown.

Figure III

Google Racially Charged Search Compared to GSS Opposition to Interracial Marriage

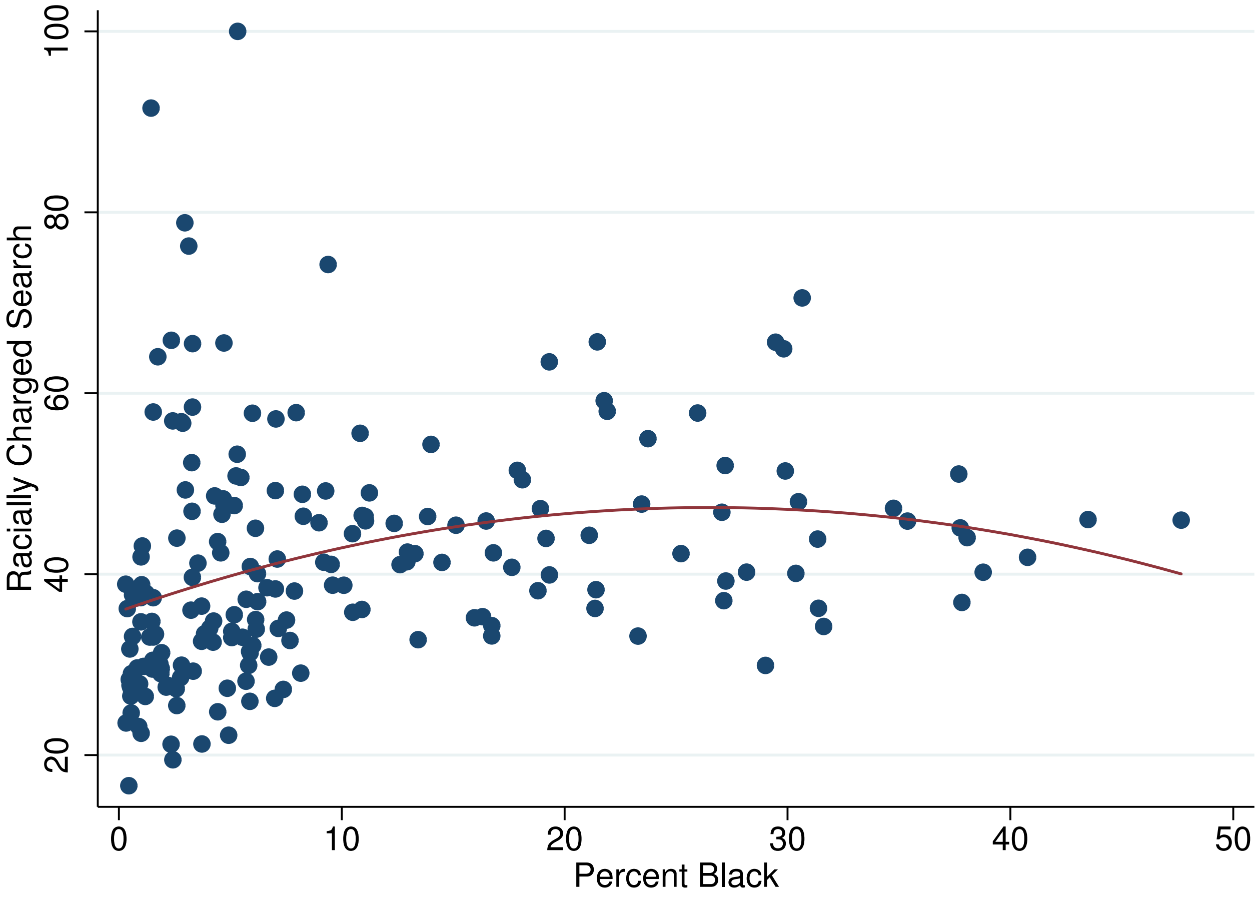


(a) Absolute

(b) Rank

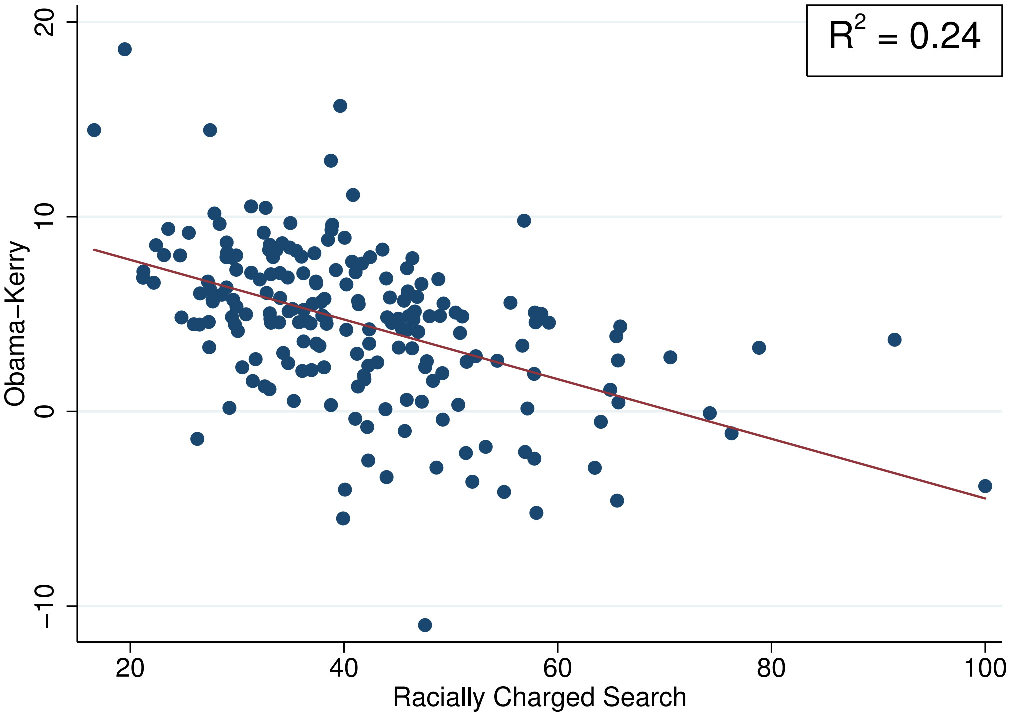
Notes: The x axis in panels (a) is the measure of racial attitudes used in Mas and Moretti (2009): percent of whites, from 1990-2004, supporting a law banning interracial marriage. The (x) axis in panel (b) is the rank of the 45 states for this measure. Thus, the value 1 in Panel (b) means that state (Kentucky) had the highest percentage of whites telling the GSS they supported a law banning interracial marriage. The y axis for panel (a) uses the unrounded number in Table A.1 for the 45 states for which GSS data are available; The y axis panel (b) is the rank of racially charged search for these 45 states.

Figure IV
Racially Charged Search and African-American Population



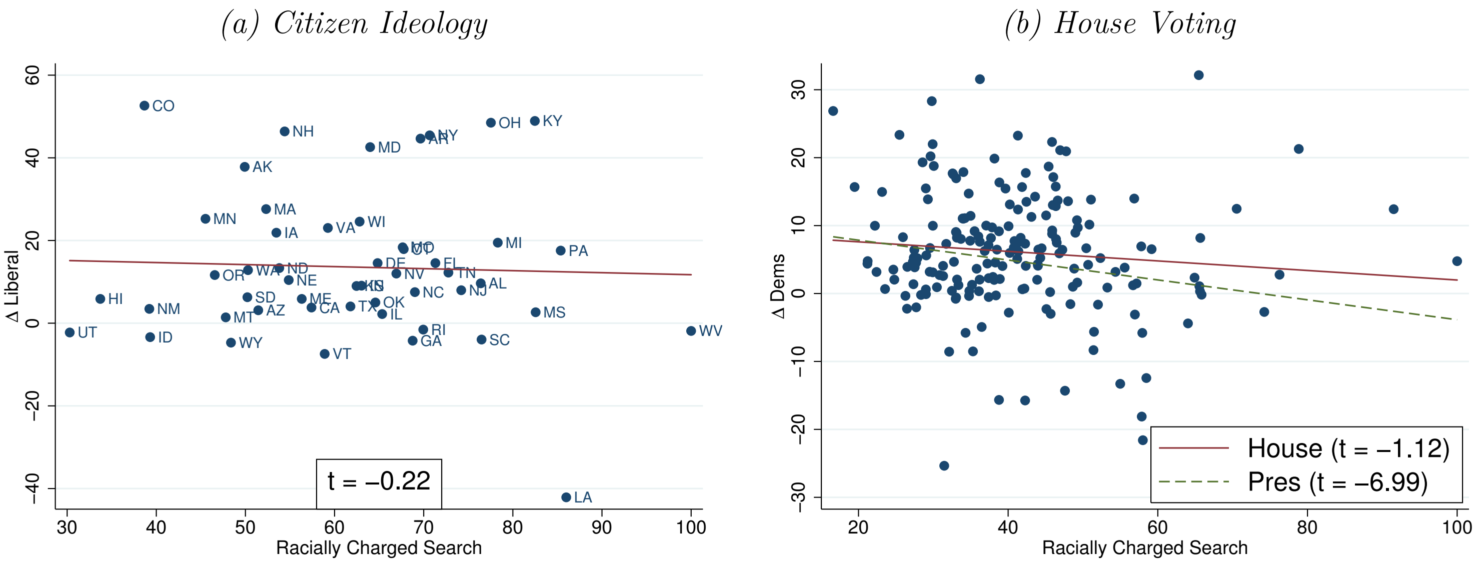
Notes: This figure shows the relationship between the size of the African-American population and racially charged search, at the media market level. Percent black is from the American Community Survey '05-'09. The preferred quadratic fit is shown.

Figure V
Obama-Kerry and Racially Charged Search



Notes: This figure plots the correlation between a media market's normalized search volume, from 2004-2007, for racially charged language and the difference between Obama and Kerry vote shares. The number on the y-axis is Kerry's percentage points of the two-party vote subtracted from Obama's percentage points of the two-party vote. The number on the x-axis is as defined in Equation 1, obtained by the algorithm described in Appendix B

Figure VI
Change in Liberalism (2004-2008) and Racially Charged Search

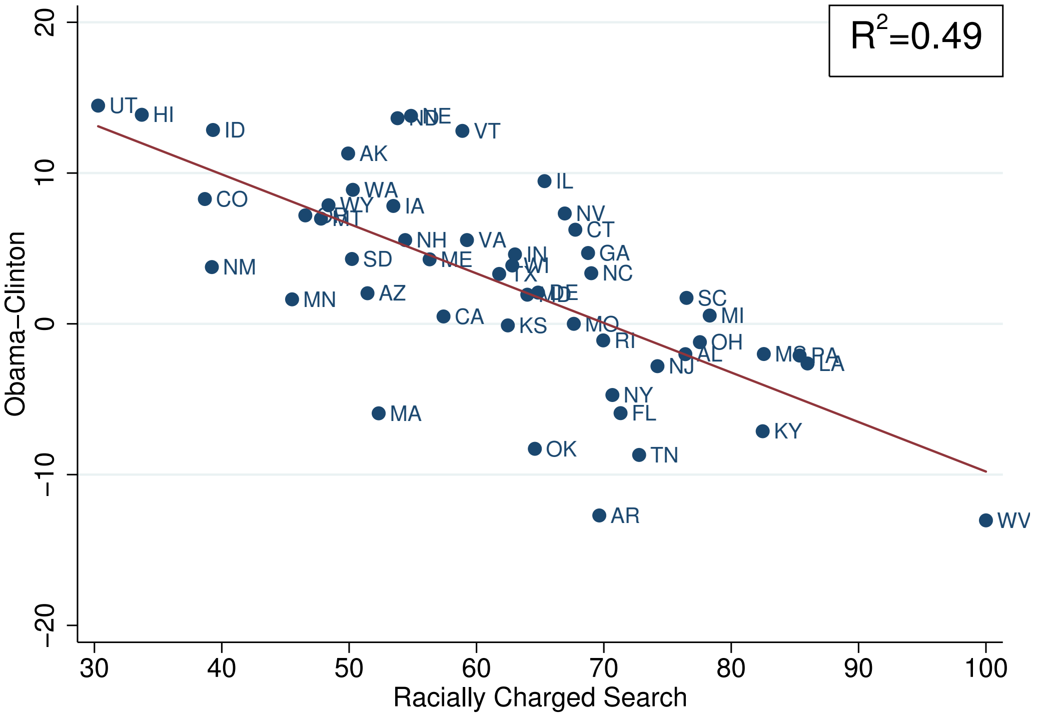


(a) Citizen Ideology

(b) House Voting

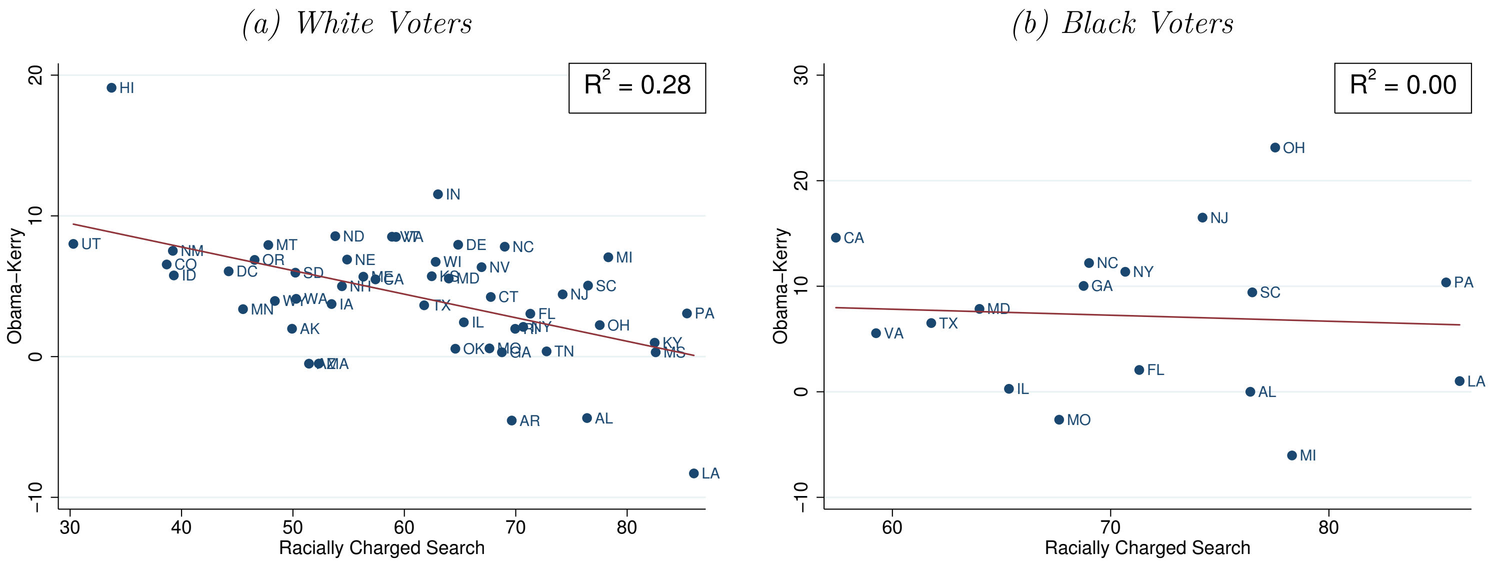
Notes: The x axis for panel (a) is the unrounded value from Table A.1. The x axis for panel (b) is the number used in Figure V. The y axis in panel (a) measures change in liberalism, from 2004 to 2008, according to the "revised 1960-2008 citizen ideology series." This commonly used proxy is described in Berry et al. (1998). To construct the y axis variable for panels (b), counties are dropped if either major party received 0 votes in House elections in 2004 and/or 2008. Vote totals are aggregated to the media market level in Panel (b), and the difference in the two-party share for Democrats is calculated. The solid line shows the best linear fit of the points. The dotted line shows the best linear fit of points (not shown) representing the difference in Obama and Kerry vote shares over the same, limited 'unopposed' sample. The relationship between racially charged search and changing Democratic House support using all races, not just races fielding candidates from both major parties, is not significant, either. An earlier draft of this paper shows that there is not a statistically significant relationship between racially charged search and Berry et al.'s (1998)'s "ADA/COPE measure of state government ideology" nor racially charged search and House voting, at the state level.

Figure VII
Obama-Clinton (Whites) and Racially Charged Search



Notes: The x-axis is the unrounded value from Table A.1. The y-axis is based on polling data from SurveyUSA. In particular, it is the difference in Barack Obama and Hillary Clinton's two-party vote shares, among whites, in a hypothetical matchup with John McCain. The polls were conducted in February 2008. They are available for all states but not the District of Columbia. The data were first used by Donovan (2010). I am grateful to the author for providing the data.

Figure VIII
Obama-Kerry and Racially Charged Search, by Race

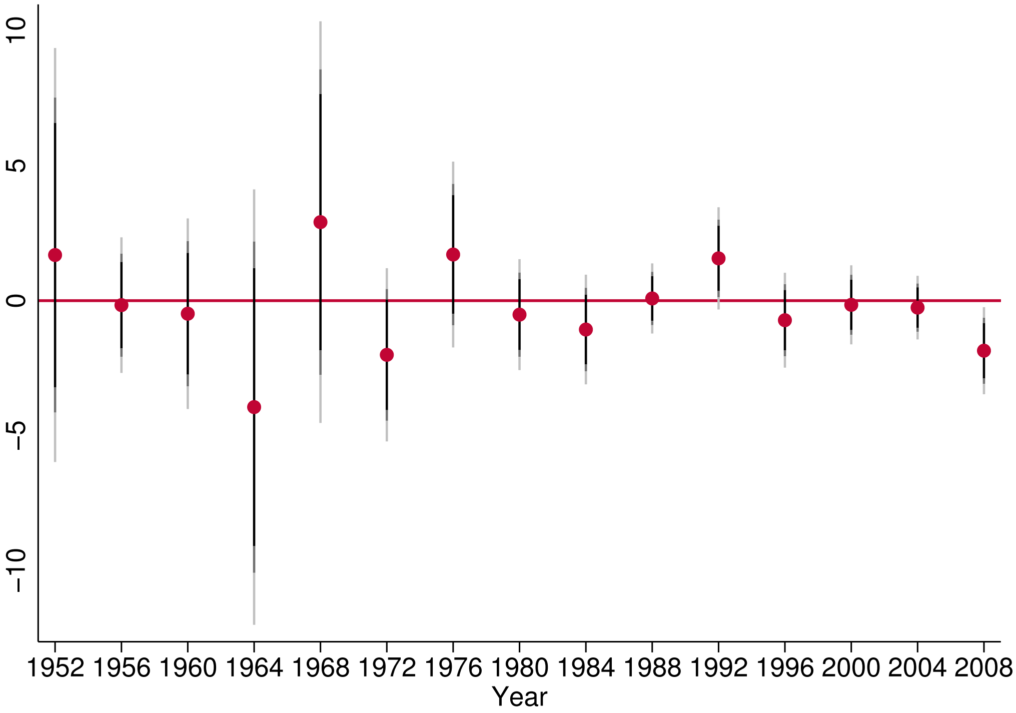


(a) White Voters

(b) Black Voters

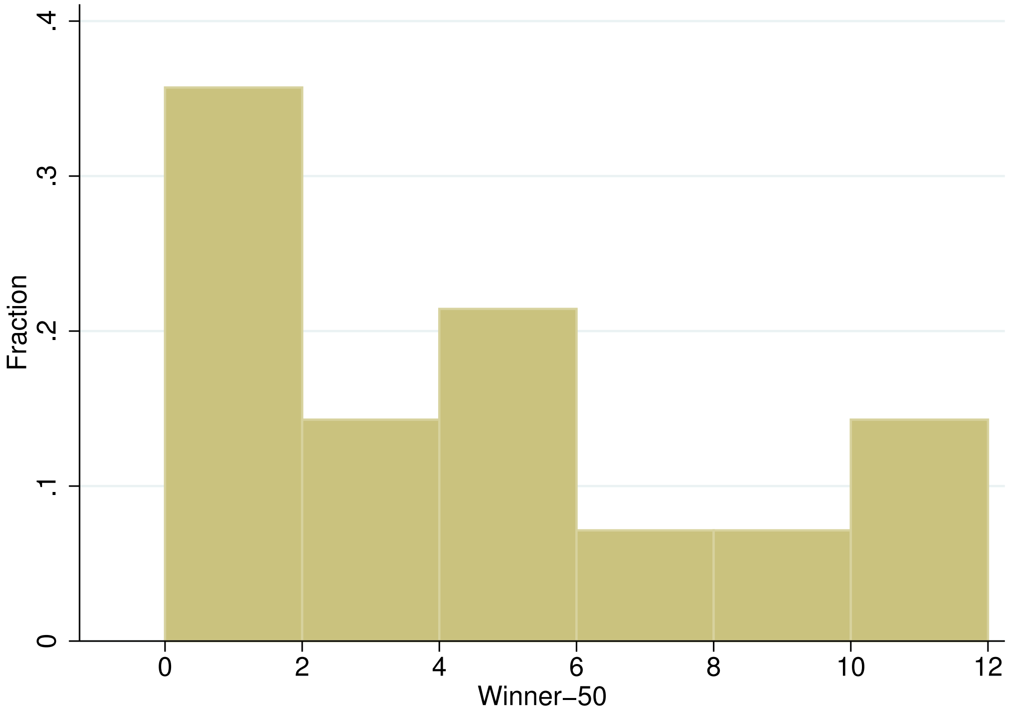
Notes: Both panels of this figure use the unrounded value from Table A.1. Panel (a) compares the difference in the two-party vote share for Obama and Kerry, according to exit polls, among whites. This data were graciously provided by Highton (2011). Exit poll data from 2004 for West Virginia are not available. Panel (b) compares the difference in the two-party vote share for Obama and Kerry, according to exit polls, among African-Americans. I do not include any states for which the number of African-Americans surveyed was fewer than 20 in either 2004 or 2008. There is no significant relationship, though far more noise, among racially charged search and black voters' support, using all states.

Figure IX
Post-War Presidential Elections and Racially Charged Search



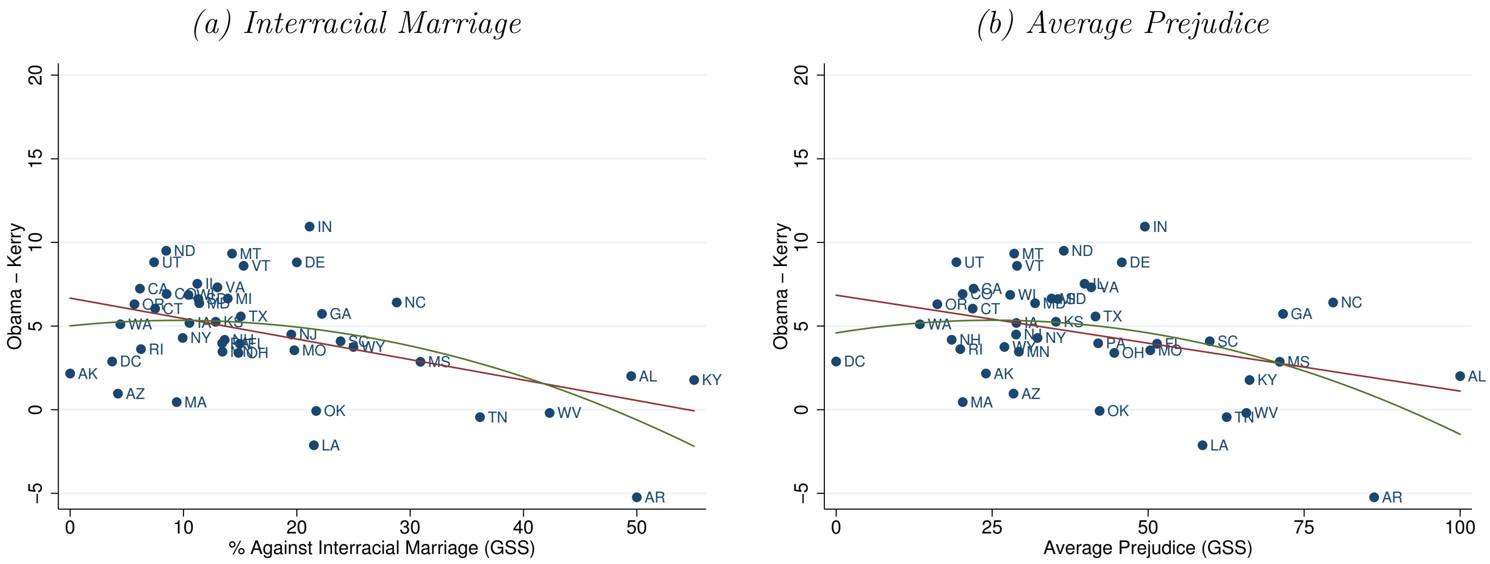
Notes This figure shows point estimate and confidence intervals for regressions of change in Democratic presidential candidate support on racially charged search. Three levels of confidence are shown (99 %; 95 %; and 90 %) with darker parts of the line showing lower confidence intervals. Racially charged search is the unrounded state-level number from Table A.1. All regressions are weighted by 2004 two-party votes and include controls for nine Census division. Large standard errors in 1964 and 1968 are due, in part, to Lyndon Johnson's not appearing on the Alabama ballot in 1964.

Figure X
Winner's Two-Party Vote Share Minus 50, Post-War Presidential Elections



Notes: This figure plots a histogram of the percentage points of the two-party vote the popular vote-winning candidate would have had to lose to have lost the popular vote for all American presidential elections between 1952 and 2004, inclusive.

Figure XI
Obama-Kerry and Racial Attitudes using GSS Data



(a) Interracial Marriage

(b) Average Prejudice

Notes: The y-axis is the state-level difference in Obama's and Kerry's shares of the two-party vote. The x-axis is a measure from the General Social Survey: Panel (a) uses the measure of Mas and Moretti (2009): percent of whites supporting a law banning interracial marriage, from 1990-2004. Panel (b) uses the measure of Charles and Guryan (2008): average prejudice, from 1972-2004, normalized so the top state scores 100. The best linear and quadratic fits are shown.

Detecting Hate Speech on the World Wide Web

William Warner and Julia Hirschberg
Columbia University
Department of Computer Science
New York, NY 10027
whw2108@columbia.edu, julia@cs.columbia.edu

Abstract

We present an approach to detecting *hate speech* in online text, where hate speech is defined as abusive speech targeting specific group characteristics, such as ethnic origin, religion, gender, or sexual orientation. While hate speech against any group may exhibit some common characteristics, we have observed that hatred against each different group is typically characterized by the use of a small set of high frequency stereotypical words; however, such words may be used in either a positive or a negative sense, making our task similar to that of words sense disambiguation. In this paper we describe our definition of hate speech, the collection and annotation of our hate speech corpus, and a mechanism for detecting some commonly used methods of evading common "dirty word" filters. We describe pilot classification experiments in which we classify anti-semitic speech reaching an accuracy 94%, precision of 68% and recall at 60%, for an F1 measure of .6375.

1 Introduction

Hate speech is a particular form of offensive language that makes use of stereotypes to express an ideology of hate. Nockleby (Nockleby, 2000) defines hate speech as "any communication that disparages a person or a group on the basis of some characteristic such as race, color, ethnicity, gender, sexual orientation, nationality, religion, or other characteristic." In the United States, most hate speech is protected by the First Amendment of the U. S. Constitution, which, except for obscenity, "fighting words" and incitement, guarantees the right to free speech, and internet commentators exercise this right in online forums such as blogs, newsgroups, Twitter and Facebook. However, *terms of service* for such hosted services typically prohibit hate speech. Yahoo! Terms Of Service [1] prohibits posting "Content that is unlawful, harmful, threatening, abusive, harassing, tortuous, defamatory, vulgar, obscene, libelous, invasive of another's privacy, hateful, or racially, ethnically or otherwise objectionable." Facebook's terms [2] are similar, forbidding "content that: is hateful, threatening, or pornographic; incites violence." While user submissions are typically filtered for a fixed list of offensive words, no publicly available automatic classifier currently exists to identify hate speech itself.

In this paper we describe the small amount of existing literature relevant to our topic in Section 2. In Section 3 we motivate our working definition of hate speech. In Section 4 we describe the resources and corpora of hate and non-hate speech we have used in our experiments. In Section 5 we describe the annotation scheme we have developed and interlabeler reliability of the labeling process. In Section 6 we describe our approach to the classification problem and the features we used. We present preliminary results in Section 7, follow with an analysis of classification errors in 8 and conclude in Section 9 with an outline of further work.

[1] Yahoo TOS, paragraph 9a http://info.yahoo.com/legal/us/yahoo/utos/utos-173.html

[2] Facebook TOS, paragraph 3.7 https://www.facebook.com/legal/terms

Proceedings of the 2012 Workshop on Language in Social Media (LSM 2012), pages 19–26,
Montréal, Canada, June 7, 2012. ©2012 Association for Computational Linguistics

2 Previous Literature

There is little previous literature on identifying hate speech.

In (A Razavi, Diana Inkpen, Sasha Uritsky, Stan Matwin, 2010), the authors look for Internet “flames” in newsgroup messages using a three-stage classifier. The language of flames is significantly different from hate speech, but their method could inform our work. Their primary contribution is a dictionary of 2700 hand-labeled words and phrases.

In (Xu and Zhu, 2010), the authors look for offensive language in YouTube comments and replaces all but the first letter of each word with asterisks. Again, while the language and the goal is different, the method may have some value for detecting hate speech. Their detection method parses the text and arranges it into a hierarchy of clauses, phrases and individual words. Both the annotation and the classification strategies found in this paper are based on the sentiment analysis work found in (Pang and Lee, 2008) and (Pang, Lee and Vaithyanathan, 2002).

3 Defining Hate Speech

There are numerous issues involved in defining what constitutes hate speech, which need to be resolved in order to annotate a corpus and develop a consistent language model. First, merely mentioning, or even praising, an organization associated with hate crimes does not by itself constitute hate speech. The name “Ku Klux Klan” by itself is not hateful, as it may appear in historical articles, legal documents, or other legitimate communication. Even an endorsement of the organization does not constitute a verbal attack on another group. While one may hypothesize that such endorsements are made by authors who would also be comfortable with hateful language, by themselves, we do not consider these statements to be hate speech.

For the same reason, an author’s excessive pride in his own race or group doesn’t constitute hate speech. While such boasting may seem offensive and likely to co-occur with hateful language, a disparagement of others is required to satisfy the definition.

For example, the following sentence does not constitute hate speech, even though it uses the word “Aryan”.

> *And then Aryan pride will be true because humility will come easily to Aryans who will all by then have tasted death.*

On the other hand, we believe that unnecessary labeling of an individual as belonging to a group often should be categorized as hate speech. In the following example, hate is conveyed when the author unnecessarily modifies bankers and workers with “jew” and “white.”

> *The next new item is a bumper sticker that reads: “Jew Bankers Get Bailouts, White Workers Get Jewed!” These are only 10 cents each and require a minimum of a $5.00 order*

Unnecessarily calling attention to the race or ethnicity of an individual appears to be a way for an author to invoke a well known, disparaging stereotype.

While disparaging terms and racial epithets when used with the intent to harm always constitute hateful language, there are some contexts in which such terms are acceptable. For example, such words might be acceptable in a discussion of the words themselves. For example:

> *Kike is a word often used when trying to offend a jew.*

Sometimes such words are used by a speaker who belongs to the targeted group, and these may be hard to classify without that knowledge. For example:

> *Shit still happenin and no one is hearin about it, but niggas livin it everyday.*

African American authors appear to use the “N” word with a particular variant spelling, replacing “er” with “a”, to indicate group solidarity (Stephens-Davidowitz, 2011). Such uses must be distinguished from hate speech mentions. For our purposes, if the identity of the speaker cannot be ascertained, and if no orthographic or other contextual cues are present, such terms are categorized as hateful.

4 Resources and Corpora

We received data from Yahoo! and the American Jewish Congress (AJC) to conduct our research on hate speech. Yahoo! provided data from its news group posts that readers had found offensive. The AJC provided pointers to websites identified as offensive.

Through our partnership with the American Jewish Congress, we received a list of 452 URLs previously obtained from Josh Attenberg (Attenberg and Provost, 2010) which were originally collected to classify websites that advertisers might find unsuitable. After downloading and examining the text from these sites, we found a significant number that contained hate speech according to our working definition; in particular, a significant number were anti-semitic. We noted, however, that sites which which appeared to be anti-semitic rarely contained explicitly pejorative terms. Instead, they presented scientifically worded essays presenting extremely anti-semitic ideologies and conclusions. Some texts contained frequent references to a well known hate group, but did not themselves constitute examples of hate speech. There were also examples containing only defensive statements or declarations of pride, rather than attacks directed toward a specific group.

In addition to the data we collected from these URLs, Yahoo! provided us with several thousand comments from Yahoo! groups that had been flagged by readers as offensive, and subsequently purged by administrators. These comments are short, with an average of length of 31 words, and lacked the contextual setting in which they were originally found. Often, these purged comments contained one or more offensive words, but obscured with an intentional misspelling, presumably to evade a filter employed by the site. For common racial epithets, often a single character substitution was used, as in "nagger", or a homophone was employed, such as "joo." Often an expanded spelling was employed, in which each character was separated by a space or punctuation mark, so that "jew" would become "j@e@w@."

The two sources of data were quite different, but complementary.

The Yahoo! Comment data contained many examples of offensive language that was sometimes hateful and sometimes not, leading to our hypothesis that hate speech resembles a word sense disambiguation task, since, a single word may appear quite frequently in hate and non-speech texts. An example is the word "jew". In addition, it provided useful examples of techniques used to evade simple lexical filters (in case such exist for a particular forum). Such evasive behavior generally constitutes a positive indicator of offensive speech.

Web data captured from Attenberg's URLs tended to include longer texts, giving us more context, and contained additional lower frequency offensive terms. After examining this corpus, we decided to attempt our first classification experiments at the paragraph level, to make use of contextual features.

The data sets we received were considered offensive, but neither was labeled for hate speech per se. So we developed a labeling manual for annotating hate speech and asked annotators to label a corpus drawn from the web data set.

5 Corpus Collection and Annotation

We hypothesize that hate speech often employs well known stereotypes to disparage an individual or group. With that assumption, we may be further subdivide such speech by stereotype, and we can distinguish one form of hate speech from another by identifying the stereotype in the text. Each stereotype has a language all its own, with one-word epithets, phrases, concepts, metaphors and juxtapositions that convey hateful intent. Anti-hispanic speech might make reference to border crossing or legal identification. Anti-African American speech often references unemployment or single parent upbringing. And anti-semitic language often refers to money, banking and media.

Given this, we find that creating a language model for each stereotype is a necessary prerequisite for building a model for all hate speech. We decided to begin by building a classifier for anti-semitic speech, which is rich with references to well known stereotypes.

The use of stereotypes also means that some language may be regarded as hateful even though no single word in the passage is hateful by itself. Often there is a relationship between two or more sentences that show the hateful intent of the author.

Using the website data, we captured paragraphs that matched a general regular expression of words relating to Judaism and Israel [3]. This resulted in about 9,000 paragraphs. Of those, we rejected those that did not contain a complete sentence, contained more than two unicode characters in a row, were only one word long or longer than 64 words.

Next we identified seven categories to which labelers would assign each paragraph. Annotators could label a paragraph as anti-semitic, anti-black, anti-asian, anti-woman, anti-muslim, anti-immigrant or other-hate. These categories were designed for annotation along the anti-semitic/not anti-semitic axis, with the identification of other stereotypes capturing mutual information between anti-semitism and other hate speech. We were interested in the correlation of anti-semitism with other stereotypes. The categories we chose reflect the content we encountered in the paragraphs that matched the regular expression.

We created a simple interface to allow labelers to assign one or more of the seven labels to each paragraph. We instructed the labelers to lump together South Asia, Southeast Asia, China and the rest of Asia into the category of anti-asian. The anti-immigrant category was used to label xenophobic speech in Europe and the United States. Other-hate was most often used for anti-gay and anti-white speech, whose frequency did not warrant categories of their own.

5.1 Interlabeler Agreement and Labeling Quality

We examined interlabeler agreement only for the anti-semitic vs. other distinction. We had a set of 1000 paragraphs labeled by three different annotators. The Fleiss kappa interlabeler agreement for anti-semitic paragraphs vs. other was 0.63. We created two corpora from this same set of 1000 paragraphs. First, the *majority* corpus was generated from the three labeled sets by selecting the label with on which the majority agreed. Upon examining this corpus with the annotators, we found some cases in which annotators had agreed upon labels that seemed inconsistent with their other annotations – often they had missed instances of hate speech which they subsequently felt were clear cases. One of the authors checked and corrected these apparent “errors” in annotator labeling to create a *gold* corpus. Results for both the original majority class annotations and the “gold” annotations are presented in Section 7.

As a way of gauging the performance of human annotators, we compared two of the annotators' labels to the gold corpus by treating their labeled paragraphs as input to a two fold cross validation of the classifier constructed from the gold corpus. We computed a precision of 59% and recall of 68% for the two annotators. This sets an upper bound on the performance we should expect from a classifier.

6 Classification Approach

We used the template-based strategy presented in (Yarowsky, 1994) to generate features from the corpus. Each template was centered around a single word as shown in Table 1. Literal words in an ordered two word window on either side of a given word were used exactly as described in (Yarowsky, 1994). In addition, a part-of-speech tagging of each sentence provided the similar part-of-speech windows as features. Brown clusters as described in (Koo, Carreras and Collins, 2008) were also utilized in the same window. We also used the occurrence of words in a ten word window. Finally, we associated each word with the other labels that might have been applied to the paragraph, so that if a paragraph containing the word “god” were labeled “other-hate”, a feature would be generated associating “god” with other-hate: “RES:other-hate W+0:god”.

We adapted the hate-speech problem to the problem of word sense disambiguation. We say that words have a *stereotype sense*, in that they either anti-semitic or not, and we can learn the sense of all words in the corpus from the paragraph labels. We used a process similar to the one Yarowsky described when he constructed his decisions lists, but we expand the feature set. What is termed log-likelihood in (Yarowsky, 1994) we will call log-odds, and it is calculated in the following way. All templates were generated for every paragraph in the corpus, and a count of positive and negative occurrences for each template was maintained. The ab-

[3] `jewish|jew|zionist|holocaust|denier|rabbi|israel|semitic|semite`

solute value of the ratio of positive to negative occurrences yielded the log-odds. Because log-odds is based on a ratio, templates that do not occur at least once as both positive and negative are discarded. A feature is comprised of the template, its log-odds, and its sense. This process produced 4379 features.

Next, we fed these features to an SVM classifier. In this model, each feature is dimension in a feature vector. We treated the sense as a sign, 1 for anti-semitic and -1 otherwise, and the weight of each feature was the log-odds times the sense. The task of classification is sensitive to weights that are large relative to other weights in the feature space. To address this, we eliminated the features whose log-odds fell below a threshold of 1.5. The resulting values passed to the SVM ranged from -3.99 to -1.5 and from +1.5 to +3.2. To find the threshold, we generated 40 models over an evenly distributed range of thresholds and selected the value that optimized the model's f-measure using leave-1-out validation. We conducted this procedure for two sets of independent data and in both cases ended up with a log-odds threshold of 1.5. After the elimination process, we were left with 3537 features.

The most significant negative feature was the unigram literal "black,", with log-odds 3.99.

The most significant positive feature was the part-of-speech trigram "DT jewish NN", or a determiner followed by jewish followed by a noun. It was assigned a log-odds of 3.22.

In an attempt to avoid setting a threshold, we also experimented with binary features, assigning -1 to negative feature weights and +1 to positive feature weights, but this had little effect, and are not recorded in this paper. Similarly, adjusting the SVM soft margin parameter C had no effect.

We also created two additional feature sets. The *all unigram* set contains only templates that are comprised of a single word literal. This set contained 272 features, and the most significant remained "black." The most significant anti-semitic feature of this set was "television," with a log-odds of 2.28. In the corpus we developed, television figures prominently in conspiracy theories our labelers found anti-semitic.

The *positive unigram* set contained only unigram templates with a positive (indicating anti-semitism) log-odds. This set contained only 13 features, and the most significant remained "television."

7 Preliminary Results

7.1 Baseline Accuracy

We established a baseline by computing the accuracy of always assuming the majority (not anti-semitic) classification. If N is the number of samples and N_p is the number of positive (anti-semitic) samples, accuracy is given by $(N - N_p)/N$, which yielded a baseline accuracy of 0.910.

7.2 Classifiers

For each of the majority and gold corpora, we generated a model for each type of feature template strategy, resulting in six classifiers. We used SVM^{light} (Joachims, 1999) with a linear kernel function. We performed 10 fold cross validation for each classifier and recorded the results in Table 2. As expected, our results on the majority corpus were not as accurate as those on the gold corpus. Perhaps surprising is that unigram feature sets out performed the full set, with the smallest feature set, comprised of only positive unigrams, performing the best.

8 Error Analysis

Table 3 contains a summary of errors made by all the classifiers. For each classifier, the table reports the two kinds of errors a binary classifier can make: false negatives (which drive down recall), and false positives (which drive down precision).

The following paragraph is clearly anti-semitic, and all three annotators agreed. Since the classifier failed to detect the anti-semitism, we use look at this example of a false negative for hints to improve recall.

> *4. That the zionists and their american sympathizers, in and out of the american media and motion picture industry, who constantly use the figure of "six million" have failed to offer even a shred of evidence to prove their charge.*

Table 1: Example Feature Templates

unigram	”W+0:america”
template literal	”W-1:you W+0:know”
template literal	”W-1:go W+0:back W+1:to”
template part of speech	”POS-1:DT W+0:age POS+1:IN”
template Brown sub-path	”W+0:karma BRO+1:0x3fc00:0x9c00 BRO+2:0x3fc00:0x13000”
occurs in ±10 word window	”WIN10:lost W+0:war”
other labels	”RES:anti-muslim W+0:jokes”

Table 2: Classification Performance

	Accuracy	Precision	Recall	F1
Majority All Unigram	0.94	0.00	0.00	0.00
Majority Positive Unigram	0.94	0.67	0.07	0.12
Majority Full Classifier	0.94	0.45	0.08	0.14
Gold All Unigram	0.94	0.71	0.51	0.59
Gold Positive Unigram	0.94	0.68	0.60	0.63
Gold Full Classifier	0.93	0.67	0.36	0.47
Human Annotators	0.96	0.59	0.68	0.63

Table 3: Error Report

	False Negative	False Positive
Majority All Unigram	6.0%	0.1%
Majority Positive Unigram	5.6%	0.2%
Majority Full Classifier	5.5%	0.6%
Gold All Unigram	4.4%	1.8%
Gold Positive Unigram	3.6%	2.5%
Gold Full Classifier	5.7%	1.6%

The linguistic features that clearly flag this paragraph as anti-semitic are the noun phrase containing *zionist ... sympathizers*, the gratuitous inclusion of *media and motion picture industry* and the skepticism indicated by quoting the phrase *"six million"*. It is possible that the first feature could have been detected by adding parts of speech and Brown Cluster paths to the 10 word occurrence window. A method for detecting redundancy might also be employed to detect the second feature. Recent work on emotional speech might be used to detect the third.

The following paragraph is more ambiguous. The annotator knew that GT stood for gentile, which left the impression of an intentional misspelling. With the word spelled out, the sentence might not be anti-semitic.

> *18) A jew and a GT mustn't be buried side by side.*

Specialized knowledge of stereotypical language and the various ways that its authors mask it could make a classifier's performance superior to that of the average human reader.

The following sentence was labeled negative by annotators but the classifier predicted an anti-semitic label.

> *What do knowledgeable jews say?*

This false positive is nothing more than a case of over fitting. Accumulating more data containing the word "jews" in the absence of anti-semitism would fix this problem.

9 Conclusions and Future Work

Using the feature templates described by Yarowsky we successfully modeled hate speech as a classification problem. In terms of f-measure, our best classifier equaled the performance of our volunteer annotators. However, bigram and trigram templates degraded the performance of the classifier. The learning phase of the classifier is sensitive to features that ought to cancel each other out. Further research on classification methods, parameter selection and optimal kernel functions for our data is necessary.

Our definition of the labeling problem could have been more clearly stated to our annotators. The anti-immigrant category in particular may have confused some.

The recall of the system is low. This suggests there are larger linguistic patterns that our shallow parses cannot detect. A deeper parse and an analysis of the resulting tree might reveal significant phrase patterns. Looking for patterns of emotional speech, as in (Lipscombe, Venditti and Hirschberg, 2003) could also improve our recall.

The order of the paragraphs in their original context could be used as input into a latent variable learning model. McDonald (McDonald et al, 2007) has reported some success mixing fine and course labeling in sentiment analysis.

Acknowledgments

The authors are grateful for the data and the support of Matthew Holtzman of the American Jewish Congress. We would also like to thank Belle Tseng, Kim Capps-Tanaka, Evgeniy Gabrilovich and Martin Zinkevich of Yahoo! for providing data as well as for their financial support. Without their support, this research would not have been possible.

References

[Choi et al 2005] Yejin Choi, Claire Cardie, Ellen Riloff, Siddharth Patwardhan, *Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns*. In *HLT '05* Association for Computational Linguistics Stroudsburg, PA, USA, pp. 355-362, 2005

[Yarowsky 1994] David Yarowsky, *Decision Lists for Lexical Ambiguity Resolution: Application to Accent Restoration in Spanish and French*. In *ACL-94*, Stroudsburg, PA, pp. 88-95, 1994

[Yarowsky 1995] David Yarowsky, *Unsupervised Word Sense Disambiguation Rivaling Supervised Methods*. In *ACL-95*, Cambridge, MA, pp. 189-196, 1995.

[Nockleby 2000] John T. Nockleby, *Hate Speech*. In *Encyclopedia of the American Constitution* (2nd ed., edited by Leonard W. Levy, Kenneth L. Karst et al., New York: Macmillan, 2000), pp. 1277-1279 (see `http://www.jiffynotes.com/a_study_guides/book_notes/eamc_03/eamc_03_01193.html`)

[Stephens-Davidowitz 2011] Seth Stephens-Davidowitz, *The Effects of Racial Animus on Voting: Evidence Using Google Search Data* `http://www.people.fas.harvard.edu/~sstephen/papers/RacialAnimusAndVotingSethStephensDavidowitz.pdf`

[McDonald et al 2007] McDonald, R. Hannan, K. Neylon, T. Wells, M. Reynar, J. *Structured Models for Fine-to-Coarse Sentiment Analysis*. In *ANNUAL MEETING- ASSOCIATION FOR COMPUTATIONAL LINGUISTICS* 2007, CONF 45; VOL 1, pages 432-439

[Pang and Lee 2008] Pang, Bo and Lee, Lillian, *Opinion Mining and Sentiment Analysis*. In *Foundations and Trends in Information Retrieval*, issue 1-2, vol. 2, Now Publishers Inc., Hanover, MA, USA, 2008 pp. 1–135

[Pang, Lee and Vaithyanathan 2002] Pang, Bo and Lee, Lillian and Vaithyanathan, Shivakumar *Thumbs up?: sentiment classification using machine learning techniques*. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, Association for Computational Linguistics, Stroudsburg, PA, USA, 2002 pp. 79-86

[Qiu et al 2009] Qiu, Guang and Liu, Bing and Bu, Jiajun and Chen, Chun *Expanding domain sentiment lexicon through double propagation*. In *Proceedings of the 21st international jont conference on Artificial intelligence*, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 2009 pp. 1199-1204

[Joachims 1999] *Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning*, B. Schlkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999.

[Koo, Carreras and Collins 2008] *Simple Semi-supervised Dependency Parsing* In *Proc. ACL/HLT* 2008

[Xu and Zhu 2010] *Filtering Offensive Language in Online Communities using Grammatical Relations*

[A Razavi, Diana Inkpen, Sasha Uritsky, Stan Matwin 2010] *Offensive Language Detection Using Multi-level Classification* In *Advances in Artificial Intelligence* Springer, 2010, pp. 1627

[Attenberg and Provost 2010] *Why Label When You Can Search?: Alternatives to active learning for applying human resources to build classification models under extreme class imbalance,* KDD 2010

[Lipscombe, Venditti and Hirschberg 2003] *Classifying Subject Ratings of Emotional Speech Using Acoustic Features.* In *Proceedings of Eurospeech* 2003, Geneva.

ICCT International Centre for Counter-Terrorism - The Hague

Blind Spot? Security Narratives and Far-Right Violence in Europe

Dr. Arun Kundnani

ICCT Research Paper
June 2012

Abstract

This paper discusses the challenges of countering far-Right political violence in the wake of the terrorist attack carried out by Anders Behring Breivik in Norway in July 2011. With brief case studies of Britain, the Netherlands, Denmark and Belgium, it argues that classic neo-Nazi groups are being supplemented by new 'counter-jihadist' far-Right movements, which use various modes of political action, including participation in elections, street-based activism and terrorist violence. Building on recent interest among scholars and practitioners in the role of narratives and performativity in counter-terrorism, this paper argues that official security discourses tend to hinder efforts to counter far-Right violence and can unwittingly provide opportunities for counter-jihadists to advance their own narratives. When leaders and officials of Western European governments narrate issues of multiculturalism and radical Islamism in ways that overlap with counter-jihadist ideology, it suggests a need for reflection on the unintended side-effects of their security discourse. The paper concludes with a discussion of how governments can rework their security narratives to oppose far-Right violence.

About the Author

Dr. Arun Kundnani was a Visiting Research Fellow at the International Centre for Counter-Terrorism – The Hague from March to May 2012. He is a former fellow at the Open Society Foundations, New York, and has written widely on race relations, multiculturalism and security. He is currently writing a book on the politics of anti-extremism in the US and the UK. His study of the UK government's Preventing Violent Extremism programme, Spooked: how not to prevent violent extremism (Institute of Race Relations, 2009), led to widespread recognition that the policy had counter-productively blurred surveillance and community engagement. He is the author of The End of Tolerance: racism in 21st century Britain (Pluto Press, 2007) and a former editor of the journal Race & Class. He holds an MA in philosophy from Cambridge University and a PhD in social science from London Metropolitan University. He can be contacted at: arun@kundnani.org.

About ICCT - The Hague

The International Centre for Counter-Terrorism (ICCT) – The Hague is an independent knowledge centre that focuses on information creation, collation and dissemination pertaining to the preventative and international legal aspects of counter-terrorism. The core of ICCT's work centres on such themes as de- and counter-radicalisation, human rights, impunity, the rule of law and communication in relation to counter-terrorism. Functioning as a nucleus within the international counter-terrorism network, ICCT – The Hague endeavours to connect academics, policymakers and practitioners by providing a platform for productive collaboration, practical research, exchange of expertise and analysis of relevant scholarly findings. By connecting the knowledge of experts to the issues that policymakers are confronted with, ICCT – The Hague contributes to the strengthening of both research and policy. Consequently, avenues to new and innovative solutions are identified, which will reinforce both human rights and security.

Contact

ICCT – The Hague
Koningin Julianaplein 10
P.O. Box 13228
2501 EE, The Hague
The Netherlands

T +31 (0)70 800 9531
E info@icct.nl

All papers can be downloaded free of charge at www.icct.nl
Stay up to date with ICCT, follow us online on Facebook, Twitter and LinkedIn

Introduction

On 22 July 2011, as news emerged of a major terrorist attack taking place in Norway, the *Wall Street Journal* went to press while the identity of the perpetrator was still unknown. On the presumption that only a Muslim could be responsible, the newspaper's editorial claimed that Norway had been targeted because it is 'a liberal nation committed to freedom of speech and conscience, equality between the sexes, representative democracy and every other freedom that still defines the West'.[1] The reflexes entrenched by nearly ten years of 'war on terror' rhetoric led the editorial writer to feel confident the attacker's motivation could already be known.

As it turned out, the worst terrorist attack in Europe since the Madrid bombings of 2004 – a car bomb in Oslo, followed by a shooting spree on the island of Utøya, leaving 77 dead – had been carried out in the name of a 'counter-jihadist' rather than jihadist ideology. Anders Behring Breivik, whose 1,500-page manifesto, *2083 – A European Declaration of Independence*, was published online on the day of the attacks, believed that European elites were pandering to multiculturalism and enabling an 'Islamic colonisation of Europe'. Like the *Wall Street Journal* editorial writer, he believed that Norway's liberal values were under threat from 'radical Islam'.

The newspaper's error was an extreme case of a much wider problem not only among journalists but also within the world of counter-terrorism policy-making and practice – that the ways in which counter-terrorism is narrated leads to a focus on particular threats, conceived in particular ways, while potentially neglecting others, and that in doing so openings may be provided for unexpected new forms of violence. While a number of scholars have begun to explore how the 'war on terror' meta-narrative is itself appropriated by jihadists to win new recruits, there has been no exploration of similar effects in enabling far-Right violence. Yet there is strong prima facie evidence that new far-Right 'counter-jihadist' movements use the meta-narrative of a global struggle against 'radical Islam' to legitimise themselves to their audiences. If this is the case, what strategies can be adopted by governments that wish to reverse this effect? How can security policies be developed and communicated to oppose both jihadist and far-Right violence?

Official security narratives in Europe are multiple, complex and contested. It is beyond the scope of this paper to fully explore the plurality of this terrain. However, across a number of contexts since 2005, a shared way of thinking has emerged that underlies various different positions taken by government ministers, security officials and other commentators. This way of thinking is reflected in a **values-identity narrative of terrorism**, which establishes fundamental ideas about how to identify the 'we' that is countering terrorism, the 'them' that is engaged in terrorism, and the terms on which the conflict between 'us' and 'them' is to be understood. There has been a debate since 9/11 as to whether terrorism is related to Islam or to a particular misinterpretation of Islam. But both sides in this debate agree that the substantial problem is a conflict between liberal values and religiously-inspired 'extremist' values and that the question of Muslim 'integration' into 'our' values is therefore part of the security picture. Thus, the official narrative, in whatever particular form it takes, tends to focus on the question of values within Muslim populations, and claims that counter-terrorism requires initiatives aimed at a wider transformation of the Muslim population beyond those who actually perpetrate or advocate violence. The case studies in this paper will examine the prevalence of this narrative in counter-terrorism thinking in four countries: Britain, the Netherlands, Denmark and Belgium.

How these narratives are received by different audiences is again not a straightforward matter. In this paper, two audiences for official security narratives are discussed. The first is those professionals working within the counter-terrorist systems themselves, which, since 2004, include not just police and intelligence officers investigating terrorism cases but teachers, youth workers, social workers, community activists, local authority managers and civil society groups who have been drawn into the counter-terrorist project as actors with a

[1] 'Terror in Oslo', *Wall Street Journal* (22 July 2011).

preventative role. It will be argued that, for this group, recognition of the threat of far-Right violence has often been hampered by its lack of fit with the prevailing values-identity narrative of terrorism – would-be perpetrators were seen as one of 'us' rather than 'them'. In addition, there is the simple matter of resource allocation and prioritisation in counter-terrorism practice. Unlike jihadist terrorism, far-Right violence is generally not seen by European security officials as a strategic threat, only as a public order problem. For example, in its 2011 *EU Terrorism Situation and Trend Report*, Europol states that right-wing extremist incidents 'raised public order concerns, but have not in any way endangered the political, constitutional, economic or social structures of any of the Member States. They can, however, present considerable challenges to policing and seriously threaten community cohesion.'[2] Shortly after the Breivik terrorist attack in Norway, it emerged that a German neo-Nazi group – the *Nationalsozialistischer Untergrund* (NSU, National Socialist Underground) – had operated for thirteen years without arrest, during which time eight people of Turkish origin, a Greek man, and a policewoman had been killed, despite federal and regional intelligence services reportedly having infiltrated the group. It remains unclear why the NSU was not intercepted earlier. However it appears that part of the problem was that efforts to counter right-wing violence rested with regional states, which did not consider it a priority, in contrast to initiatives to counter the threat of jihadist violence, which were well-resourced and centrally co-ordinated at the federal level.

The second audience for official security discourse is the far-Right milieu itself. This paper proposes the thesis that, in some contexts, the circulation of the values-identity narrative of terrorism has the unintended consequence of creating discursive opportunities for far-Right actors who are able to blend official narratives into their own discourses, enabling them to creatively update their existing belief systems and draw renewed legitimacy by bringing their ideologies into closer proximity to mainstream views. Breivik is one example. His manifesto makes clear that he believes Islam to be a totalitarian political ideology that aims at infiltrating national institutions to impose sharia law on Muslims and non-Muslims, and that this process of 'Islamisation' has been enabled by elites in Western countries, through their weakening of immigration controls and introduction of multiculturalist policies – views that, as we shall see, have significant overlaps with official discourse. He hoped his violence would 'penetrate the strict censorship regime' of 'cosmopolitan' elites, so that European citizens would see the need to defend their liberal values against multiculturalism.[3] Others, such as the English Defence League (EDL), share the same definition of the 'problem' but employ different tactics, favouring demonstrations and street-based activism, often involving public disorder, racist violence and incitements to anti-Muslim hatred. Both examples demonstrate how the ideological basis for far-Right violence has grown more complex, as new actors appropriate narrative elements from official security discourses in innovative ways, potentially making far-Right threats harder to identify.

Every perception has a blind spot, the area that cannot be seen because it is part of the mechanism of perception itself. This paper considers whether, since 9/11, the far-Right has been the blind spot of counter-terrorism, the problem that could not be perceived clearly because it had begun to absorb significant elements from official security narratives themselves. After the Breivik case, it has become harder to believe that these unintended consequences can be ignored. Moreover, as demonstrated in Annex 1, if the level of threat is measured in terms of the number of people who have lost their lives as a result of far-Right violence, it is incorrect to see jihadism as representing a greater danger to European citizens. Since 1990, at least **249** persons have died in incidents of far-Right violence in Europe, compared to **263** who have been killed by jihadist violence, indicating that both threats are of the same order of magnitude.[4] That both these numbers are tiny relative to the

[2] Europol, *EU Terrorism Situation and Trend Report* (2011), p. 29.

[3] Anders Behring Breivik, *2083 – A European Declaration of Independence* (2011), p. 822.

[4] The total for jihadists is based on adding the murder of 'Abd al-Baqi Sahraoui in Paris in 1995, the killing of 8 persons in the bomb attack on the Paris metro in 1995 (although it remains disputed on whose behalf this attack was carried out), the murder of Stephen Oake in Manchester in 2003, the 191 persons killed in Madrid in 2004, the murder of Theo Van Gogh in Amsterdam in 2004, 52 persons killed in the 7/7 attack in London in 2005, two US servicemen killed at Frankfurt Airport in 2011 and 7 persons killed in Toulouse in 2012. This list was compiled with the help of the chronology in Petter Nesser, 'Chronology of Jihadism in Western Europe 1994–2007: planned, prepared, and executed terrorist attacks', *Studies in Conflict & Terrorism* (Vol. 31, 2008).

population of Europe suggests perceptions of the threat of jihadist violence have been over-inflated and should now be brought down to the same level as that of far-Right violence.

1. Conceptualising far-Right violence

1.1. Current trends in far-Right ideology

In April 2012, Kenny Holden, of South Shields in north-east England, used his mobile phone to post on Facebook a threat to carry out a pipe-bomb attack on the town's Ocean Road, a street with a number of Asian convenience stores and restaurants. His post included the line, 'Give me a gun and I'll do you all Oslo style', a reference to Anders Behring Breivik, who was then on trial in Norway.[5] Breivik no doubt hoped to use the trial to further publicise his cause and Holden's threat suggested an audience among some in England.

Holden is an activist with the EDL, an organisation formed in Luton, Bedfordshire, in 2009, ostensibly to combat Islamist 'extremism'. In March 2009, Anjem Choudary – the leader of a small radical Islamist group that has had various names since its original incarnation, al-Muhajiroun, was disbanded in 2004 – organised a protest against a parade through Luton town centre of British troops recently returned from Afghanistan. There was a furious reaction from bystanders; a coalition of angry locals, members of football 'firms' and seasoned far-Right activists came together and soon grew to form the EDL. Making good use of the online and offline networks that already linked football firms and the far-Right across the country, and picking up a significant number of young people who seemed to relate, via Facebook and YouTube, to its style of politics, the EDL was soon organising demonstrations in several towns and cities, attracting up to 2,000 people.

The 'counter-jihadist' ideology of the EDL differs markedly from the traditional far-Right. Its two main targets are Islam, which it regards as an extremist political ideology, and multiculturalism, which is presented as enabling 'Islamification'. Rhetorically, the EDL embraces values of individual liberty, freedom of speech, gender equality and gay rights, and rejects colour-based forms of racism and anti-Semitism, in favour of a civilisational discourse – talking of defending Western values rather than the white race.[6] Indeed, fairly wide sections of the English population which would have rejected neo-Nazism and overt colour-based racism are nevertheless supportive of this discourse.

The EDL has become Europe's most significant 'counter-jihadist' street movement, inspiring copycat Defence Leagues in a number of other countries and prompting an attempt to form a European Defence League, launched, somewhat feebly, at Aarhus, Denmark, in March 2012. Since its formation in 2009, there have been Nazi salutes, racist chants and incidents of racial violence at EDL demonstrations.[7] Activism for the EDL overlaps significantly with membership of the racist British National Party (BNP). Indeed, both of the EDL's senior leaders, Stephen Yaxley-Lennon (aka Tommy Robinson) and his cousin Kevin Carroll, are former members of the BNP and have been convicted of criminal violence. Members of the 'West Midlands Division' of the EDL have taken photographs of themselves standing in front of Ulster Volunteer Force flags, carrying imitation firearms.

At a demonstration on 3 September 2011 through the largely Muslim area of Tower Hamlets, east London (a favourite location for far-Right mobilisation since the 'Battle of Cable Street' in 1936), Yaxley-Lennon told the crowd:

> 'We are here today to tell you, quite loud, quite clear, every single Muslim watching this video on YouTube: on 7/7, you got away with killing and maiming British citizens. You got away with it. You better understand that we have built a network from one end of this country to the other end.

[5] Paul Clifford, 'Shields man arrested over Facebook "threat" to bomb town's Muslims', *Shields Gazette* (23 April 2012), http://www.shieldsgazette.com/news/crime/shields-man-arrested-over-facebook-threat-to-bomb-town-s-muslims-1-4475638.

[6] English Defence League mission statement, http://englishdefenceleague.org.

[7] Ryan Erfani-Ghettani, 'From portrayal to reality: examining the record of the EDL', *IRR News* (8 December 2011), http://www.irr.org.uk/news/from-portrayal-to-reality-examining-the-record-of-the-edl/.

> We will not tolerate it. And the Islamic community will feel the full force of the English Defence League if we see any of our citizens killed, maimed or hurt on British soil ever again.'[8]

This threat of violence against all Muslims in Britain used the same implicit logic as that of his purported enemies. Mohammed Siddique Khan, leader of the suicide bombers who carried out the 7/7 attacks on London's transport system, told the British people: 'Until we feel security, you will be our targets.' In this way, both Khan and Yaxley-Lennon seek to justify violence against a whole community which they hold collectively responsible for the violence of some of its members.

In Breivik's manifesto, *2083 – A Declaration of Independence*, he praises the EDL for being the first youth movement to transcend the old-fashioned race hate and authoritarianism of the far-Right in favour of an identitarian defence of Western values against Islam. He urges 'conservative intellectuals' to help ensure the EDL continues to reject 'criminal, racist and totalitarian doctrines'. But he also considers their faith in democratic change 'dangerously naïve'.[9] Breivik claimed to have hundreds of EDL members as Facebook friends and there has been speculation over his links to members of the organisation.

Certainly, his manifesto shares with the EDL a counter-jihadist ideology. Claiming to be a member of a secret group of new 'crusaders' founded in London in 2002 by representatives from eight European countries, he says his aim is to 'free indigenous peoples of Europe and to fight against the ongoing European Jihad'.[10] Much of the document consists of advice to fellow far-Right terrorists on weapons, bomb-making, body armour, physical training and the potential use of chemical, biological and nuclear weapons. One section of *2083* describes the ranks, organisational structure, initiation rites, uniforms, awards and medals to be used by this secret 'Knights Templar' group. These parts of the manifesto – and a section in which he interviews himself, narcissistically listing his favourite music, clothes and drinks – appear to be its only original content.

The bulk of the document constitutes a compilation of texts mainly copied from US far-Right websites. Its opening chapters are plagiarised from *Political Correctness: a short history of an ideology*, a book published online in 2004 by the Free Congress Foundation – a Washington-based lobby group founded by Paul Weyrich, one of the most influential activists of the US Christian Right. In this section, Breivik has replaced references to 'America' in the original text with 'Western Europe'. Apart from this, the writers Breivik cites most often are prominent counter-jihadists: Robert Spencer, the American Islamophobic blogger whose *Jihad Watch* website, a subsidiary of the David Horowitz Freedom Center, receives close to a million dollars of funding from wealthy backers;[11] Ba'et Yor, the Swiss author of the 'Eurabia' conspiracy theory, which claims that the European political establishment is involved in a secret plot to facilitate Muslim immigration and transform the continent into an Arab colony;[12] and 'Fjordman', a Norwegian, who blogs for the US-based *Gates of Vienna* and *Jihad Watch* websites. These writers are paranoid conspiracy theorists who claim Islam is a totalitarian political ideology that aims at infiltrating national institutions in order to subject society to sharia law. Like Breivik, they blame Western elites for pandering to multiculturalism and enabling 'Islamic colonisation of Europe' through 'demographic warfare'.

This multiculturalist elite, says Breivik, has prevented the possibility of democratic opposition and the clock is ticking: 'We have only a few decades to consolidate a sufficient level of resistance before our major cities are completely demographically overwhelmed by Muslims.'[13] Hence, he justifies his violence as 'a pre-emptive war'.[14] In a 2007 blog post by Fjordman, entitled 'A European Declaration of Independence', which Breivik reproduces and whose title he borrows, Fjordman writes: 'We are being subject to a foreign invasion, and aiding

[8] http://www.youtube.com/watch?v=jAGnzu5V3zE.
[9] Anders Behring Breivik, *2083 – A European Declaration of Independence* (2011), pp. 1241, 1436.
[10] Ibid., p. 817.
[11] Wajahat Ali, Eli Clifton, Matthew Duss, Lee Fang, Scott Keyes, and Faiz Shakir, *Fear, Inc.: the roots of the Islamophobia network in America* (Center for American Progress, August 2011).
[12] Matt Carr, 'You are now entering Eurabia', *Race & Class* (Vol. 48, no. 1, 2006).
[13] Anders Behring Breivik, *2083 – A European Declaration of Independence* (2011), p. 9.
[14] Ibid., p. 766.

and abetting a foreign invasion in any way constitutes treason. If non-Europeans have the right to resist colonisation and desire self-determination then Europeans have that right, too. And we intend to exercise it.'[15]

In the conventional neo-Nazi doctrine of 'race war', whites are called upon to rise up against governments seen as secretly controlled by Jews and whose aim is to dilute white racial purity by enabling black immigration. Breivik reframes this doctrine by substituting culture for race, Muslims for blacks, and multiculturalists for Jews. Rejecting the 'race war' concept, he calls instead for a 'cultural war' in which 'absolutely everyone will have the opportunity to show their loyalty to our cause, including nationalist European Jews, non-European Christians or Hindu/Buddhist Asians'.[16] Yet he also speaks of his 'opposition to race-mixing' and wants to prevent the 'extinction of the Nordic genotypes'.[17] Of Jews, he writes that 'we must embrace the remaining loyal Jews as brothers' and that there is no 'Jewish problem in Western Europe' as their numbers are small. Yet he goes on to say that the UK, France and US do have a 'considerable Jewish problem'.[18] Casting Jews as both potential allies (if they join in fighting Islam) and a demographic threat (if there are too many), Breivik is simultaneously anti-Semitic and supportive of right-wing Zionism.

The Breivik case demonstrates that the new counter-jihadist far-Right is as compatible with terrorist violence as older forms of neo-Nazism. And, whereas neo-Nazism is a fringe phenomenon, many elements of the counter-jihadist ideology attract wide support, including among mainstream politicians, newspaper columnists and well-funded think-tanks. The major theme of Breivik's manifesto is the argument that multiculturalism has weakened national identity and encouraged Islamist 'extremism', bringing European nations to a crisis point. As Breivik himself correctly noted in the first week of his trial, this view is held by 'the three most powerful politicians in Europe' – Nicolas Sarkozy, Angela Merkel and David Cameron.[19] The uncomfortable truth is that the central plank of a terrorist's narrative is shared by heads of Western European governments – an unprecedented situation that begs the central questions of this paper: Has mainstream security discourse created an enabling environment for counter-jihadist violence? And does the proximity of the counter-jihadist narrative to views that are acceptable in mainstream discourse make it harder to identify persons willing to carry out violence in its name?

1.2. Historical background

The content of far-right narratives has continually evolved since the end of the second world war, while maintaining a consistent formal structure. The post-war neo-fascism of groups such as Britain's National Front (NF) was never just a matter of hating minorities. It was also an ideology that sought to explain and exploit social dislocation felt by working classes, through a rival narrative to that of the Left. To achieve this, it presented non-white immigration as corrupting the racial purity of the nation; but it paid equal attention to the ruling elite that had allowed this to happen, a betrayal which far-Right ideology explained in terms of a Jewish conspiracy theory. What appeared to be a national ruling class was, according to this narrative, a mirage; real power lay with the secret Jewish cabal that pulled the strings of international finance, the media and the revolutionary Left, as supposedly revealed in *The Protocols of the Learned Elders of Zion*, the forged Tsarist document purporting to show how Jews manipulated world events to their advantage.[20]

While far-Right street activism involved racist violence against non-whites, far-Right ideology saw the real problem as lying elsewhere: the Jews and their hidden agenda of destroying national identity by fostering the immigration and mixing of other races. As David Edgar put it in his 1977 analysis of the politics of Britain's NF, the far-Right 'blames the Jews for the blacks'.[21] Even as popular racism against Asian, Middle Eastern and African

[15] Fjordman, 'Native revolt: a European declaration of independence ', *The Brussels Journal* (16 March 2007), http://www.brusselsjournal.com/node/1980.
[16] Anders Behring Breivik, *2083 – A European Declaration of Independence* (2011), p. 1259.
[17] Ibid., pp. 1161, 1190.
[18] Ibid., p. 1163.
[19] Helen Pidd, 'Anders Behring Breivik claims victims were not innocent', *Guardian* (17 April 2012).
[20] Norman Cohn, *Warrant for Genocide: the myth of the Jewish world conspiracy and the Protocols of the Elders of Zion* (Harmondsworth, Penguin, 1970).
[21] David Edgar, 'Racism, fascism and the politics of the National Front', *Race & Class* (Vol. 19, no. 2, 1977).

immigrant communities was the means by which young people were recruited, anti-Semitism remained a necessary ideological component, because only Jews could play the role of the secret source of economic and political power that had weakened and corrupted the nation. To this extent, post-war fascist parties such as the NF were correctly described as Nazi in their ideology. For the same reason, the far-Right, with an ideology that had been completely discredited by its association with the holocaust, struggled to advance in the post-war period.

However, from the 1980s, the French *Front National* (FN) began to achieve a higher level of support by downplaying its neo-Nazi legacy and speaking of the need to preserve cultural identity, defined as an unchanging national 'way of life', rather than in overtly racial terms. In this 'New Right' narrative, identity was seen to be under threat from a ruling elite that enabled excessive immigration of persons with different cultures and that promoted policies of multiculturalism, giving immigrants licence to maintain their own cultural identities. The FN argued that this process of *mondialisation* (globalisation) was being imposed by an 'all-powerful oligarchy', which, with the end of communism, was advancing a new utopianism: instead of a 'red paradise', the aim was a 'society without differences … a *café au lait* paradise … a melting pot'.[22] Thus, instead of explicit talk of a Jewish conspiracy, there was the idea that those in power were too 'cosmopolitan' to have the real interests of the native people at heart. This message resonated effectively with many voters and soon other far-Right parties in Europe began to emulate the FN strategy. The success of this approach largely depended on the extent to which far-Right parties could convincingly distance themselves from their neo-Nazi pasts. Thus, in the 1990s, an analyst of trends in far-Right politics distinguished two different kinds of groups:

> 'One type is nostalgic, backward-looking neo-fascist aggregations, parties whose raison d'être is a revival of fascist or Nazi ideas. … The other type consists of a class of parties described as right-wing populist – such as France's National Front, the Austrian Freedom Party, Flemish Bloc in Belgium, and the various Scandinavian Progress parties – that have done relatively well at the polls.'[23]

Following 9/11, a new version of this identitarian narrative began to circulate, first in the Netherlands and later in other European countries, often promoted by new political actors without the usual neo-Nazi baggage. In the 'counter-jihadist' narrative, the identity that needs to be defended is no longer a conservative notion of national identity but an idea of liberal values, seen as a civilisational inheritance. Islam becomes the new threat to this identity, regarded as both an alien culture and an extremist political ideology. Multiculturalism is seen as enabling not just the weakening of national identity but 'Islamification', a process of colonisation leading to the rule of sharia law. European governments are regarded as weak and complicit in the face of this totalitarian threat. Old-style racism, anti-Semitism and authoritarianism are rejected; right-wing Zionism is taken to be a potential ally. Unlike the traditional far-Right, these new movements rhetorically embrace what they regard as Enlightenment values of individual liberty, freedom of speech, gender equality and gay rights. In moving from neo-Nazism to counter-jihadism, the underlying structure of the narrative remains the same, but the protagonists have changed: the identity of Western liberal values has been substituted for white racial identity, Muslims have taken the place of blacks and multiculturalists are the new Jews.

The counter-jihadist narrative has been advanced by a trans-Atlantic movement, including think-tanks, bloggers, street-based movements and political parties. At the heart of the movement are websites such as *Gates of Vienna*, *Politically Incorrect* and *The Brussels Journal*, and think-tanks, such as the International Free Press Society and the David Horowitz Freedom Center, which fund and facilitate international linkages for the movement. In both the Netherlands and Denmark, counter-jihadist political parties have firmly entered the political mainstream; until recently in both countries, their support was necessary for governments to secure a working parliamentary majority. Geert Wilders, leader of the Dutch *Partij Voor de Vrijheid* (PVV, Freedom Party),

[22] Front National, *300 Mesures pour la renaissance de la France: programme de gouvernement* (Paris, Editions Nationales, 1993).
[23] Jeffrey Kaplan and Tore Bjørgo, *Nation and Race: the developing Euro-American racist subculture* (Boston, Northeastern University Press, 1998), p. 10.

shares with other counter-jihadist groups the same core beliefs but is careful to reject the methods of non-state violence or extra-parliamentary agitation favoured by street-based movements like the EDL. With a similar stance, the *Dansk Folkeparti* (DF, Danish People's Party) continues to be the third largest political party in Denmark. In Belgium, the *Vlaams Belang* (VB, Flemish Interest) has sought to move away from an older tradition of far-Right politics with roots in neo-Nazism and anti-Semitism to embrace counter-jihadist rhetoric. In Germany, the *Bürgerbewegung Pax Europa* (Pax Europa Citizens' Movement) is a counter-jihadist social movement and think-tank.

Just as the older far-Right narrative had a structural need for a Jewish conspiracy theory in order to explain the purported complicity of national governments with their enemies, so too the counter-jihadist movement tends towards conspiracy theory. After all, one might ask, why the need for popular mobilisation for the counter-jihadist cause when European governments already take a tough stance on fighting 'radical Islam'? The answer must be that government rhetoric about fighting Islamist 'extremism' is mere appearance; behind the scenes, ruling elites are secretly in league with the Islamic enemy. Hence the indispensability of the Eurabia conspiracy theory, outlined in Bat Ye'or's 2005 book *Eurabia: the Euro-Arab axis*. Her claim is that the Euro-Arab Dialogue – a programme initiated by the European Community's political establishment following the 1973 oil crisis, to forge closer links with Arab nations – was actually a secret plot by European politicians and civil servants to facilitate Muslim immigration, subjugate Europe and transform the continent into an Arab colony, Eurabia. Like the Jewish conspiracy theory of the *Protocols*, no evidence is offered. Nevertheless, through the mainstream conservative writing of Oriana Fallaci, Niall Ferguson and Melanie Phillips, the term 'Eurabia' has been associated with an image of Europe as cowardly and weak in the face of Islamic intimidation, allowing itself to be 'colonised' by an increasing Muslim presence.[24] The focus of such conspiracy theories is the moral corruption of the West's own leaders, academics and journalists, who are accused of a lack of pride in Western culture, which has led to relativism and appeasement of radical Islam. As the counter-jihadist Ned May puts it on his *Gates of Vienna* blog: 'the Jihad is just a symptom … the enemy lies within. This war is a civil war within the West, between traditional Western culture and the forces of politically correct multicultural Marxism that have bedevilled it for the last hundred years.'[25]

For the new conspiracy theorists, Islamist terrorism is just the visible tip of a hidden jihad iceberg. Alongside the use of violence is the strategy of 'stealth jihad' which aims at the infiltration of national institutions and the assertion of Muslim demands through the legal system. Muslims advocating for their civil rights or seeking to win political office are therefore to be regarded not as fellow citizens but as agents of a secret plan to impose totalitarian government on the world. Non-Muslims who stand with Muslims in challenging discrimination are 'dhimmis', the twenty-first century equivalent of the Cold War's 'fellow travellers', who have already internalised the status of second-class citizenship within an 'Islamo-fascist' state. The provision of halal food, sharia-compliant finance or prayer breaks in workplaces is 'creeping sharia', the first steps towards a society ruled by Islam. And since the Islamic doctrine of *taqiyya* supposedly sanctions systematic lying to non-Muslims to help advance sharia government, Muslims who say they interpret Islam as a religion of peace and tolerance are not to be trusted.

For neo-Nazis, members of other races could never be integrated; for counter-jihadists, Muslims can only be tolerated if they explicitly reject their Islamic culture and embrace 'our' values; until then, they are considered suspect at best, agents of sedition at worst. For both neo-Nazis and counter-jihadists, members of ruling elites are viewed, with few exceptions, as traitors who have betrayed Western civilisation. And whereas neo-Nazis view mixing with Jews and non-whites as undermining racial purity, counter-jihadists view them as potential allies so long as they declare their allegiance to 'our' values.

[24] Matt Carr, 'You are now entering Eurabia', *Race & Class* (Vol. 48, no. 1, 2006).

[25] 'The emperor is naked', *Gates of Vienna* (26 September 2006), http://gatesofvienna.blogspot.com/2006/09/emperor-is-naked.html.

Thus, counter-jihadism exists alongside more familiar forms of neo-Nazi discourse and, within the wider far-Right milieu, there is a continuing debate about ways forward: where to hold on to long-standing principles and positions, and where to compromise and adapt to new trends. For some, Jews are still the prime enemy; for others, Jews are now seen as allies in opposing Islam. For still others, counter-jihadist themes and neo-Nazi narratives merge in a cocktail of hatred.[26]

It has normally been assumed that the threat of far-Right violence is predominantly from those with a neo-Nazi ideology, whereas right-wing populists, rhetorically focusing on culture and identity rather than race, engage in the democratic process. In other words, hardcore racial identity politics is seen as fostering violence while the 'extremism lite' of values-based identity politics is thought to go hand in hand with democratic participation. What is striking about Breivik and the EDL is that, in their different ways, they have used violence to advance a values and identity ideology, rather than old-style racism.

2. Narratives and performativity

Narratives are the stories we tell ourselves and others about the world in which we live. We can use the term 'meta-narrative' to refer to the larger public narratives that persist over a longer period of time and which appear in a wide range of different settings. For example, the Cold War, the 'war on terror' and the monotheistic religions are all meta-narratives. The first two of these are also security meta-narratives in that they tell stories that are primarily about the threats we face and how to protect ourselves against them. And they are official meta-narratives because they are produced by states and embodied in government policies.

Narratives have plots, within which events are given significance and explained in terms of particular causes. They also have protagonists who are given particular identities. Events and protagonists are relational, in that they only make sense in relation to other actual and potential protagonists and other actual and potential events. And narratives are necessarily selective, reflecting choices about what is relevant and irrelevant, and foregrounding particular events and protagonists as opposed to others. Usually, narrative plots involve their protagonists being confronted with a disturbance or conflict which needs to be resolved through some course of action.

Recent scholarship in terrorism studies has stressed the question of what governments can do or not do to undermine jihadist narratives. For example, in June 2009, the National Coordinator for Counterterrorism in the Netherlands and the Centre for Terrorism and Counterterrorism at Leiden University convened an expert meeting to examine the narratives that jihadists use and what kinds of counter-narrative might be effective in response. While the exact causal relationships between narratives and acts of violence are highly opaque, this work proceeded on the assumption that some kind of relationship was plausible and that therefore governments could expect to reduce the potential for violence by advancing counter-narratives crafted with this aim in mind. President Obama's Cairo speech in the same month was cited as an example of the kind of counter-narrative that governments could deploy against al-Qaeda.[27]

A related question that has recently come to the fore is the performative power of counter-terrorism, by which is meant the intended and unintended consequences of the ways that governments communicate their security policies to the public. Official security narratives 'set the tone for the overall discourse regarding terrorism and counterterrorism – thereby mobilising (different) audiences for its purposes'. Beatrice de Graaf and Bob de Graaff argue that this communicative component to counter-terrorism may ultimately determine its effectiveness.[28] This is because the messages generated by counter-terrorism policies are themselves

[26] José Pedro Zúquete, 'The European extreme Right and Islam: new directions?', *Journal of Political Ideologies* (Vol. 13, no. 3, 2008).

[27] National Coordinator for Counterterrorism, *Countering Violent Extremist Narratives* (2010), p. 8.

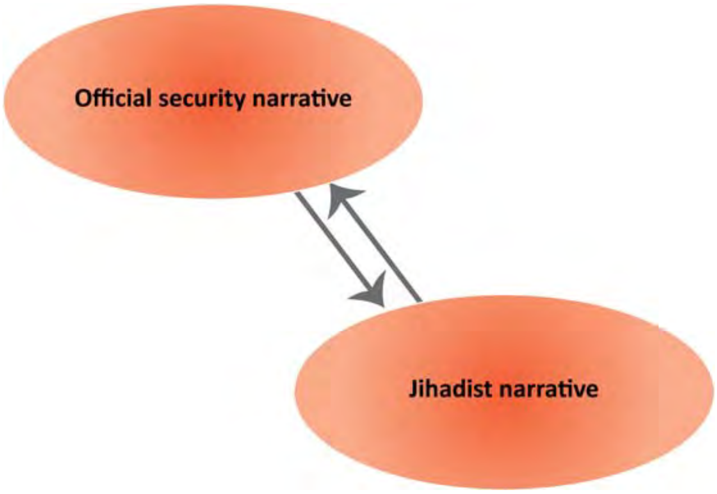
[28] Beatrice de Graaf and Bob de Graaff, 'Bringing politics back in: the introduction of the "performative power" of counterterrorism', *Critical Studies on Terrorism* (Vol. 3, no. 2, 2010), p. 267.

appropriated by terrorists who 'try to distort them and subsequently use them to fuel sentiments of oppression and injustice' in a battle of legitimacy engaged against governments.[29] From this perspective, terrorists and states are conducting 'a battle to convince and persuade different target audiences to rally behind them'.[30]

In a historical survey of the communicative aspects of counter-terrorism, Beatrice de Graaf defines the performative power of counter-terrorism as 'the extent to which the national government, by means of its official counterterrorism policy and corresponding discourse (in statements, enactments, measures and ministers' remarks) aims to mobilize public and political support and in the last instance, wittingly or unwittingly, assists the terrorists in creating social drama'.[31] Performative power can be measured by considering the extent to which terrorism is high on the political agenda and the level of perceived crisis; whether the terrorist threat is considered temporary and limited or wide-ranging and ongoing; whether governments attempt to mobilise society in opposition to terrorism; whether social conventions are seen as needing modification to deal with terrorism; and whether there is dialogue and the potential for recognition of terrorists' demands or government intransigence.[32] She concludes that a low-key approach to counter-terrorism which minimises perceptions of injustice and oppression in the population being targeted for recruitment is most likely to 'take the wind out of the sails that keep terrorists afloat'.[33]

The interactive narrative relationship between official security policies and jihadists is represented in figure 1 – arrows flow in both directions to represent the influence of government security narratives on jihadists and vice versa.

Figure 1



Terrorism studies scholars have also recently identified the danger of 'cumulative extremism' – the possibility that right-wing extremism and radical Islamism reinforce each other through a dynamic in which each one's narrative encourages support in the opposing group, in a spiral of fear and mutual demonisation.[34] Others use the term 'tit for tat radicalisation' to describe the same process.[35] For example, a key part of the EDL's counter-jihadist narrative is the need to oppose the threat of radical Islamist groups such as al-Muhajiroun and its

[29] Beatrice de Graaf, *Why Communication and Performance are Key in Countering Terrorism* (International Center for Counter-Terrorism – The Hague, February 2011), p. 7.

[30] Ibid., p. 7.

[31] Beatrce de Graaf, *Evaluating Counterterrorism Performance* (London, Routledge, 2011), p. 12.

[32] Ibid. p. 12.

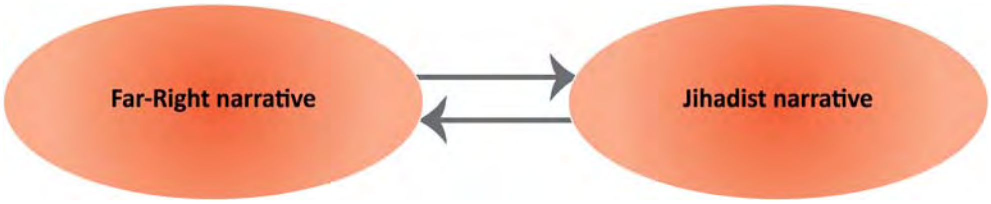
[33] Ibid., pp. 248–50.

[34] Roger Eatwell and Matthew J. Goodwin, eds, *The New Extremism in 21st century Britain* (London, Routledge, 2010), p. 243.

[35] Paul Jackson, *The EDL: Britain's 'new far Right' social movement* (University of Northampton Radicalism and New Media Research Group, 2011).

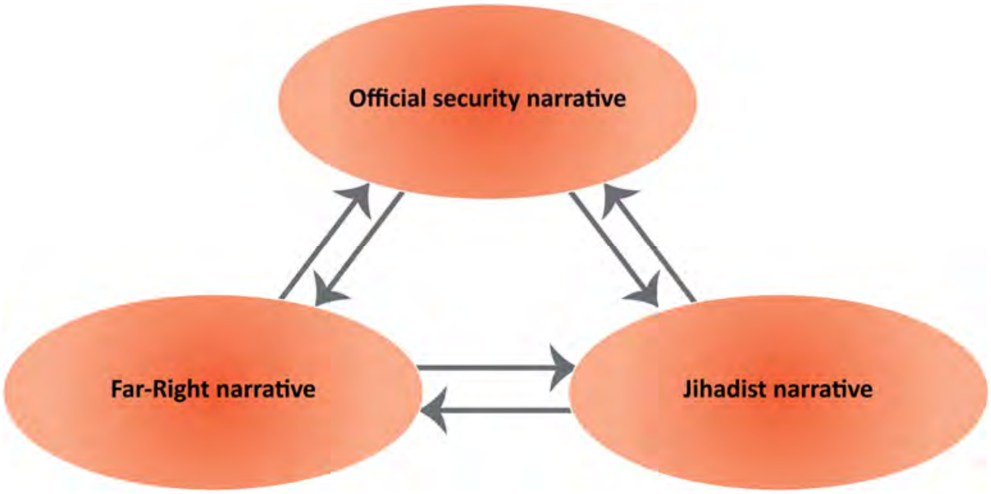
descendants. These groups, in turn, win support with the narrative that they need to exist to defend Muslims against the kind of Islamophobia represented by the EDL. The narratives of the EDL and radical Islamist groups therefore become mutually reinforcing – as represented in figure 2. As counter-jihadist narratives become more prominent among far-Right networks, this phenomenon becomes more relevant.

Figure 2



While there has been substantial research on counter-narratives to challenge the messages of jihadists, there has been little consideration given to counter-narrative strategies to undermine the far-Right. One way of opening up this question is to combine the notion that counter-terrorism has a crucial performative dimension with the concept of cumulative extremism to produce a complex relational picture of the mutual influences between the narratives of government security policy, jihadists and right-wing counter-jihadists. Represented graphically, this would mean combining figures 1 and 2 and adding a third side to a triangle of influences between the three actors, as shown in figure 3, with arrows representing lines of influence from one actor's narrative to another actor.

Figure 3



On this triangular model:

- Government security activity is influenced by the narratives of the jihadist terrorists it is attempting to counter, and to a lesser extent the narratives of the far-Right;
- Jihadist activity is influenced by official security narratives and by counter-jihadist narratives;
- Counter-jihadist activity is influenced by jihadist narratives and by official security narratives.

As we have seen, existing research has explored two sides of the above triangle – that between official and jihadist narratives and that between jihadist and counter-jihadist narratives. But the third side – that between official security narratives and counter-jihadist narratives – has not been explored. Consideration has not been given to whether the discursive frames of counter-terrorism policies are encouraging or discouraging far-Right narratives and whether governments could contribute to reducing far-Right violence by changing the way that counter-terrorism is narrated. This research paper seeks to pose these questions.

The relationship between counter-jihadism and security narratives is complicated by the probability that counter-jihadists respond not just to counter-terrorism narratives on far-Right violence but also to counter-terrorism narratives on jihadism. For example, support for the EDL may be influenced by government communication about the threat of 'radical Islam' as well as by government communication on the threat of far-Right violence. If terrorists and states are conducting 'influence warfare', then, given the above triangle of relationships, it is important to ask whether counter-terrorism policy narratives, in aiming to tackle jihadism, are leading to unintended consequences by reinforcing far-Right counter-jihadist narratives.

A precedent for such a triangular relationship between the state and opposed violent actors existed in Northern Ireland, at least during the 1970s and 1980s, when the narratives of the Provisional IRA, Ulster Unionist terrorists and the UK state were mutually reinforcing. While the UK government condemned the violence of both nationalists and Unionists, its rhetoric until the early 1990s – that there could be no negotiation on the basic question of sovereignty – reinforced the Unionist narrative of 'no surrender'; in addition, there was direct collusion between elements of the UK state and Unionist terrorist groups. Another example is the 'strategy of tension' employed in the 1970s by elements of the Italian security apparatus, which involved sponsoring right-wing terrorist groups in an effort to discredit and undermine the Left.

In a context in which society faces various forms of political violence, politicians and officials ought to be aware that the implicit 'stories' they tell about security policies and practices have audiences among potential jihadist supporters who are the target of the policy as well as among far-Right groups who may appropriate these 'stories' to reinforce their own message of violence.

Of course, in an information age, states are very far from having a monopoly on public communication. All those who are involved in communicating about terrorism (journalists, think-tanks, bloggers, civil society groups) share the same responsibility to reflect on the ways in which the narratives they circulate encourage or discourage violence. But states are responsible for a society's counter-terrorism policy and practice, and are the dominant force in shaping the ways society understands terrorism. Even in an age of new media and transnational integration, nation-states remain the critical actors in determining what points of view are considered sensible, realistic and legitimate.

3. Case studies

In an effort to explore the narrative relationships described above, in each of the four brief case studies under consideration (Britain, the Netherlands, Denmark and Belgium), the following questions are posed:

- **Official narratives**: What are the ways in which the problem of terrorism is narrated in official discourse? How are the protagonists and 'plots' in these narratives constructed? To what extent is terrorism seen as a political priority? How far is terrorism seen as causing a state of crisis in society? Is the terrorist threat perceived as limited or a wide-ranging problem? Do governments attempt to mobilise society to oppose terrorism? Are existing conventions seen as needing to be amended to oppose terrorism? Is there any recognition of the terrorists' political demands?
- **Far-Right narratives**: How do different far-Right actors narrate their political aims and strategy? How are the protagonists and 'plots' in these narratives constructed? Have official security discourses been appropriated by

the far Right? Have official security discourses contributed to far-Right violence being neglected as an issue? If far-Right ideology has developed beyond familiar forms of neo-Nazism to include 'counter-jihadist' and 'crisis of multiculturalism' elements borrowed from official security narratives, has it become harder for government agencies to recognise it and respond to it?

As well as examining counter-terrorist policy documents, ministerial statements and far-Right campaign material in each country under consideration, interviews were conducted with a small number of analysts and civil servants in each case.

It is beyond the scope of this paper to consider the underlying causes of the emergence of counter-jihadist and 'crisis of multiculturalism' discourses. Such an analysis would have to include a much broader historical account of racism and nationalism in Europe and their reworking in an age of neoliberalism.[36] However, it is suggested that, in some contexts, official rhetoric has played a role in encouraging far-Right narratives and that, where there is an overlap with government discourse, the state's ability to reduce the threat of far-Right violence is inhibited.

3.1. Britain

3.1.1. Official values-identity narrative strongly asserted and reflected in counter-terrorism policy

In the UK, the government considers the most serious terrorist threat to be from 'Al Qa'ida, its affiliates and like-minded organisations' and resources are targeted accordingly, whether in terms of the institutional focus of policing and intelligence agencies, government funding of preventative measures or ministerial leadership.[37] Until recently, the UK's counter-radicalisation policy was entirely focused on tackling Islamist 'extremism'.[38] (Although, over the last decade, more deaths probably resulted from the conflict in Northern Ireland than from jihadist violence in the UK.[39]) The threat from Islamist 'extremism' has been narrated in a series of ministerial speeches over the last six years. Speeches by Prime Minister Tony Blair (2006), Home Secretary Jacqui Smith (2008), Communities Minister Hazel Blears (2009) and Prime Minister David Cameron (2011) have been the major statements of government thinking on security matters since the 7/7 terrorist attacks on the London transport system in 2005. All these speeches present essentially the same story-line, despite a change in government in 2010 and some differences over policy details.[40] The key elements of this story-line are that:

- Our identity is based on liberal values of gender equality, freedom of speech, secularism, etc.;
- There are two kinds of Muslims: moderates who practise their religion in a peaceful way and share our values, and extremists/Islamists who interpret Islam as a political ideology, believe in rejecting our values and aim to impose sharia law on Muslims and non-Muslims;
- Political correctness and multicultural tolerance have weakened the defence of our values and thereby aided extremist Muslims;
- We have suffered terrorism because of Islamist extremism;
- We now need to put aside multicultural sensitivities, assertively defend our liberal values and be tougher in opposing Islamist extremism.

[36] This question is explored further in: Liz Fekete, *A Suitable Enemy: racism, migration and Islamophobia in Europe* (London, Pluto Press, 2009); Arun Kundnani, *The End of Tolerance: racism in 21st century Britain* (London, Pluto Press, 2007); and Alana Lentin and Gavan Titley, *The Crises of Multiculturalism: racism in a neoliberal age* (London, Zed Books, 2011).
[37] HM Government, *Prevent Strategy* (June 2011), p. 5.
[38] Arun Kundnani, *Spooked: how not to prevent violent extremism* (Institute of Race Relations, 2009).
[39] The Conflict Archive on the Internet (CAIN) at the University of Ulster lists 62 deaths probably related to the conflict in Northern Ireland from January 2002 to December 2011. There have been 53 deaths as a result of jihadist violence in the UK over the same period. See http://cain.ulst.ac.uk/sutton/.
[40] Tony Blair, speech to the World Affairs Council in Los Angeles, 1 August 2006; Jacqui Smith, 'Our shared values – a shared responsibility', speech to the International Conference on Radicalisation and Political Violence, 17 January 2008; Hazel Blears, 'Many voices: understanding the debate about preventing violent extremism', speech to London School of Economics, 25 February 2009; David Cameron, speech at Munich Security Conference, 5 February 2011.

The 'we' in this values-identity narrative is defined differently depending on the context – the people of Britain, the people of Europe or the people of 'the West' – but since 'our' identity is defined in terms of 'liberal values' which are assumed to be shared across Western societies, these distinctions are of limited importance. The significance of this narrative is that it introduces three protagonists (us, moderate Muslims and extremist Muslims), whose identities are defined in specific ways (whether or not they share our values), a disturbance (terrorist violence), an explanation for the cause of the disturbance (extremism) and a suggested resolution (rejecting multiculturalism and asserting our values more forcefully).

Among the policy impacts of such a narrative are immigration rules that seek to exclude foreign nationals from the UK on the basis of their not sharing 'our values'. The Prevent counter-radicalisation policy, introduced from 2007, embeds this narrative in a range of local settings, making available hundreds of millions of pounds of public money for initiatives to counter the circulation of Muslim 'extremism' and to encourage professionals such as teachers, lecturers, health workers and youth workers to be aware of the content of such ideology.

The major policy disagreement in the UK in this area has been over how to define the 'moderate' Muslims whom the government wants to recruit to a project of defending 'our values'. Some civil servants at the Home Office, departing from the narrative articulated by ministers, have argued privately that some of the most effective Muslim partners in preventative work are Salafi Muslims, who may hold 'extremist' values but nevertheless oppose violence in the UK. On this dissenting view, the 'us' that is to be united against 'them' is constituted by all those who oppose violence against Britain, rather than defined in terms of sharing British values. Since June 2011, however, when a review of Prevent policy was published, that analysis has been firmly rejected and the values-identity narrative of counter-terrorism now completely dominates official discourse.[41]

One consequence of the foregrounding of a values-identity narrative is that questions of identity, values and multiculturalism have been strongly linked to the issue of national security. David Cameron's speech at the Munich Security Conference in February 2011, given on the same day that the EDL marched through Luton, conveyed this strong sense that Britain was facing a generational problem of jihadist violence because of a legacy of misguided 'multiculturalist' policies that had failed to 'integrate' young Muslims into mainstream society. In this way, the terrorist problem was constructed not as consisting of a few individuals engaged in violence but as a symptom of a much deeper cultural malaise in the British Muslim population. This implied that the solution must involve not just a focus on a small number of Muslims but a mobilisation of a broader population to embrace a stronger sense of national identity based on 'muscular liberalism'. Hence the Prevent policy adopted a very broad approach, directed at the entire Muslim community, not just specific individuals or a few neighbourhoods. The values-identity narrative had already led funding for Prevent projects to be allocated in proportion to the number of Muslim residents in each local authority area – reflecting the policy's sole focus on Muslim 'extremism' and the assumption that it needed to address the entire Muslim population.[42] Like the previous ministerial statements, Cameron's speech also spelled out that existing conventions of 'multicultural tolerance' needed to be weakened to deal with the perceived crisis of identity and values.

In these ways, the official security narrative in the UK has a high degree of performativity: it presents the problem of terrorism as a major generational crisis, as rooted in a wide-ranging problem of identity, and as needing to be fought by mobilising whole sections of society and dispensing with existing social conventions.

3.1.2. Far-Right movements appropriate values-identity narrative

The main far-Right political party in the UK, the British National Party (BNP), enjoyed an increase in support from 2001 onwards – driven by multiple factors. In 1999, Nick Griffin assumed the leadership of the party and sought to downplay its neo-Nazi legacy, which included his own 1997 pamphlet claiming that Jews secretly controlled the

[41] HM Government, *Prevent Strategy* (June 2011).
[42] Arun Kundnani, *Spooked: how not to prevent violent extremism* (Institute of Race Relations, 2009).

media.[43] Re-modelling the party along the lines of more successful European counterparts such as the *Front National* in France, he used the language of defending British cultural identity (rather than white racial identity) against a ruling elite that wants to destroy it through immigration and multiculturalism. Instead of talk of a Jewish conspiracy, there was the idea that those in power are too 'cosmopolitan' to have the real interests of the British people at heart; and, after 9/11, Islamic militancy was invoked to illustrate the alleged dangers of immigration.

Despite the BNP's active membership remaining dominated by long-standing neo-Nazis and violent racists, from 2001 it was able to dramatically increase its electoral support. In June 2004, the BNP secured 808,200 votes across the UK in European elections; by 2009, the BNP had won two seats in the European parliament. From 2003 to 2010, the BNP had at least ten councillors in office at any one time, with the real possibility of winning control of a borough or city council, such as Burnley's. However, in the last few years, the BNP's organising capacity has been severely reduced, firstly by the leaking of its membership list and, secondly, by the financial burden of defending itself against a legal challenge to its racist membership policy. But these tactics targeted the organisation not the message, allowing other far-Right groups to pick up from where the BNP had left off.

As it turned out, the English Defence League (EDL) was well placed to do so. It does not organise as a conventional political party and has no formal members, so it is less vulnerable to the tactics that have been partially effective against the BNP. More significantly, the EDL has been able to better tailor its ideology to current circumstances, because it owes its entire outlook to the 'war on terror'. The BNP's opportunistic exploitation of Islamophobia after 9/11 carried it to a level of electoral support unimaginable in the 1990s. But, by virtue of its core membership, the party remains tethered to the neo-Nazi tradition and so, unlike the EDL, cannot fully realise the potential of the post-9/11 context.

There are multiple strands of opinion within the EDL, including conventional colour-based racism and straightforward opposition to the presence of Muslims in Britain. But the narrative it seeks to foreground, for example in its mission statement, has a different character. It can be summarised as follows:

- The West is at war against Islamic extremism;
- Unlike other groups in society, among Muslims, there is a problem of those who reject modern Western liberal and democratic values;
- It is wrong to assume that this is true of all Muslims. Rather, there is a conflict going on within Muslim communities between reformists who oppose orthodoxy and radicals who believe in a fundamentalist form of Islam;
- These radicals dominate Muslim organisations, remain key figures in British mosques, and are steadily increasing their influence;
- Cultural diversity is to be welcomed but parts of other cultures that conflict with liberal values cannot be tolerated in the name of multiculturalism;
- People of all races, religions and lifestyles should unite to oppose the growing power of radical Islam;
- The government has systematically failed to oppose Islamic extremism in Britain.

This narrative constructs subject positions of who 'we' are (those who share liberal values) and who 'they' are (those Muslims who reject them) and attempts to unite a diverse range of groups against Islamic extremism. It sets out a relationship of conflict between 'us' and 'them'. And it sets out the obstacles to victory in that conflict, such as failed government policies. And with this, the street activism of the EDL is legitimised: demonstrations against mosques, marches through Muslim communities, demands for tougher action against 'Islamic extremism'.

The EDL narrative differs from that of the traditional far-Right (for whom the 'we' is members of the 'white race') and even recognises a distinction between 'moderate' Muslims and 'extremist' Muslims. 'Race war'

[43] *Who are the Mind Benders? The people who rule Britain through control of the mass media* (London, Steven Books, 1997).

has been swapped for the idea of a global conflict between Western liberal civilisation and 'radical Islam'. For the EDL, Western civilisation is liberal, secular and modern. And this civilisation is open to Muslims to join; it does not completely close the door on them. Indeed, the question for the EDL is whether Muslims choose to join this civilisation or whether they choose to remain locked into what the EDL regards as the barbarity of traditional Islamic culture. If they do embrace Western liberal values, they can be seen as moderate Muslims. If they do not, then they are extremists. And passing this test is hugely significant, because moderates are potential allies, while extremists are people against whom a war is being fought. The EDL thus sees Muslims as people whose Islamic identity determines their whole being unless they can prove that they have freed themselves from it and embraced liberal civilisation. Politics can therefore be reduced to a conflict between the regressive cultural identities of traditional Islam and Western liberal values. And the only acceptable agency for Muslims is the rejection of their cultural practices in order to become 'free like us'.[44]

Of course, having to pass this values test is a flimsy basis for equality, as it means acceptance of Muslims as fellow citizens is conditional on meeting the moral or political approval of others who regard you with suspicion. And using the language of culture and values to define a 'Muslim problem' can produce the same outcomes that more obviously racial discourses once achieved; cultural tropes, such as wearing a hijab, can serve as signifiers of who belongs and who does not, in the same way that skin colour does. Yet precisely because this narrative differs from familiar patterns of racialisation, it can present itself as the defence of a liberal 'way of life' and appear 'post-racial'. This explains the paradox of a far-Right organisation that is able to tentatively include supporters from a variety of ethnic backgrounds and invoke 'liberal ideals', such as women's rights, gay rights, and 'democratic accountability'.

With its focus on whether Muslims share 'our values', the EDL's definition of the 'problem' is strikingly similar to Britain's official security narrative. The EDL takes literally government statements that there is a conflict between 'our values' and 'Islamic extremism'. From counter-terrorism programmes, it absorbs the notion that the enemy in this conflict is not a few individuals engaged in violence but an ideology embedded in Muslim communities. Likewise, the notion that Muslims can be categorised as extremist or moderate, according to their allegiance to Western values, has been taken from the official narrative. And from ministerial speeches, the EDL borrows the belief that 'state multiculturalism' is holding back the fight against Muslim 'extremism'. The EDL would take a tougher view on the extent to which Muslim communities embrace liberal values, seeing a more widespread rejection than official discourse would allow. But the main difference between the EDL narrative and the official narrative lies elsewhere: the EDL holds that the politicians running the domestic 'war on terror' are too soft and cowardly, still too caught up in multicultural platitudes to fight it properly; this is where a new far-Right street movement will fill the gap with its own form of militancy. It is this last element – government failure – that justifies the need for a social movement willing to fight the enemy on the streets, and gives the EDL its militancy and distance from the liberal state.

This suggests that Britain's official security narrative has been strongly performative, providing discursive opportunities for new far-Right actors whose ideologies significantly overlap with government discourse, and which are therefore harder to counter. The claim is not that, in drafting their mission statement, EDL leaders studied ministerial speeches and policy documents. Rather, the argument is that the government, through its leadership role in public discourse on terrorism, has been able to entrench a values and identity narrative as the prevalent way in which terrorism is understood in society, and that this narrative – amplified by popular newspapers, such as the *Mail*, *The Sun* and *Star*[45] – has been ripe for appropriation by the far-Right.

[44] Alana Lentin and Gavan Titley, *The Crises of Multiculturalism: racism in a neoliberal age* (London, Zed Books, 2011).

[45] For example: Ruth Dudley Edwards, 'Will Britain one day be Muslim?', *Daily Mail* (5 May 2007). See also further examples listed in Ryan Erfani-Ghettani, 'Strangers in our own land', *IRR News* (23 March 2012), http://www.irr.org.uk/news/strangers-in-our-own-land.

3.1.3. Far-Right violence neglected in official security narrative

In post-7/7 Britain, there has been a consistent problem of 'fitting' the threat of far-Right violence into official security narratives. For most of its existence, Prevent policy has completely neglected the far-Right as an issue. Prevent practitioners interviewed in the first half of 2009 were unable to cite any examples of work specifically aimed at tackling the far-Right.[46] As of the end of 2010, less than 10 per cent of individual interventions designed to prevent radicalisation, as part of Prevent's Channel programme, were directed at the far-Right; over 90 per cent of the programme's focus was on Muslims.[47] The June 2011 Prevent policy review publicly recognised the existence of a far-Right threat but it was strongly downplayed: there was only a 'small number of relevant cases' and there were no 'extreme right-wing terrorist organisations and formal groups'.[48] Above all, the far-Right threat was not conceived to be part of a wider social drama; whereas Islamist terrorism was seen as symptomatic of a generational conflict over values, multiculturalism and identity, far-Right violence was seen as involving no more than a few isolated 'lone wolves'.

With regard to the EDL, there is a reluctance by many officials and advisors to recognise the group as a significant threat. For example, in April 2011, Adrian Tudway, the police's National Co-ordinator for Domestic Extremism, wrote in an email to Muslim groups that: 'In terms of the position with EDL, the original stance stands, they are not extreme right wing as a group, indeed if you look at their published material on their web-site, they are actively moving away from the right and violence with their mission statement etc.'[49] Similarly, in January 2011, Douglas Murray, the associate director of the Henry Jackson Society, which influences the government on national security policy, stated that, in relation to the EDL: 'If you were ever going to have a grassroots response from non-Muslims to Islamism, that would be how you'd want it, surely.'[50] Both these statements suggest that 'counter-jihadist' ideologies, through reworking far-Right narratives and appropriating official discourse, are able to evade categorisation as a source of far-Right violence.

3.2. Netherlands

3.2.1. Official values-identity narrative strongly asserted and reflected in counter-terrorism policy

Like in Britain, counter-terrorism discourse in the Netherlands has strongly focused on jihadist terrorism, understood largely through a values-identity narrative in the public and political debate, in which 'lack of integration' with purported 'Dutch liberal values' such as gender equality, freedom of speech and secularism is seen as causing a generational problem of 'polarisation' between young Muslims and the rest of Dutch society, which in turn is thought to create a breeding ground for violent extremism. While some Dutch counter-terrorism officials and politicians have spoken more about 'grievances' than about values, they have done so in the shadow of a prevailing security narrative of 'polarisation', understood as a problem of values and identity. A mainstream public debate has taken place on whether terrorism is the product of Islam itself or of particular extremist interpretations of Islam. But both sides in this public and political debate have held a common set of assumptions:

- Our Dutch identity is based on liberal values of gender equality, freedom of speech, secularism, etc.;
- Either all followers or some minority of followers interpret Islam as a political ideology (Islamism), believe in rejecting our liberal values and aim to impose sharia law on Muslims and non-Muslims;
- Political correctness and multicultural tolerance have weakened the defence of our values and thereby aided the polarisation that leads to Islamist extremism;
- We have suffered terrorism because of Islamist extremism;

[46] Arun Kundnani, *Spooked: how not to prevent violent extremism* (Institute of Race Relations, 2009).
[47] Home Office, 'Channel data 2007–2010', document released under the Freedom of Information Act, May 2011.
[48] HM Government, *Prevent Strategy* (June 2011), p. 20.
[49] Vikram Dodd and Matthew Taylor, 'Muslims criticise Scotland Yard for telling them to engage with EDL', *Guardian* (2 September 2011).
[50] http://www.youtube.com/watch?v=1wgAliHwrNo.

- We now need to put aside multicultural sensitivities, assertively defend our liberal values and be tougher in opposing Islamist extremism.

The more radical voices in this public debate, such as the politicians Geert Wilders and Ayaan Hirsi Ali, have argued that Islam can only be interpreted in an extremist fashion, whereas most other politicians and commentators have held that a moderate Islam, compatible with Dutch values, is possible and that only a minority of Muslims are extremists. This latter view has allowed for a narrative with three protagonists (us, moderate Muslims and extremist Muslims), while the former leads to a simpler narrative in which the 'moderate' protagonist is dropped from the story-line. But both narratives define the identities of their protagonists on the basis of whether or not they share 'our values', give an explanation for the cause of the terrorism 'disturbance' (it is a product of Islamist extremism bred in a context of polarisation) and suggest a resolution (rejecting multiculturalism and asserting 'our values' more forcefully).

While these narratives became especially important following the murder of filmmaker Theo Van Gogh in 2004, the groundwork had already been laid some years earlier. Indeed, the Netherlands was a pioneer in advancing this kind of 'new realism' narrative in Europe.[51] In 1991, a speech by Frits Bolkestein, the leader of the *Volkspartij voor Vrijheid en Democratie* (VVD, Liberal Party), on the need for minorities to integrate to the values of Dutch liberalism, prompted a national debate on cultural integration.[52] The *Partij van de Arbeid* (PvdA, Labour Party) also began to quietly move towards a more restrictive immigration policy and away from multiculturalism while it was in government in the 1990s.[53] In 1997, Pim Fortuyn published his book, *Tegen de islamisering van onze cultuur* (*Against the Islamisation of our Culture*), which provided the clearest statement yet of the values-identity narrative.[54] His subsequent electoral success in municipal elections in Rotterdam, before his murder in 2002, demonstrated its potential appeal to a section of voters. Geert Wilders ploughed the same furrow after he left the VVD in 2004 to establish his own party. Other parties responded to the emergence of this new 'radical Right' with a good measure of emulation, following Wilders in a lot of his rhetoric. While they did not embrace full-on Islamophobia, the other parties did largely support the notion that 'multiculturalism' is in crisis because of the failure of Muslim immigrant groups to 'integrate', placing a values-identity narrative firmly in the mainstream. Over the last decade, a series of new 'integration' measures directed at Dutch Muslims have been proposed, culminating in the recently drafted niqab and burka ban, new barriers to immigration have been brought in, for example through tighter rules on family reunion, and multiculturalism has been officially pronounced a failure.[55] Though not specifically focused on counter-terrorism, these developments have further entrenched a values-identity narrative in relation to Dutch Muslims.

Like in Britain after 7/7, the Van Gogh murder was interpreted as both an individual act of terrorism and symptom of a wider problem – the failure of significant numbers of Muslims to integrate and adopt Dutch values. For example, *From Dawa to Jihad: the various threats from radical Islam to the democratic legal order*, a major policy study published in 2004 by the *Algemene Inlichtingen- en Veiligheidsdienst* (AIVD, General Intelligence and Security Service), argued that Dutch democratic values are under threat from Muslim 'extremists' and that Muslim 'moderates' who share these values need to be assisted by wider society to defend them.[56] In response, alongside the investigative efforts of the police and intelligence agencies focused on individual criminals, counter-radicalisation policies were implemented by municipal authorities in an attempt to mobilise a wide range of 'partners' to address Muslim 'extremism' at the community level. Thousands of 'front-line' workers, such as teachers, police officers and youth workers, were given training on spotting the 'warning signs' that a young person was rejecting Dutch values and embracing an 'extremist' Islamic identity.

[51] Baukje Prins, *Voorbij de onschuld: het debat over integratie in Nederland* (Amsterdam, van Gennep, 2004), pp. 34–6.
[52] Rob Witte, 'The Dutch Far Right: from "classical outsiders" to "modern insiders"', paper presented at CSTPV Workshop, 16–17 May 2011.
[53] Tim Bale, Kurt Richard Luther, Christoffer Green-Pedersen, André Krouwel, Nick Sitter, 'If you can't beat them, join them? Explaining social democratic responses to the challenge from the populist radical Right in Western Europe', *Political Studies* (Vol. 58, 2010), p. 416.
[54] Pim Fortuyn, *Tegen de islamisering van onze cultuur. Nederlandse identiteit als fundament* (Utrecht, Bruna, 1997).
[55] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 'Aanbieding visie op integratie' (June 2011).
[56] Dutch General Intelligence and Security Service, *From Dawa to Jihad: the various threats from radical Islam to the democratic legal order* (2004).

By 2010, some individual practitioners were arguing privately for a separation of the questions of values, identity and polarisation from the discussion of counter-terrorism, and questioned the direct connection between lack of Muslim integration and terrorism.[57] However, they did so against the continuing backdrop of a political context in which Muslim 'lack of integration' was seen as driving a generational problem of jihadist violence.

3.2.2. Far-Right violence neglected in official security narrative

The counter-radicalisation initiatives introduced from 2004 were at first focused entirely on Muslim communities. In parliament, a White Paper discussing the policy had been questioned because of its failure to address other forms of extremism. However, in implementing the policy, at least initially, there was often insufficient interest in seeing the far-Right as a problem. For example, Rotterdam city council did not initially believe it had a problem of right-wing extremism, even though it is known to be home to a strong far-Right movement.[58] Other municipal authorities informed ministries that the far-Right was, in fact, the only problem of extremism in their area. But civil servants reportedly believed that the professionals working on counter-radicalisation, being predominantly white, would find it straightforward to identify 'warning signs' of right-wing extremism, as they would have a sense of what is 'normal' and what is not for white young people, whereas they would find it hard to distinguish mainstream Islamic theology from the radical fringe. For these reasons, they felt there was no need to introduce specific training programmes to tackle far-Right ideology.[59] However, the assumption that far-Right extremism is familiar enough for most people to identify and challenge becomes less plausible in a situation where mainstream political leaders and perpetrators of far-Right violence share much of the same counter-jihadist narrative.

Counter-radicalisation initiatives appear to have focused 90 per cent of their resources on Muslim extremism from 2004 until 2008, when the policies began to lose funding.[60] Efforts by at least one civil society counter-radicalisation partner to persuade civil servants that the far-Right also needed to be addressed were rebuffed because such concerns did not fit with the prevailing focus on Muslim extremism. According to some observers, there is an additional problem in that, among local police officers and municipal authorities, there is a tendency to downplay the problem of far-Right violence because of a fear of their town acquiring a negative association with right-wing extremism. Instead, far-Right violence is labelled as a 'youth problem' of delinquency and public order, and its political dimension downplayed.[61] For this reason, even where national resources are available for tackling far-Right extremism, municipal authorities are reluctant to make use of them, due to this desire to protect a town's reputation.

Today, the office of the National Coordinator for Counterterrorism and Security has a small programme of work on far-Right extremism but it is marginal to the organisation as a whole, which continues to focus predominantly on Islamist extremism. To the extent that there has been attention directed towards the far-Right, it has largely been couched in terms of the 'lone wolves' model, which emphasises individuals with mental health problems, rather than addressing the wider social and political context. Whereas officials have investigated how jihadist narratives might be publicly countered in various ways, there has been no attempt to pursue counter-narrative initiatives in relation to far-Right violence. In general, the political influence of the *Partij Voor de Vrijheid* (PVV) has been an additional barrier to focusing more attention on the question of far-Right violence.

3.2.3. Counter-jihadist politics in the mainstream

According to the Dutch intelligence service, the AIVD, there are no more than 300 active followers of 'right-wing extremism and the extreme right' in the Netherlands.[62] This refers to activists in the classic neo-Nazi and racial

[57] Interview with counter-terrorism policy-makers, Den Haag, 7 July 2011.
[58] Jaap van Donselaar and Peter R. Rodrigues, eds, *Racism and Extremism Monitor Eighth Report* (Anne Frank Stichting/Leiden University, 2008), p. 12.
[59] Interview with counter-terrorism policy-maker, Den Haag, 10 May 2012.
[60] Interview with counter-radicalisation practitioner, Netherlands, 9 May 2012.
[61] Interviews with counter-radicalisation practitioners, Netherlands, 4 April 2012, 1 May 2012, 9 May 2012.
[62] General Intelligence and Security Service, *Right-wing Extremism and the Extreme Right in the Netherlands* (March 2011), p. 11.

supremacy tradition. The emergence of informal counter-jihadist networks has begun to attract some limited official attention following the Oslo-Utøya attacks in July 2011 but is considered largely a matter of 'keyboard activism' at present. A recent attempt to establish a Dutch Defence League on the model of the EDL did not succeed, and it folded within a year. There was little enthusiasm among Dutch right-wingers to attend the Aarhus counter-jihadist meeting in March 2012. While a civil war narrative is circulated online, officials say there is no evidence of individuals acquiring weapons in anticipation, as has happened in other countries. However, unlike in Britain, a counter-jihadist narrative is strongly articulated in the mainstream political process by the PVV.

Prior to the emergence of the new 'radical right' of Pim Fortuyn and Geert Wilders' PVV, the most successful far-Right parties came from the 'Centre movement' of the 1990s, which included the *Centrum Democraten* (CD, Central Democrats), *Centrum Partij* (Central Party), CP'86, *Nationale Alliantie* (National Alliance), *Nieuwe Nationale Partij* (New National Party) and *Nieuw Rechts* (New Right). The CD won a seat in parliament in 1989 and achieved 2.5% of the national vote in 1994, including 12% of the vote in Rotterdam. However, it was plagued by internal rifts and fell foul of Dutch laws on the incitement of discrimination.[63] In 1993, five leaders of the breakaway CP'86 were arrested for conducting a campaign of racist violence in the name of the 'Nijmegen Liberation Front'.[64]

As analyst Rob Witte has argued, many of the issues these classic far-Right groups pioneered – such as asylum-seeking, immigration and cultural integration – became increasingly prominent in mainstream political and public discourse during the 1990s, often using similar terminology and arguments.[65] By the turn of the millennium, the classic far-Right parties, such as the Centre Democrats, had dropped out of the picture for the most part and new political parties such as the *Lijst Pim Fortuyn* (LPF, Pim Fortuyn List), and later the PVV, were able to pick up the themes they had focused on. Like the classic far-Right parties, these parties argued that elite 'multiculturalism' had allowed 'immigrants', particularly Muslims, to undermine Dutch identity. With Wilders, a fully counter-jihadist narrative emerged that viewed Islam as a totalitarian ideology intent on introducing sharia law through violence and subversion. However, the new parties did not emerge from the existing far-Right milieu, and differed in how they defined Dutch identity, which they described in terms of cultural values of freedom of expression, secularism and gender equality, rather than in terms of race. Because this new counter-jihadism did not fit the usual image of neo-Nazism, it was able to normalise an Islamophobic, identitarian discourse within the mainstream political process.[66] The older far-Right vote was largely swallowed up by the new parties, which nevertheless took care to ensure that neo-Nazis and old-fashioned racists were excluded from active participation. Wilders' strong support for the Israeli right-wing helped to demarcate a clear distinction between himself and the far-Right tradition; since the end of the Second World War, anti-Semitism had been considered the key test in the Netherlands of whether far-Right politics had crossed the line into public unacceptability.

By 2010, the normalisation of the new far-Right was completed with the establishment of a government that depended on the support of the PVV for a parliamentary majority. Though not a member of the cabinet, Wilders was able to strongly influence policy on security, integration, migration and asylum. In this new climate, attempts to label the PVV a far-Right party became increasingly difficult in the public sphere. An annual academic speech due to be given by the historian Thomas von der Dunk was cancelled when it emerged that he intended to draw an analogy between the PVV and pre-war pro-Nazi parties in the Netherlands. Similarly, the punk band *Jos en de Tosti's*, scheduled to play at the annual festival commemorating Dutch liberation from Nazi occupation, was reportedly asked not to perform its song '*Mussolini van de Lage Landen*', which placed Wilders within the history of fascism. Rob Witte notes that these attempts to silence critical voices came from 'members of established political elites, uncomfortable with outspoken criticism of the extremist elements'.[67] Meanwhile, with stunts such

[63] Marcel Lubbers, *Exclusionistic Electorates: extreme right-wing voting in Western Europe* (Katholieke Universiteit Nijmegen, 2001), pp. 16–17, 185.
[64] Liz Fekete, 'Centrum Partie '86 and racial violence', *European Race Bulletin* (March 1994).
[65] Rob Witte, 'The Dutch Far Right: from "classical outsiders" to "modern insiders"', paper presented at CSTPV Workshop, 16–17 May 2011.
[66] Tjitske Akkerman, 'Anti-immigration parties and the defence of liberal values: the exceptional case of the List Pim Fortuyn', *Journal of Political Ideologies* (Vol. 10, no. 3, 2005).
[67] Rob Witte, 'The Dutch Far Right: from "classical outsiders" to "modern insiders"', paper presented at CSTPV Workshop, 16–17 May 2011.

as his call for banning the Quran and his video *Fitna*, Wilders became an international icon of the counter-jihadist movement. He declared himself an admirer of Ba'et Yor's 'Eurabia' conspiracy theory and received support from Islamophobic movements in Europe, Israel and the US. According to Dutch newspaper reports, Wilders receives substantial funding from the US-based David Horowitz Freedom Center, which, with an annual budget of around $5million, is a major financier of the counter-jihadist movement, including websites such as Robert Spencer's *Jihad Watch*.[68]

In the short-term, it appears that the success of the PVV in mainstream Dutch politics has stolen the thunder of the street-based far-Right movements. Immigration and identity are now central debates within the democratic process. But mainstream acceptability also brings with it greater dangers. Margaret Thatcher's incorporation of National Front rhetoric into her 1979 general election campaign rendered the NF irrelevant in British politics, reversing its rise in popularity over the previous decade. But by opening political space for the NF's narrative within her party, she also helped contribute to the social acceptance of cultural racism.

In the Netherlands, the costs of Wilders' presence at the centre of Dutch politics may prove similarly damaging. Some observers express concern that young Muslims are withdrawing from political engagement after Wilders' success, leading to a generation of Dutch Muslim citizens who are alienated from public life, constantly talked about but never talking back. A 2010 survey of young Muslims in Amsterdam found that the 'increasing anti-Islam climate' and 'public insults' associated with Wilders 'have led to fear, frustration and anger'; almost all of the participants had 'at some point in their lives experienced feelings of injustice, stigmatization and discrimination'.[69] In general, there is a lack of Dutch Muslim voices in the public sphere able to articulate the community's experiences and advocate on its behalf.

At the same time, the problem of racist and Islamophobic harassment and violence continues.[70] The sociologist Ineke van der Valk documented 117 attacks on Dutch mosques from 2005 to 2010.[71] Halim el Madkouri, a programme director of Forum, the Institute for Multicultural Affairs, estimates the actual number to be much higher – perhaps five attacks on mosques each month in the Netherlands – and says verbal abuse against women wearing headscarves is common, particularly in smaller towns. He is clear that Islamophobic violence is fuelled by the rhetoric of mainstream politicians: 'I have never seen Islam on the street – I see only Muslims. So if you say you want to get Islam out of Europe, it means getting rid of Muslims.' Media and public interest in the victims of such violence is minimal and, in the absence of an official response, el Madkouri notes, young people experiencing it have taken to defending themselves. Because of a fear of publicity, these conflicts are classified by authorities as generic 'youth problems' and their racial dimension ignored.

The official security narrative in the Netherlands has had a high degree of performativity: like in the UK, the problem of terrorism has been presented as a major generational crisis, rooted in a wide-ranging problem of identity, and as needing to be fought by mobilising whole sections of society and dispensing with existing social conventions of tolerance. The salience of this 'story-line' in policy-making discourse has been sustained both by the PVV and other political parties that have absorbed a values-identity narrative. One consequence has been counter-terrorism policy neglecting the threat of far-Right violence, which does not 'fit' the official narrative. There have been some positive signs of change following the Breivik case in Norway and the decreased direct influence of Wilders after the collapse of the coalition government in April 2012. For example, in its latest annual report, the AIVD has added a category of 'anti-Islam(ism)' in its section on 'radicalism and extremism'.[72] However, it remains to be seen whether the ways in which counter-terrorism is understood in the Netherlands are amenable to substantial change.

[68] Tom-Jan Meeus, 'Grote geld voor Wilders ligt in VS', *NRC Handelsblad* (25 January 2012).

[69] Mare Stijntje Visser, 'Consequences of the PVV: an immigrant perspective', *Re-public: reimagining democracy* (16 January 2012), http://www.re-public.gr/en/?p=5194.

[70] Rob Witte, *Racist Violence in the Netherlands* (European Network Against Racism, 2011).

[71] 'Ruim 100 geweldsincidenten bij moskeeën', *De Telegraaf* (30 December 2011).

[72] Algemene Inlichtingen- en Veiligheidsdienst, *Jaarverslag 2011* (April 2012), p. 17.

3.3. Denmark

3.3.1. Official values-identity narrative strongly asserted and partially reflected in counter-terrorism policy

Since the 1980s, Danish politics has been progressively transformed by far-Right anti-immigrant and Islamophobic movements, which have constructed a narrative of Danish liberal identity threatened by Muslim immigrants seen as bringing an incompatible set of cultural values.[73] Beginning with the formation of the *Den Danske Forening* (DDF, The Danish Society) anti-refugee protest group in 1986, led by the priest Søren Krarup, through to the 2001 election, in which the far-Right *Dansk Folkeparti* (DF, Danish People's Party) became the key partner of the Liberal-Conservative governing coalition (an arrangement that lasted until 2011), a values-identity narrative has been fully normalised in Danish public culture. From the 2001 election, the DF placed culture and values at the centre of its programme, arguing that Denmark had been betrayed by a political elite that had favoured multiculturalism and immigration, threatening the very substance of Danish identity by importing a Muslim culture that was incompatible with European modernity. As analyst Susi Meret notes:

> 'After 9/11, the Danish People's Party clearly radicalized its rhetoric against Islam. The difference between Islam and Islamism (radical Islam) at times disappeared from the party political discourses and Islam was more and more often directly associated with a totalitarian and violent ideology, whose destructive effects were seriously jeopardising Western democratic principles and values from within.'[74]

By 2007, Krarup, by then a DF member of the Danish parliament, was advancing the counter-jihadist notion of Islam as a form of totalitarianism: 'The (Muslim) veil is a totalitarian symbol that can be compared to the symbols we know from the Nazi swastika and from communism.'[75] However, Anders Fogh Rasmussen, the then Liberal Party prime minister, was also advocating a values-identity narrative in a slightly different form, claiming in 2005 'an aggressive practice of Islam as the greatest challenge to the cohesive force in Danish society'.[76] He had earlier spoken of the need to launch a 'cultural war of values' to transform Danish society in a neoconservative direction, a project for which influential allies existed in the print media.[77] Brian Mikkelsen, the Conservative Party minister of cultural affairs, already announced a crisis of multiculturalism in 2005: 'We have gone to war against the multicultural ideology that says that everything is equally valid.'[78] He added: 'In Denmark, we have seen the appearance of a parallel society in which minorities practise their own medieval values and undemocratic views. This is the new front in our cultural war.'[79] Analyst Peter Hervik notes that the process of other parties absorbing narratives from the far-Right had already begun in the 1990s with social democrats adopting the rhetoric of the DF's predecessor, the Progress Party, in order to retain voters or capture new support.

From 2001 to 2011, the DF was able to directly influence policy-making, particularly on matters of integration and immigration, as the government was dependent on its consent to secure a working majority. For example, following the election, a new Ministry for Refugees, Immigrants and Integration was established and tighter restrictions on immigration policy were brought in, especially with regard to family union. Reflecting the DF's identitarian politics, new 'integration contracts' for permanent residents were introduced, requiring would-be immigrants to declare their allegiance to 'Danish values' of self-sufficiency, gender equality, freedom of

[73] Karen Wren, 'Cultural racism: something rotten in the state of Denmark?', *Social & Cultural Geography* (Vol. 2, no. 2, 2001).

[74] Susi Meret, *The Danish People's Party, the Italian Northern League and the Austrian Freedom Party in a Comparative Perspective: party ideology and electoral support* (Aalborg University, SPIRIT PhD Series Thesis no. 25, 2009), p. 127.

[75] 'Outrage in Denmark after MP compares Muslim veil to swastika', *Agence France Presse* (19 April 2007).

[76] Quoted in Ferruh Yilmaz, *Ethnicized Ontologies: from foreign worker to Muslim immigrant: how Danish public discourse moved to the Right through the question of immigration* (University of California, San Diego, PhD Thesis, 2006), p. 194.

[77] Peter Hervik, 'Ending tolerance as a solution to incompatibility: the Danish "crisis of multiculturalism"', *European Journal of Cultural Studies* (Vo. 15, no. 2, 2012), p. 218.

[78] Stefan Theil, 'The end of tolerance: farewell, multiculturalism: a cartoon backlash is pushing Europe to insist upon its values', *Newsweek* (6 March 2006).

[79] Martin Burcharth, 'Capture the flag', *New York Times* (12 February 2006).

speech, and so on. The Danish Centre for Human Rights, the Board for Ethnic Equality and the Documentation Centre on Racial Discrimination were all closed by the government.[80]

While integration policy is highly performative in Denmark and completely dominated by a values-identity narrative, the picture with counter-terrorism policy is more complex. In ministerial speeches on terrorism, a familiar values-based story-line has often been articulated. For example, Prime Minister Rasmussen spoke in 2006 of terrorism being an aspect of 'a global value struggle' between 'sensible enlightenment and fundamentalist darkening'. This 'global value struggle takes place in Denmark too' where:

> 'fortunately … the great majority of Danes with an immigrant background … are contributing positively to the Danish society. But there are also a few extremists who seem to hate the society which has secured their political freedom and material safety. … We must demand respect for the very fundamental rules of the game in Danish society … We must not, out of naïve and happy-go-lucky tolerance, show understanding towards or facilitate religious fanaticism or political extremism.'[81]

Here we have the usual narrative with protagonists of 'us', 'moderate Muslims' and 'extremist Muslims' defined in terms of allegiance to 'our values', the explanation of terrorism as a problem of 'fanaticism' or 'extremism', and the danger of excessive 'tolerance' allowing extremism to advance. Earlier Rasmussen had stated that an 'active integration policy at home' was a part of the counter-terrorism strategy, underlining the perceived linkage between rejection of Danish values and terrorism.[82] Unsurprisingly, where this values-identity narrative has been in the foreground, the question of far-Right violence has not arisen. For example, the Danish government's 2011 report on counter-terrorism only names one form of terrorist threat, that from 'networks, groups and individuals that subscribe to a militant Islamist ideology', and none of the initiatives it mentions under its 2009 action plan to 'prevent radicalisation' are explicitly directed at the far-Right.[83] (Interestingly, the majority of terrorism prosecutions in Denmark listed in the report relate to groups that are neither Islamist nor far-Right, but Leftist national liberation movements, but this is not reflected in the report's overall narrative.[84])

However, in practice, Danish counter-radicalisation policies have not focused solely on Muslims and the language of values has been less prevalent in policy-making discourse than in the Netherlands or Britain. Counter-radicalisation policy literature written for local partners and practitioners rather than for a political audience is fairly rigorous in giving equal weight to different forms of extremism, focusing on left-wing, right-wing and 'Islamist' extremisms. Civil servants note that the particular focus will vary by location. The emphasis is less on cultural values and integration, and more on personal relationships, belonging and participation.[85] Until the new government of 2011, the DF had 'quite a heavy voice in policy-making', as one civil servant put it, and there was strong pressure to focus policies more on 'Islamists' and to link the issue of terrorism to the wider issue of the perceived failed integration of Muslim communities.[86] But this seems to have been partially resisted. Analyst Ulrik Pram Gad suggests that Danish counter-radicalisation policy has tried to position itself as a third way between a parliamentary opposition that favours 'self-reform' of Muslims and the DF, which does not believe Muslims even capable of reform. The resulting strategy has been to promote reform of Muslim values through engaging in a process of 'two-way dialogue'. Although this 'dialogue' has strict limits, it opens a space for a values-identity narrative to be partially challenged, even if it remains the basic framework within which the policy is conceived.[87]

[80] Peter Hervik, 'Ending tolerance as a solution to incompatibility: the Danish "crisis of multiculturalism"', *European Journal of Cultural Studies* (Vo. 15, no. 2, 2012), p. 216–7.
[81] Quoted in Ulrik Pram Gad, 'It takes two to tango: Danish concepts of dialogue as counterterrorism' (Norsk Utenrikspolitisk Institutt, Working Paper 747, 2008), p. 5.
[82] Quoted in Ibid., p. 8.
[83] *Government Report on Counter-Terrorism Efforts* (Danish Government, May 2011), p. 4.
[84] Ibid., p. 9.
[85] *A Common and Safe Future: an action plan to prevent extremist views and radicalisation among young people* (Government of Denmark, January 2009); *Preventing Extremism: a Danish handbook series* (Ministry of Social Affairs and Integration, 2011).
[86] Interview with counter-terrorism policy-maker, Copenhagen, 10 May 2012.
[87] Ulrik Pram Gad, 'It takes two to tango: Danish concepts of dialogue as counterterrorism' (Norsk Utenrikspolitisk Institutt, Working Paper 747, 2008), p. 19.

3.3.2. Counter-jihadist politics in the mainstream and on the streets

Under the new government which is no longer dependent on the DF, and in the aftermath of the Breivik case in Norway, civil servants say it has become easier to focus attention on the far-Right. They believe the old neo-Nazi networks have become marginal in the far-Right milieu and are finding it difficult to recruit new members. Counter-jihadist groups, on the other hand, are thriving, although there appears to be a high turnaround of supporters. The Danish Defence League, modelled on the EDL and based in Aarhus, is growing and was able to host the attempt to launch a European Defence League there in March 2012. With a similar politics, Stop the Islamisation of Denmark has organised several demonstrations in Denmark against 'Islamism' and against mosques. Its founder, Anders Gravers, is also leader of the Stop the Islamisation of Europe organisation which seeks to co-ordinate similar efforts across Europe.

The *Vederfølner*, an anti-immigrant, anti-Muslim group of around 200 to 300 members, has also gained ground and is strong in Aarhus. It includes ex-members of White Pride, a neo-Nazi group linked to football supporters. The newly formed *Danskernes Partie* (Danes' Party) also includes former members of neo-Nazi parties and focuses on Islam. The narrative of these street-based far-Right groups appears to be a more radical form of the counter-jihadist narrative of the DF. It is unclear whether the presence of counter-jihadist politics in the mainstream of the political process for at least the last ten years has helped these groups to legitimise themselves to wider audiences, or whether the success of the DF has dragged support away. It is possible that, with the DF no longer playing a supporting role in government, the attraction of the street-based far-Right will increase.

Denmark has also been an intellectual centre for the international counter-jihadist movement. A key organisation has been the Danish Free Press Society, founded by Lars Hedegaard, which has strong links with the DF, with US neoconservatives and with Geert Wilders. First established in 2005, it has since branched out to other countries under the name of the International Free Press Society since 2009. Its aim is to informally influence public debate, in order to entrench a counter-jihadist narrative, using its connections with leading politicians and journalists. Its board of advisors includes conspiracy theorist Bat Ye'or, Belgian counter-jihadist blogger Paul Beliën, US-based 'sharia' conspiracy theorists Frank Gaffney, Daniel Pipes, Robert Spencer, Allen West and Mark Steyn, and leader of the Dutch PVV, Geert Wilders.

Denmark's integration narrative has involved a high degree of social drama, polarisation and calls for existing 'multicultural' conventions to be swept aside. This trend has, to some extent, carried over to counter-terrorism policy, although there have been counter-currents which have sought to de-link questions of identity and values from counter-terrorism, and recognise a range of different extremisms. However, these counter-currents have run up against the wider prominence of the values-identity narrative in Danish politics, promoted not only by the DF but by other mainstream parties and prominent journalists connected to counter-jihadist networks. With the values-identity narrative dominating Danish public culture, it seems unlikely that the ways in which counter-terrorism is understood will be substantially changed. The prospect of a growing street-based far-Right milieu, fuelled by over a decade of anti-Islam, anti-immigrant rhetoric, is a potential danger for the future, especially if the DF begins to lose electoral support.

3.4. Belgium

3.4.1. Official values-identity narrative asserted and partially reflected in counter-terrorism policy

Like the other European countries considered in this paper, a values-identity narrative of Muslim 'failed integration' has pervaded the Belgian public sphere for at least a decade. From the late 1990s, a policy of *inburgering* (making citizens) came to be premised on the idea that immigrants and their descendants, irrespective of their actual citizenship status, needed to undergo a cultural transformation to acquire support for

European values of gender equality, freedom of expression and secularism.[88] The debates in the Netherlands and France on 'integration' had a strong influence in Belgium, too, which, at around the same time as France, introduced a ban on women wearing the niqab in public. Analyst Sami Zemni notes that, over the last decade:

> '[T]he idea of a failure of integration – presumably caused by Islam's rejection of, or resistance to, modernity – emerged. The idea of a failure of integration and multiculturalism is now widespread across the political spectrum and is popular with opinion leaders and other intellectuals. … The proponents of this thesis share the conviction that the root causes of specific societal problems are to be found in the cultural permissiveness and naive cultural relativism that lay at the heart of Belgium's multicultural policy and that this is what made Islam thrive.'[89]

Belgium has developed a number of initiatives to respond to the security implications of this perceived failure of integration. A recurring pattern has been a government search for 'moderate Muslims' who could partner with the state to counter extremism while, at the same time, acting to undermine the possibility of such partners emerging. The Belgian case illustrates the tendency of policies based on a values-identity narrative to both call for 'moderate Muslim' partners, defined by their embrace of European values, while also paradoxically casting suspicion on whether those partners actually subscribe to these values.

Uniquely among western European nations, Belgium has sought to foster the creation of a national representative body of Muslims, with officially administered elections to decide its leaders. This Belgian Muslims' Executive (EMB) was to become the key partner organisation for the Belgian state in its efforts to integrate Muslims and prevent extremism. However, the partnership was not entirely equal. In 2004, Justice Minister Laurette Onkelinx demanded fresh elections to the EMB because of concerns that elected leaders were 'extremists' who did not embrace European values, and a system of vetting of candidates by state security was initiated.

The emergence in Antwerp of a genuinely popular youth movement, the Arab European League (AEL), prompted similar concerns. Led by its charismatic leader Dyab Abu Jahjah, the AEL advocated a proud pan-Arab and Islamic identity inspired by Malcolm X rather than the Islamist tradition. Following the racist murder of a teacher, the AEL established local self-defence patrols. The authorities responded by attempting to criminalise the movement, accusing its leaders of links to Hizbullah and of inciting street violence. It took five years before they were acquitted, after which a demoralised Abu Jahjah decided to leave Belgium for Lebanon. During this time, newspaper articles published a series of false claims that he was involved in money laundering and child pornography, and AEL leaders' bank accounts were frozen. A movement that was able to articulate the political demands of a marginalised section of Belgian society, and present its own version of 'integration', was broken. As Sami Zemni notes: 'What should have been seen as a good example of integration, a well-organised group democratically demanding members' rights within a participatory public sphere, was thus criminalised and depicted as proof of the failure of integration.'[90] Some observers argue that the criminalisation of the AEL left a vacuum that has now been filled by radical Salafi groups such as 'Sharia4Belgium', modelled on Britain's al-Muhajiroun and its successor networks.

More recently, the Belgian government has developed 'a comprehensive anti-radicalisation plan' which, as in the other countries examined in this paper, seeks to establish partnerships between municipal authorities, local police, social workers, schools and 'moderate Muslims' to combat different forms of extremism. According to Rik Coolsaet and Tanguy Struye de Swielande:

> 'Belgian authorities have gone to great lengths to make clear that Islam is not being targeted, but instead extremist and terrorist acts of whatever origin or justification. The plan concentrates on

[88] Sami Zemni, 'The shaping of Islam and Islamophobia in Belgium', *Race & Class* (Vol. 53, no. 1, 2011), p. 31.
[89] Ibid., pp. 37–8.
[90] Ibid., p. 37.

> the surveillance of violent, racist, xenophobic and anti-Semite messages, conveyed through the internet, radio, television programmes, imams, cultural centres, propaganda and groups.'[91]

Alongside, the Belgian federal police has developed a training programme for police officers that aims to help them recognise signs of radicalisation 'signals', whether it be in the form of 'Islamist', right-wing, left-wing or animal rights extremism. This Community Policing and Prevention of Radicalisation (COPPRA) programme began as a Belgian initiative before becoming an EU-wide project. These initiatives seem to recognise an equal threat of far-Right violence and appear to avoid explicitly narrating their efforts in terms of values and identity, although such conceptions are implicit in the wider political context from which they emerge.

Since 2009, concerns about Salafism have been prominent in the way that the chief of state security, Alain Winants, has narrated counter-terrorism. In September 2009, an Imam from Antwerp, Nordin Taouil, made a statement defending the right of women to wear a veil: 'If you ban the veil then we have no other choice but to open up our own schools.' The statement quickly provoked a public controversy. Interviewed by a television journalist, Winants commented that: 'Monsieur Taouil is an extremist Muslim and a dangerous man.' The journalist responded by asking: 'Is Monsiour Taouil accused of something?' To which Winants replied: 'No, but I can tell you that he is a Salafist, he is an extremist Muslim and a dangerous man.' Taouil's wife lost her accreditation to run a nursery school soon afterwards. Four years earlier, he had failed the security vetting to be a candidate for the EMB, due to his allegedly 'extreme' views.[92]

In 2011, Winants gave another interview in which he stated:

> 'I believe that political Salafism is something that, in the long run, is a greater danger than Salafism of the terrorist tendency. Its destabilising effects are being felt now: women are being spat on because they do not wear the veil in public; a teacher is kicked by a 10-year-old child because she gives a lesson on the theory of evolution and the parents take the child's side; traders are threatened because they sell alcohol […] In some quarters or districts, a completely separate life is led, with schools, a banking system, weddings, shops, separate media […] Moreover, this extremism can beget another.'[93]

These comments suggest an alternative conception of security that helpfully pays attention to the need to protect citizens from the low-level harassment of community agitators. But they also frame these concerns within the familiar narrative of failed integration. Moreover, efforts to tackle these kinds of security issues are hampered by viewing autonomous community initiatives such as the AEL, not as potential partners in a process of political empowerment but as 'extremist' threats, because they do not 'fit' the official security narrative.

3.4.2. Counter-jihadist politics in the mainstream

Far-Right politics in Belgium has been dominated by the *Front National* in the French-speaking region and by the more successful *Vlaams Blok* (Flemish Block) and its successor, the *Vlaams Belang* (Flemish Interest, VB), in Flanders. The neo-fascist *Vlaams Blok* was opposed to the Belgium state and favoured the creation of an independent Flanders with Brussels as its capital, eventually to include the Netherlands and South Flanders (a small area of northwest France). From the later 1980s, its programme focused on immigration, with the slogan '*Eigen volk eerst*' ('Our own people first'). In November 1991, on what came to be known as 'black Sunday', the *Vlaams Blok* achieved an electoral breakthrough, securing 6.6% of the vote (10.7% of the Flemish vote). By 1999,

[91] Rik Coolsaet and Tanguy Struye de Swielande, *Belgium and Counterterrorism Policy In The Jihadi Era, 1986–2007* (Egmont – Royal Institute for International Relations, 2007), p. 22.

[92] *Islamophobia, Human Rights and the Anti-Terrorist Laws*, (Institute of Race Relations, 2011), p. 10; 'Schade-eis van imam Taouil tegen Staatsveiligheid afgewezen', *Gazet van Antwerpen* (7 October 2011).

[93] Marie-Cécile Royen, 'Alain Winants « Le salafisme politique est plus grave qu'un attentat »', *Le Vif-L'Express* (8 April 2011). Author's translation.

the party was achieving 15% of the Flemish vote and had become the largest party in Antwerp, winning 33% of the vote there in the 2000 municipal elections.[94]

Unlike in the other countries considered here, the mainstream political parties in Belgium responded to the rise of the *Vlaams Blok* by declaring a *cordon sanitaire* around the party, refusing to enter into coalitions with it at either the municipal or national level. The strategy has remained in place since it was launched in the 1990s and, though it has been criticised as 'anti-democratic', there is good reason to believe it has been important in blocking the advance of the far-Right. In 2004, anti-racist groups successfully used racial discrimination laws to prevent not-for-profit foundations from funding the *Vlaams Blok*, which effectively rendered it ineffective. The party then had to reconstitute itself with new supporting foundations and give itself a new name – the *Vlaams Belang*.

The VB's leader, Filip Dewinter, has sought to circumvent accusations that his party is fascist and racist by reinventing it as a counter-jihadist movement focused on Islam. By building links with the Israeli Right, he has even succeeded in gaining the support of a minority of Antwerp's Jewish voters, despite the party's roots in anti-Semitism and neo-Nazism: in 1988, Dewinter paid his respects to the tens of thousands of Nazi soldiers buried in Belgium and, in 2001, he opened a speech with an oath used by the SS.[95] There continue to be old-style racists and anti-Semites among VB activists. But counter-jihadism has substituted for anti-Semitism as the chief mobilising message and Dewinter makes visits to Israel to meet right-wing members of the Knesset. In 2005, he told the Israeli newspaper *Ha'aretz*:

> 'Islam is now the No. 1 enemy not only of Europe, but of the entire free world. After communism, the greatest threat to the West is radical fundamentalist Islam. There are already 25–30 million Muslims on Europe's soil and this becomes a threat. It's a real Trojan horse. Thus, I think that an alliance is needed between Western Europe and the State of Israel. I think we in Western Europe are too critical of Israel and we should support Israel in its struggle to survive. I think we should support Israel more than we do because its struggle is also very important for us.'[96]

While the *cordon sanitaire* has offered some resistance to the promotion of this new far-Right narrative, it nevertheless overlaps with much mainstream discourse. Belgian social democrats and liberals tend not to mount a definitive riposte to such views, and Islamophobic rhetoric can be found even in quality newspapers. For example, Wim van Rooy, author of the bestseller *The Malaise of Multiculturalism*, wrote in an op-ed article for *De Standaard* that: 'Muslims are people like non-Muslims, but they are conditioned to hostility towards non-Muslims by the ideology that Mohammed captures in the Qu'ran.'[97]

Faced with recent competition from the *Nieuw-Vlaamse Alliantie* (New Flemish Alliance) party, which is picking up its Flemish nationalist agenda, while softening its outright racism, the VB has become more militant and provocative in its Islamophobia. A recent campaign featured Dewinter's 19-year-old daughter, An-Sofie, posing in a niqab and bikini. The poster for her father's 'Women against Islamisation' campaign urged women to choose between freedom and Islam, and appeared on billboards on the streets of Antwerp.

Within the ranks of the VB, there remains a more militant strand, the *Vlaamse Militanten Order* (Flemish Militant Order). Though officially illegal, it is said to still function with a small number of members who appear on VB demonstrations.[98] In 1993, the group sent a letter to various newspapers, warning that they plan to take unspecified action against their opponents. Following a police raid on a fascist cafe in Antwerp, left-wing centres were attacked by the far-Right.

Belgium is also home to a leading outlet of counter-jihadist online propaganda, *The Brussels Journal*, edited by Paul Beliën, the husband of a leading figure in the VB. Breivik's manifesto took its title from an article by

[94] Marcel Lubbers, *Exclusionistic Electorates: extreme right-wing voting in Western Europe* (Katholieke Universiteit Nijmegen, 2001), p. 14.
[95] Adar Primor, 'The unholy alliance between Israel's Right and Europe's anti-Semites', *Ha'aretz* (12 December 2010).
[96] Adi Schwartz, 'Between Haider and a hard place', *Ha'aretz* (28 August 2005).
[97] Wim Van Rooy, 'Islamofobie en waar ze vandaan komt', *De Standaard* (29 January 2009).
[98] Interview with analyst of far-Right, Belgium, 23 April 2012.

'Fjordman' published on this website, which defined 'multiculturalists' as traitors to Europe and called for 'resistance'.[99] Beliën is a Catholic conservative with strong links to counter-jihadist movements in the US, the Netherlands and Scandinavia. In 2006, he wrote an op-ed article for *De Standaard*, entitled 'Give us arms', in which he wrote that: 'Muslims are predators who have learned from childhood … during the yearly feast of the sacrifice … how to slaughter warm herd animals.'[100] The article was a response to the robbery and murder of a young schoolboy, Joe van Holsbeek, in Brussels, initially thought to have been carried out by North Africans. It was later discovered that the perpetrators were in fact Polish.

4. Conclusions and recommendations

The foregoing has demonstrated that, to varying degrees, Britain, the Netherlands, Denmark and Belgium have narrated their counter-terrorism efforts according to a framework of values, identity, Muslim generational crisis and high social drama. In each case, security has been largely understood through a lens in which 'Islamist' terrorism is the primary threat, its cause has been taken to be a culture of extremism within Muslim communities, facilitated by multicultural policies that have undermined European values; in response, a stronger assertion of liberal values against extremism has been called for, brushing aside what are perceived to be conventions of political correctness and naïve tolerance of cultural difference.

The communication of this narrative of the 'Islamist' terrorist threat has had two consequences. First, security practitioners have tended to neglect the danger of far-Right violence, failing to take the threat seriously in their analyses and not allocating sufficient resources to countering it. While the jihadist threat is seen as 'strategic', the far-Right threat is regarded more as a public order problem, a problem of 'lone wolves' or disturbed individuals; governments have thus absolved themselves of a broader reflection on the social and political contexts from which far-Right violence draws its sustenance. Whereas the murder of Theo Van Gogh, for example, was taken to be symbolic of a wider problem with young Dutch Muslims, murders carried out by the far-Right have been seen as one-offs that are not indicative of social issues.

With the Breivik case and groups like the EDL, we see a trend of groups and individuals who have appropriated the official narrative of the 'war on terror' and chosen to open a domestic 'front' against fellow citizens. With the prevalence of a similar values-identity narrative in its thinking, the counter-terrorism system has provided the far-Right with an enabling environment and is itself in danger of becoming an unintentional ally of the new counter-jihadist movements.[101] The proximity of these new far-Right narratives to official security discourse means they occupy a blind spot in the vision of the counter-terrorism system. While the propaganda of classic neo-Nazi groups is easily condemned by everyone, the emergence of the counter-jihadist far-Right, overlapping with mainstream politics of all shades, is in general publicly accepted.

In the following, a series of recommendations are made to address these problems. It is beyond the scope of this paper to examine the wider social causes of far-Right violence; as such, the recommendations presented here are principally focused on reconceptualising security threats and how they are communicated by politicians and officials.

4.1. A new approach to assessing security threats

It is time to engage in a process of rethinking security from an objective and neutral standpoint. The cursory survey of deaths resulting from far-Right violence since 1990, presented in Annex 1, suggests that, in Europe as a

[99] Fjordman, 'Native revolt: a European declaration of independence ', *The Brussels Journal* (16 March 2007), http://www.brusselsjournal.com/node/1980.
[100] Sami Zemni, 'The shaping of Islam and Islamophobia in Belgium', *Race & Class* (Vol. 53, no. 1, 2011), p. 38.
[101] Luk Vervaet, *Le Making-Of D'Anders B. Breivik: Oslo-Utøya 2011: islamophobie et sionisme, les nouvelles guerres de l'extrême droite* (Egalité Editions, 2012).

whole, there is no reason to elevate the harm of jihadist violence beyond that of far-Right violence – both have taken about the same number of lives in Europe.[102]

What reasons might be given for conceptualising jihadist violence to be a fundamentally different order of threat from the far-Right, despite involving a similar level of murderous violence? It may be that the threat of jihadism is considered more serious, because it overlaps with strategic military interests overseas – for example, British troops combating the Taliban in Afghanistan. The problem with this is that the stated reason British troops are in Afghanistan is to reduce the threat of terrorism in the UK, so the argument is circular. Alternatively, it might be argued that the 9/11 attacks, though they did not occur in Europe, nevertheless demonstrated the willingness of jihadists to carry out violent attacks against civilians in Western cities on a much larger scale than the far-Right. However, while this argument might have been plausible ten years ago, the tactical differences between al-Qaeda and the far-Right have since diminished, as the Breivik case demonstrates, and especially as jihadists have focused more recently on low-level targets. Finally, it could be held that jihadist terrorism warrants a higher priority because it is an international problem, whereas the far-Right is solely a domestic matter. Yet the far-Right thrives through multiple international connections across Europe, the United States and elsewhere.

Even before the Breivik case, terrorism studies scholars Robert Lambert and Jonathan Githens-Mazer made a similar argument about the need to take the far-Right threat more seriously, specifically in relation to the UK:

> 'Arguments we have heard from politicians and public servants involved in Prevent policy that the threat from violent extremist nationalists in the UK is local and lesser when compared to the al-Qaeda threat which is global and greater are not compelling now and likely to become less so during this new decade as it unfolds. In fact, the evidence is already sufficiently clear to conclude that violent extremist nationalists in the UK take inspiration from propaganda that is every bit as global in nature as that which promotes al-Qaeda. More importantly, violent extremist nationalists in the UK have a present capacity to inflict death and destruction on a scale that is broadly comparable to their UK counterparts who are inspired instead by al Qaeda. Whereas the latter group sometimes have links to al-Qaeda affiliates or franchises in countries in the Middle East, Gulf and South East Asia that may assist them in terrorist training so too can members of the former group sometimes rely on long-standing links to violent extremist nationalists in countries in Europe, Scandinavia and North America.'[103]

Ultimately, the decisive difference between the far-Right and jihadist threats is not the harm they are each capable of inflicting on the people of Europe, or the geographical spread of their activities, but the fact that jihadist movements are using violence to radically oppose the foreign policies of European governments, whereas far-Right groups are using violence to pressure for demographic and cultural changes to European societies. It is for this reason that the former is considered a 'strategic' threat whereas the latter is considered a 'public order' threat. Yet this distinction is only valid if one holds foreign policies to be more sacrosanct than the rights of minority ethnic citizens. If one takes the preservation of the constitutional democratic order as the baseline for defining security threats, then violence aimed at removing the rights of minorities is at least as serious a threat to the fundamental well-being of European societies as violence aimed at opposing foreign policies. But neither jihadist nor far-Right violence represent strategic threats to the survival of European democracies in their current forms. The dramatic, fear-inducing 'crisis' rhetoric of much counter-terrorism discourse of the last ten years has been unwarranted and unhelpful.

A more objective approach to counter-terrorism would move away from the current state-centred agenda, and shift to an approach that calibrates threats on the basis of how politically motivated violence

[102] Of course, this comparison does not take account of a significant number of jihadist terrorist plots that have been intercepted in Europe since 1990.

[103] Robert Lambert and Jonathan Githens-Mazer, *Islamophobia and Anti-Muslim Hate Crime: UK case studies 2010* (University of Exeter and European Muslim Research Centre, 2010), p. 79.

impinges on the people of Europe in their everyday lives. With such an approach, the concept of social solidarity would be more important than national identity in narrating security, and issues such as organised racist violence or the intimidation of women by community agitators would attract more visibility.

4.2. Recognise racist violence as potentially a form of terrorism

Minority ethnic communities victimised by far-Right violence contend not only with occasional spectacular campaigns of violence directed at them, such as the David Copeland nail-bombing campaign in London in 1999, but also with ongoing low-level harassment which inflicts a different but no less powerful form of terror. There are strong arguments for considering all racially motivated violence as a kind of terrorism; it certainly fits the standard definition of terrorism as violence aimed at instilling fear in a population to advance a political cause (in this case, the preservation of a racially unequal society or the creation of an ethnically homogenous society). Analyst Randy Blazak has written that terrorists and perpetrators of hate crimes work in the same way. They select their targets randomly:

> 'to make a political, social, or religious point (workers in the World Trade Centers or black residents in a White neighborhood) and actors (who are often anonymous) violate the law in hopes of advancing some larger goal (removing US military bases from Saudi Arabia or getting black residents to move out of White neighborhoods) and send a larger message of fear to the wider community (of Americans or African-Americans). Cross burnings, gay bashings, and spray-painted swastikas are all designed to send a terroristic message.'[104]

Blazak points out that the Ku Klux Klan was the first group recognised as a terrorist organisation by the US federal government (although effective action against the Klan did not occur until relatively recently).

From this perspective, racial violence would appear to be a deep and perennial problem of terrorism in European societies that certainly matches in scale other categories of threat. In fact, the majority of perpetrators of racist violence are not directly affiliated with far-Right groups or networks.

In addition, violence motivated by Islamophobia should be recognised as a form of racism, analogous in many ways to anti-Semitism, and combated forcefully. A first step is the compilation of robust data on racial and Islamophobic violence, if necessary by properly resourced independent monitoring groups, backed up with research on the impact of such violence.

4.3. Recognise the complexity of far-Right ideology

It would be possible for national security practitioners to accept that the threat of far-Right violence is as serious as that of jihadism but nevertheless argue that the same levels of attention and resources are not needed. For example, it could be claimed that, whereas governments need to raise awareness of jihadist ideology in order that civil society can recognise and effectively counter it, no such initiative is required in the case of far-Right violence because European civil society is already familiar with far-Right ideology and actively involved in challenging it, so further intervention from government is unnecessary. Hence, European civil society 'self-corrects' for its far-Right movements but does not do the same for jihadism. This appears to have been an argument made in some Dutch counter-radicalisation settings.

But this simplifies far-Right ideology. While many believe that they 'know racism when they see it' or that neo-Nazi ideology is straightforward to identify, in fact, organised racism and far-Right politics present at least as many problems of definition and recognition as jihadist ideology. Far-Right politics has not remained static since the post-war period; with the counter-jihadist turn of far-Right ideology, and the significant overlaps of this strand with mainstream views, the problems of definition and recognition have become yet more challenging. For example, Europol's 2012 *EU Terrorism Situation and Trend Report*, declines to categorise Breivik's attacks as

[104] Randy Blazak, 'Isn't every crime a hate crime?: The case for hate crime laws', *Sociology Compass* (Vol. 5, no. 4, 2011), p. 248.

'right-wing terrorism', suggesting instead that they were motivated by 'a personal mix of elements from different ideologies'. Europol defines right-wing extremism in terms of neo-Nazism and, presumably for this reason, Breivik does not fall under this category.[105] Yet Breivik is clearly a right-wing terrorist. That the attempt to define Breivik's motivation generates these conflicting official responses indicates there is no consensus on the nature of the far-Right. To believe European civil society conceptually well-equipped to counter far-Right ideology is rather complacent. On the contrary, there is a need for much greater awareness of the complex and changing nature of far-Right movements.

4.4. Embrace a new security narrative

Official security narratives need to be reworked to avoid the performative effects on far-Right movements that have been highlighted in this paper and break the triangle of mutually reinforcing narratives illustrated in figure 3. To achieve this, official narratives that talk of 'war' against 'Islamist extremism', that imply Muslims pass a test of their values before they are considered equal citizens, that see Muslims as presenting a kind of cultural threat to European societies need to be abandoned.

The construction of three 'protagonists' in the prevailing security narrative – 'us', 'moderate Muslims' and 'extremist Muslims' – is unhelpful. Firstly, it positions the 'we' that is leading the fight against terrorism as the non-Muslim majority, while Muslims themselves have to prove they are on the right side by demonstrating they share 'our' liberal values. Secondly, by introducing the concept of national values, it unhelpfully confuses the question of violence with the question of cultural identity.

Counter-terrorism is better narrated by avoiding the distinction between 'moderate' and 'extremist' Islam and instead distinguishing between those who carry out or support acts of violence to advance their cause, and those who do not. The 'we' for this narrative is all those who oppose violence against fellow citizens. The counter to this violence is not an abstract set of 'values' associated with national, European or 'Western' identity but democratic participation and social solidarity. On this view, multiculturalism – as the principle that everyone has the right to full political participation irrespective of their perceived cultural values – is an asset in preventing violence rather than a barrier to be swept aside. An official security narrative on these lines would do a far better job of denying discursive opportunities to the far-Right and minimising the blind spot in Europe's optics of counter-terrorism.

This does not imply that we think of terrorists as 'lone wolves' without enabling environments. On the contrary, it draws attention to the wider conditions in European societies which encourage support for far-Right violence – the demonising rhetoric of the mainstream counter-jihadists who deny Muslims the equal right to shape the societies of which they are a part, the longer legacies of institutional racism in Europe, and the impact of ten years of 'war on terror' rhetoric.

[105] Europol, *TE-SAT 2012: EU terrorism situation and trend report* (2012), pp. 4, 9, 29, 43.

5. Annex – major incidents of far-Right violence in Europe since 1990

In this section, a number of major incidents of far-Right violence or threatened violence in Europe are listed by country.[106] Based on the following cases, it can be provisionally estimated that 249 persons have been killed in Europe as a result of far-Right violence since 1990. For the purposes of this survey, Europe is defined to include the countries that are currently members of the European Economic Area.[107] Cases were included in this count only if all the following conditions were met:

- the incident was reported in a mainstream newspaper or newswire;
- the perpetrator was clearly affiliated with far-Right politics;
- the incident was politically or racially motivated, rather than arising from some other dispute.

The actual number of persons who have died as a result of far-Right violence is likely to be higher than the figure given here, due to reporting that is vague about motive and political affiliation. For example, over this period, there were dozens of Roma victims of racist murders by gangs in Bulgaria, the Czech Republic, Hungary, Romania and Slovakia. However, hardly any of these incidents have been included, as newspaper reports tend to describe the perpetrators simply as ‘skinheads’, which, while suggestive of far-Right political views, is insufficient to ascribe a clear affiliation to far-Right networks or groups.

Finally, it should be remembered that the number of persons who have died as a result of racially motivated violence in Europe is higher than the number given here because there are many racist murders carried out by people who are not clearly affiliated to far-Right politics. In Britain, for example, most racist murders are carried out by individuals who are not linked to far-Right groups or networks.

5.1. Austria

In 1993 a letter-bomb campaign was launched, which over the course of four years led to four Roma being killed and a dozen persons injured, including the Social Democrat mayor of Vienna. The perpetrator, Franz Fuchs, was a neo-Nazi who targeted minorities and those supporting them.[108]

5.2. Belgium

In Antwerp in May 2006, Belgian far-right activist Hans Van Themsche went on a racist killing spree, first murdering a Malian woman and then the young girl she was looking after, before shooting at a Turkish woman, who was sitting on a nearby bench.[109]

5.3. Britain

The neo-Nazi David Copeland carried out a nail-bombing campaign in 1999, targeting African-Caribbean, Asian and gay heartlands in London. He killed 3 people in an attack on the Admiral Duncan pub in Soho and injured more than 100.[110]

The Institute of Race Relations has documented 109 killings in the UK between 1991 and 2011 with a suspected racial element, although only one of these appears to have been carried out by someone linked to the

[106] The work of Liz Fekete and her colleagues at the Institute of Race Relations’ European Race Audit was an indispensable resource for collating the material in this section.

[107] Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, France, Finland, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Romania and the United Kingdom.

[108] ‘Austrian “Unabomber” convicted of murder commits suicide, police say’, *Agence France Presse* (26 February 2000).

[109] ‘Life for racist murderer of nanny and girl’, *The Times* (12 October 2007).

[110] ‘Racist nail bomber found guilty of murders’, *Associated Press* (30 June 2000).

far-Right: in 2000, neo-Nazi Robert Stewart murdered Zahid Mubarek in Feltham Young Offenders Institute, west London, in a racially motivated attack.[111]

As of June 2011, there were seventeen people serving prison sentences for terrorism-related offences who are known to have been associated with far-Right groups.[112] These include Terence Gavan, who was jailed in 2010 for assembling 54 explosive devices, including nail bombs, pipe bombs and a booby-trapped cigarette packet, as well as 12 firearms, in preparation for what he considered to be an upcoming 'race war',[113] and the white supremacist Neil Lewington, jailed in 2009 after his bomb-making factory was discovered by chance following his abuse of a train conductor.[114] In the summer of 2010, police arrested a 16-year-old in Tamworth who had constructed a viable nail-bomb and possessed literature from the BNP and EDL, together with Nazi emblems.[115]

5.4. Czech Republic

A 2012 interior ministry report estimates there are 4,000 militant neo-Nazis, who, according to experts, are turning to terrorist campaigns under the influence of far-Right movements in Germany, Italy and Russia. Neo-Nazi gangs have gained access to weapons and firearms training by infiltrating the police force and private security firms, obtaining military-grade explosives. In the six months up to March 2012, there were 23 reported attacks on Czech Roma, which left three people dead.[116]

5.5. Denmark

In 1997, Danish police arrested seven neo-Nazis accused of organising an international letter-bomb campaign. One was also charged with the attempted murder of a policeman.[117] The Danish *Politiets Efterretningstjeneste* (Police Intelligence Service) assessed in November 2011 that 'part of the far-right community is preparing for a future race war in Denmark and in that context is willing to use violence'.[118]

5.6. France

In 1995, 17-year-old Ibrahim Ali was shot dead in Marseilles when he confronted *Front National* members putting up posters.[119] Later that year, four skinheads who had come to Paris for a *Front National* rally killed 29-year-old Brahim Bouraam by pushing him into the Seine.[120] In July 1995, Philippe Vignaud, an activist in the far-Right *Parti nationaliste français et européen*, and Vincent Parera, a sympathiser, were convicted of kidnapping and murdering a Toulouse car dealer, Guy Levy. They had targeted Levy because of his Jewish name and assumed he was among industrialists who had brought North African immigrants to France in the early 1960s.[121]

5.7. Germany

A study by the newspapers *Tagesspiegel* and *Die Zeit* found that 137 people were killed by right-wing extremists from 1990 to 2010, almost three times the official figure of 47, which only counts cases where the motivation has

[111] Cahal Milmo, 'Prison killer sent boasting letters to racist gangmates', *Independent* (20 November 2004). http://www.irr.org.uk/news/deaths-with-a-known-or-suspected-racial-element-1991-1999/; http://www.irr.org.uk/news/deaths-with-a-known-or-suspected-racial-element-2000-onwards/.

[112] HM Government, *Prevent Strategy* (June 2011), p. 15.

[113] 'Bomb cache bus driver Gavan "obsessed with weapons"', *BBC News* (15 January 2010), http://news.bbc.co.uk/2/hi/uk_news/8462205.stm.

[114] Sean O'Neill, 'White supremacist Neil Lewington guilty of "non-British" bomb plot', *The Times* (16 July 2009).

[115] Nick Britten, 'Boy made nailbombs with chemicals bought on eBay', *Daily Telegraph* (26 June 2010).

[116] 'Some 4000 militant neo-Nazis operate in Czech Rep – study', *CTK National News Wire* (1 March 2012); Pavol Stracansky, 'Czech Republic: neo-Nazis taking to terror', *IPS - Inter Press Service* (9 March 2012).

[117] Jan M. Olsen, 'Seven neo-Nazis charged with terrorism after police raids', *Associated Press* (18 January 1997).

[118] 'Danish right-wing extremists eye "race war": police', *Agence France Presse* (17 November 2011).

[119] 'Ibrahim Ali, 17, shot dead', *Independent* (23 February 1995).

[120] 'French rightwingers jailed for drowning Moroccan', *Agence France Presse* (16 May 1998).

[121] Nivelle Pascale, 'La virée meurtrière d'un Hitler en chambre. Avec un autre militant fasciste, Vincent Parera est jugé pour meurtre', *Libération* (24 June 1998).

been established beyond doubt by a court.[122] In late 2010, the Office for the Protection of the Constitution, Germany's domestic intelligence agency, estimated there to be roughly 25,000 right-wing extremists in Germany, of which a third are prone to violence.[123]

5.8. Greece

In May 2011, far-Right activists in Greece launched a series of pogrom-like attacks on immigrants in the downtown Athens area, leaving one person dead and dozens injured.[124]

5.9. Hungary

Six men, mainly far-Right activists, are currently on trial for the murder of six Roma. For a year from July 2008, they allegedly committed twenty attacks in small towns and villages in central and eastern Hungary.[125]

5.10. Italy

In Italy, far-Right activist Gianluca Casseri gunned down two Senegalese street vendors and wounded three others in Florence in December 2011.[126] Casseri had links to the far-Right *Casa Pound* organisation, a neo-fascist youth movement founded in Rome in 2003, which mobilises through cultural, sporting and musical activities and takes its name from the poet Ezra Pound, known for his anti-Semitism and support for Mussolini.

5.11. Netherlands

In the Netherlands, the *Anne Frank Stichting* counted 148 incidents of 'racial and right-wing extremist violence' registered by the National Police Services Agency and anti-discrimination bodies in 2009 – almost certainly an under-estimate of the actual total, due to under-reporting.[127]

5.12. Norway

In July 2011, Anders Behring Breivik detonated a car bomb in Oslo before travelling to the island of Utøya and carrying out a mass shooting at a Labour Party youth camp, leaving 77 dead.

Two young Norwegian neo-Nazis were convicted in 2002 of stabbing to death 15-year-old Benjamin Hermansen in Oslo in a racially motivated attack.[128] In 1997, five neo-Nazis were arrested on suspicion of plotting to assassinate leading Norwegian public figures and attack national institutions.[129]

5.13. Spain

In 2007, 16-year-old anti-fascist protestor Carlos Javier Palomino was stabbed to death in a Madrid metro station by a Spanish neo-Nazi activist.[130] Earlier, in 1992, Lucrecia Perez was killed by four hooded men who burst into an abandoned Madrid disco and opened fire on her and two other Dominican immigrants. Luis Merino, a policeman accused of firing the shots that killed Perez and seriously injuring one of her companions, had Nazi flags and fascist literature among his belongings.[131]

[122] Frank Jansen, Heike Kleffner, Johannes Radke and Toralf Staud, 'Eine furchtbare Bilanz', *Die Zeit* (16 September 2010).

[123] Barbara Hans, Benjamin Schulz, and Jens Witte, 'Facts and myths about Germany's far-Right extremists', *Spiegel Online* (18 November 2011), http://www.spiegel.de/international/germany/0,1518,798682,00.html.

[124] Asteris Masouras, 'In Greece, wave of racist attacks on immigrants in Athens', *Ground Report* (13 May 2011).

[125] Nicholas Kulish, 'Stalked by killers, Hungary's Roma live in fear', *International Herald Tribune* (April 27, 2009).

[126] Nick Squires, 'Florence street vendors shot dead by lone gunman', *Daily Telegraph* (13 December 2011).

[127] Jaap van Donselaar and Peter R. Rodrigues, eds, *Racism and Extremism Monitor, Eight Report* (Anne Frank Stichting / Leiden University, 2008), pp. 8, 17.

[128] 'Appeals court sentences two neo-Nazis in racial killing that shocked Norway', *Associated Press* (4 December 2002).

[129] 'Neo-Nazis arrested in suspected plot to kill Norwegian leaders', *Associated Press* (12 April 1997).

[130] Pilar Álvarez, 'Puñalada mortal en el metro de Madrid', *El Pais* (10 May 2009).

[131] Adela Gooch, 'Murder trial highlights rise of racism in Spain', *Guardian* (11 June 1994).

5.14. Sweden

Peter Mangs was charged in 2010 with conducting a string of racist shootings over a seven-year period. The three murders and ten attempted murders in Malmö bore a chilling similarity to the case of the 'Laserman' gunman who carried out a series of racially motivated shootings in Stockholm in the early 1990s, leaving one person dead and ten others wounded.[132]

In 1999, a trade unionist, Bjorn Söderberg, was shot dead outside his Stockholm apartment by neo-Nazis. In the same year, three neo-Nazis were accused of killing two policemen during a bank robbery intended to finance their political activities.[133] In 1997, Swedish police arrested a 23-year-old member of a neo-Nazi group accused of sending a letter bomb to Justice Minister Laila Freivalds.[134]

6. Acknowledgements

The author would like to thank Tijtske Akkerman, Edwin Bakker, Frank Bovenkerk, Beatrice De Graaf, Liz Fekete, Halim el Madkouri, Renée Frissen, Peter Hervik, Omar Ramadan, Luk Vervaet, Jaap van Donselaar, Rob Witte and Sami Zemni for assistance in the preparation of this paper.

[132] 'Racist sniper suspect remanded: Swedish prosecutor', *Agence France Presse* (21 December 2010).

[133] Carol J. Williams, 'Clerk's killing galvanizes Swedes; inspires backlash against neo-Nazis', *The Record* (12 December 1999).

[134] 'Man arrested after letter bomb sent to Swedish justice minister', *Agence France Presse* (15 August 1997).

Bibliography

Akkerman, Tjitske, 'Anti-immigration parties and the defence of liberal values: the exceptional case of the List Pim Fortuyn', *Journal of Political Ideologies* (Vol. 10, no. 3, 2005)

Algemene Inlichtingen- en Veiligheidsdienst, *Jaarverslag 2011* (April 2012)

Ali, Wajahat, Clifton, Eli, Duss, Matthew, Fang, Lee, Keyes, Scott, and Shakir, Faiz, *Fear, Inc.: the roots of the Islamophobia network in America* (Center for American Progress, August 2011)

Bale, Tim, Luther, Kurt Richard, Green-Pedersen, Christoffer, Krouwel, André, Sitter, Nick, 'If you can't beat them, join them? Explaining social democratic responses to the challenge from the populist radical Right in Western Europe', *Political Studies* (Vol. 58, 2010)

Bartlett, Jamie, and Littler, Mark, *Inside the EDL: populist politics in a digital age* (Demos, 2011)

Blazak, Randy, 'Isn't every crime a hate crime?: The case for hate crime laws', *Sociology Compass* (Vol. 5, no. 4, 2011)

Breivik, Anders Behring, *2083 – A European Declaration of Independence* (2011)

Carr, Matt, 'You are now entering Eurabia', Race & Class (Vol. 48, no. 1, 2006)

Cohn, Norman, *Warrant for Genocide: the myth of the Jewish world conspiracy and the Protocols of the Elders of Zion* (Harmondsworth, Penguin, 1970)

Coolsaet, Rik, and de Swielande, Tanguy Struye, *Belgium and Counterterrorism Policy In The Jihadi Era, 1986–2007* (Egmont – Royal Institute for International Relations, 2007)

de Graaf, Beatrce, *Evaluating Counterterrorism Performance* (London, Routledge, 2011)

de Graaf, Beatrice, and de Graaff, Bob, 'Bringing politics back in: the introduction of the "performative power" of counterterrorism', *Critical Studies on Terrorism* (Vol. 3, no. 2, 2010)

de Graaf, Beatrice, *Why Communication and Performance are Key in Countering Terrorism* (International Center for Counter-Terrorism – The Hague, February 2011)

Dutch General Intelligence and Security Service, *From Dawa to Jihad: the various threats from radical Islam to the democratic legal order* (2004)

Dutch General Intelligence and Security Service, *Right-wing Extremism and the Extreme Right in the Netherlands* (March 2011)

Dutch National Coordinator for Counterterrorism, *Countering Violent Extremist Narratives* (2010)

Eatwell, Roger and Goodwin, Matthew J., eds, *The New Extremism in 21st century Britain* (London, Routledge, 2010)

Edgar, David, 'Racism, fascism and the politics of the National Front', *Race & Class* (Vol. 19, no. 2, 1977)

Europol, *EU Terrorism Situation and Trend Report* (2011)

Europol, *EU Terrorism Situation and Trend Report* (2012)

Fekete, Liz, *A Suitable Enemy: racism, migration and Islamophobia in Europe* (London, Pluto Press, 2009)

Fekete, Liz, ed., *Islamophobia, Human Rights and the Anti-Terrorist Laws* (Institute of Race Relations, 2011)

Fekete, Liz, *Pedlars of Hate: the violent impact of the European far Right* (Institute of Race Relations, 2012)

Fortuyn, Pim, *Tegen de islamisering van onze cultuur. Nederlandse identiteit als fundament* (Utrecht, Bruna, 1997)

Front National, *300 Mesures pour la renaissance de la France: programme de gouvernement* (Paris, Editions Nationales, 1993)

Goodwin, Matthew, *Right Response: understanding and countering populist extremism in Europe* (Chatham House, 2011)

Goodwin, Matthew, *The New Radical Right: violent and non-violent movements in Europe* (Institute for Strategic Dialogue, February 2012)

Government of Denmark, *A Common and Safe Future: an action plan to prevent extremist views and radicalisation among young people* (January 2009)

Government of Denmark, *Government Report on Counter-Terrorism Efforts* (May 2011)

Government of Denmark, *Preventing Extremism: a Danish handbook series* (Ministry of Social Affairs and Integration, 2011)

Griffin, Nick, *Who are the Mind Benders? The people who rule Britain through control of the mass media* (London, Steven Books, 1997)

Hervik, Peter, 'Ending tolerance as a solution to incompatibility: the Danish "crisis of multiculturalism"', *European Journal of Cultural Studies* (Vo. 15, no. 2, 2012)

HM Government, *Prevent Strategy* (June 2011)

Jackson, Paul, *The EDL: Britain's 'new far Right' social movement* (University of Northampton Radicalism and New Media Research Group, 2011)

Kaplan, Jeffrey and Bjørgo, Tore, *Nation and Race: the developing Euro-American racist subculture* (Boston, Northeastern University Press, 1998)

Kundnani, Arun, *Spooked: how not to prevent violent extremism* (Institute of Race Relations, 2009)

Kundnani, Arun, *The End of Tolerance: racism in 21st century Britain* (London, Pluto Press, 2007)

Lambert, Robert, and Githens-Mazer, Jonathan, *Islamophobia and Anti-Muslim Hate Crime: UK case studies 2010* (University of Exeter and European Muslim Research Centre, 2010)

Lentin, Alana, and Titley, Gavan, *The Crises of Multiculturalism: racism in a neoliberal age* (London, Zed Books, 2011).

Lubbers, Marcel, *Exclusionistic Electorates: extreme right-wing voting in Western Europe* (Katholieke Universiteit Nijmegen, 2001)

Meret, Susi, *The Danish People's Party, the Italian Northern League and the Austrian Freedom Party in a Comparative Perspective: party ideology and electoral support* (Aalborg University, SPIRIT PhD Series Thesis no. 25, 2009)

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 'Aanbieding visie op integratie' (June 2011)

Mudde, Cas, 'The populist radical Right: a pathological normalcy', *West European Politics* (Vol. 33, no. 6, 2010)

Nesser, Petter, 'Chronology of Jihadism in Western Europe 1994–2007: planned, prepared, and executed terrorist attacks', *Studies in Conflict & Terrorism* (Vol. 31, 2008)

Prins, Baukje, *Voorbij de onschuld: het debat over integratie in Nederland* (Amsterdam, van Gennep, 2004)

Ulrik Pram Gad, 'It takes two to tango: Danish concepts of dialogue as counterterrorism' (Norsk Utenrikspolitisk Institutt, Working Paper 747, 2008)

van Donselaar, Jaap, and Rodrigues, Peter R., eds, *Racism and Extremism Monitor Eighth Report* (Anne Frank Stichting/Leiden University, 2008)

Vervaet, Luk, *Le Making-Of D'Anders B. Breivik: Oslo-Utøya 2011: islamophobie et sionisme, les nouvelles guerres de l'extrême droite* (Egalité Editions, 2012)

Visser, Mare Stijntje, 'Consequences of the PVV: an immigrant perspective', *Re-public: reimagining democracy* (16 January 2012), available online at http://www.re-public.gr/en/?p=5194. Accessed 5 April 2012

Witte, Rob, 'The Dutch Far Right: from "classical outsiders" to "modern insiders"', paper presented at CSTPV Workshop, 16–17 May 2011

Witte, Rob, *Racist Violence in the Netherlands* (European Network Against Racism, 2011)

Wren, Karen, 'Cultural racism: something rotten in the state of Denmark?', *Social & Cultural Geography* (Vol. 2, no. 2, 2001)

Yilmaz, Ferruh, *Ethnicized Ontologies: from foreign worker to Muslim immigrant: how Danish public discourse moved to the Right through the question of immigration* (University of California, San Diego, PhD Thesis, 2006)

Zemni, Sami, 'The shaping of Islam and Islamophobia in Belgium', *Race & Class* (Vol. 53, no. 1, 2011)

Zúquete, José Pedro, 'The European extreme Right and Islam: new directions?', *Journal of Political Ideologies* (Vol. 13, no. 3, 2008).

Intro Breivik Organisations Top dozen players Country guide Breivik trial blog Monitoring HnH site Donate

Introduction

By Nick Lowles

Welcome to HOPE not hate’s report into the ‘Counter-Jihad’ movement. This report is the largest and most comprehensive survey of groups and individuals who comprise the ‘Counter-jihad’ movement to date.

The report covers the right-wing political parties, who are increasingly using anti-Muslim rhetoric to garner votes. It also explores the websites and bloggers who propagate scare stories about Islam. It covers the street gangs, like the English Defence League (EDL), and the like-minded groups they inspire around Europe. It also investigates the funders and the foundations which bankroll the network.

Perhaps most interestingly, it reveals the inter-connections between the different strands of this movement.

The ‘Counter-Jihad’ movement, as we define it, is a broad alliance of people and ideas embracing sections of neo-Conservatives, Christian evangelicals, hard-line racists, football hooligans, nationalists, right wing populists and some former leftists. Some are hard-line, others less so. Some are openly racist, others are not. Few represent anything more than a minority following in the religious or political traditions they claim to represent. But in all cases the rhetoric used, either explicitly or by implication, leads us to question whether the target is merely radical Islam and Islamist extremist groups, or if it goes wider, criticising Islam as a faith and Muslims as a people. In many instances, this criticism leads to hatred.

We believe that criticism of Islam is perfectly acceptable, as it should be of all religions. We believe that people should also speak out against Islamist extremism and those who carry out terrorism and violence in the name of Islam. Indeed, we ourselves have publicly condemned Islamist extremist groups and we will do so with increasing frequency in the further.

What marks the ‘Counter-Jihad’ movement out is that in the name of opposing Islamist extremism they make generalisations about an entire faith and many fail to differentiate between the actions of a few and the vast majority of Muslims who also reject the extremists. Many of those who we include in the ‘Counter-Jihad’ report actually contribute to heightening tensions between communities and the whipping up of fear and suspicion.

In many ways the ‘Counter-Jihad’ movement represents the new face of the political right in Europe and North America. Replacing the old racial nationalist politics of neo-Nazi and traditional far right parties, with the language of cultural and identity wars, it presents itself as more mainstream and respectable. And as we have seen in countries such as Denmark, the Netherlands and Switzerland these new right-wing populist parties, with an anti-Muslim and anti-immigration message, can garner support from far broader swathes of the population than the old-style racist parties.

But the ‘Counter-Jihad’ movement is more than just right-wing populist political parties. As this report shows, the bloggers, radio hosts and journalists are increasingly shaping and poisoning the wider political and media discourse. Over the next few years, as economic hardship bites and insecurity breeds fear of the ‘Other’, the strength and impact of the ‘Counter-Jihad’ movement will probably grow. More work needs to be done in this field and as a result HOPE not hate is establishing a permanent ‘Counter-Jihad’ Monitoring Unit. We will follow these organisations and individuals, produce further reports into their activities and develop the tools we need to defeat them politically.

The ideas of the ‘Counter-Jihad’ movement are largely based around the belief that Islam poses a serious threat to Western civilisation. Many of its adherents also fail to distinguish between the hardline radical Islamists, such as Al-Muhajiroun, and the overwhelming majority of Muslims who reject these extremist views

The 'Counter-Jihad' movement

The ‘Counter-Jihad’ movement
The global trend feeding anti-Muslim hatred
HOPE not hate
www.hopenothate.org.uk

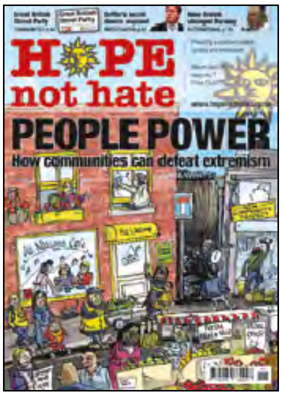
This report is available from HOPE not hate: UK £20 (UK); Europe (£22); Rest of the World (£25) - price includes postage

Click here to order your copy

Share this page

89

HOPE not hate



Click here to subscribe to *HOPE not hate*
6 issues for just £20

and just want to live quietly and in peace. Immigration and multiculturalism are seen by many as the Trojan Horses through which Islam is gaining a foothold in the West.

Their numbers are numerically small but their influence is much bigger. Their anti-Muslim rhetoric poisons the political discourse, sometimes with deadly effect.

Norwegian killer Anders Behring Breivik was inspired by many of the 'Counter-Jihadists' we profile in this report. Perhaps he would have gone on a killing spree without reading their work but it is clear their writings had an important impact on the creation of his political mindset. In his choice of targets it is obvious that he had accepted much of their hatred towards the political establishment.

He too believed that Islam was a threat to Western Europe. He too believed that the immigration and multiculturalist policies of many Western Governments were allowing Islam to go unchallenged and to prosper. He believed all this because he read what they wrote. He read it and he digested it. In his Manifesto he regurgitated it - sometimes word for word.

Almost a quarter of Breivik's 1,518-page Manifesto comprise of quotes from other people - the overwhelming majority from people featured in this report. Half of these 375 pages of quotes came from just one man, the 'Counter-Jihadist' blogger Fjordman.

Yes, Breivik feared and hated Islam, but it was the Norwegian establishment who were his real enemy. More specifically it was social democracy that he and the 'Counter-Jihad' movement blamed for encouraging and promoting immigration and multiculturalism. He bombed Government buildings and he shot young members of the ruling Labour Party.

The 'Counter-Jihadists' were desperate to distance themselves from his actions. Many did so because they were genuinely appalled by what he did. Others were worried about how it would impact on them.

As our report highlights the 'Counter-Jihad' movement is a loose network of bloggers, political activists, street gangs and foundations. Sometimes they act alone, sometimes they join together. The individuals sit on each other's boards and the organisations share platforms and co-host events.

Many of the key players and organisations have never actually met. Some operate under pseudonyms and others do not exist beyond the internet or a blog site.

The 'Counter-Jihad' movement manifests itself in different ways, in different countries, but its underlying message is the same. Sometimes it is focused around the single issue of Islam, but in other situations it becomes interwoven with wider politics of immigration, culture, loss and identity.

In the United States, five state legislatures have already banned Sharia law from being practised. It is being debated by another twenty. In Switzerland, people voted for a ban on Minarets despite the fact that there were only four in the country. In France, politicians of the centre and right tried to outbid each other in the 2012 Presidential election to prove how hardline they are on Muslim practices and extremism. The fear of Islam is playing an increasingly important role in the political discourse in many countries.

In many ways there is a symbiotic relationship between the 'Counter-Jihad' movement and those they claim to oppose. Both point to the other to prove the need for their own existence. On an intellectual level both view the other as proof of the incompatibility of the religions to co-exist. On a street level both use the activities of the other as a recruiting tool. In December 2010 Detective Superintendent John Larkin, of the West Midlands Counter-Terrorism Unit, told the BBC that the activities of the English Defence League were providing recruiting grounds for Islamist extremists.

"In some areas, we have evidence that once they [the EDL] have gone and the high-profile policing of the event has occurred, there's fertile ground for [extreme Islamist] groups who would come in to encourage people to have this reality - this is the way white Western society sees us," he told the BBC. "And that's a potential recruiting carrot for people and that's what some of these radicalisers look for - they look for the vulnerability, for the hook to pull people through and when the EDL have been and done what they've done, they perversely leave that behind.”

In Norway, Anders Breivik’s defence team are calling some of the leaders of Islamist extremist groups, including one man who was recently imprisoned for five years, to give evidence in his court case because these people will say that they want to turn Europe into an Islamic state.

On every level, the ‘Counter-Jihad’ movement is one that we cannot afford to ignore. For this reason we have produced this ‘Counter-Jihad’ report and are establishing the ‘Counter-Jihad’ Monitoring Unit.

Nick Lowles
Chief Executive
HOPE not hate

Updated 24 August 2012 | top |