CHAPTER 2

Silenced by hate? Hate speech as a social boundary to free speech

Audun Fladmoe, PhD, Senior Research Fellow, Institute for Social Research
Marjan Nadim, PhD, Senior Research Fellow, Institute for Social Research

Hate speech is central in discussions of the legal and social boundaries of freedom of speech. On the one hand, any ban on hate speech is a limitation of free speech. On the other hand, hate speech may in itself pose a social boundary on free speech through inciting fear and silencing individuals. Based on a large-scale survey among Norwegian adults, the chapter studies experiences of hate speech and other unpleasant comments in social media, and whether hate speech discourages people from voicing their opinions. The results suggest, first, that people of immigrant backgrounds are more exposed to hate speech directed towards legally protected grounds, but that the majority population are as equally exposed as immigrants to other more general unpleasant comments in social media. Second, the results suggest that more general unpleasant messages may have consequences similar to those of hate speech, in terms of willingness to voice opinions publicly. However,

women and people of immigrant background are more affected by hate speech directed towards legally protected grounds than other groups. The chapter thus demonstrates how hate speech may have negative democratic consequences by silencing certain groups.

Introduction

Free speech and the protection of minorities are not usually incompatible values; nevertheless they can come in conflict. Liberal democracies constantly engage in delimiting the legal boundaries between preserving freedom of speech and combating racism, harassment and discrimination (cf. Bleich, 2011). And hate speech – persecutory, hateful or degrading speech directed towards certain group attributes – is a core issue in discussions of the boundaries of free speech.

Legislating against hate speech and harassment means that some utterances are deemed unacceptable and unlawful. This can be problematic because it entails a constraint on freedom of speech and can potentially limit public discussions through a so called *chilling effect* – i.e. that individuals might be discouraged from engaging in legitimate political debate by threat of legal sanction (cf. Gelber & McNamara, 2015 p. 640). A further argument against hate speech regulation, and in favour of allowing such utterances, is that discriminatory and hateful speech is best met by counter-arguments in public debate. Thus, freedom of speech can in itself be seen as a tool to combat hate speech and discrimination through what can be called the 'cleansing function' of public debate (NOU, 1999 p. 10).

On the other hand, hate speech can have negative consequences for society and the targeted individuals. Allowing hate speech in public debate can contribute to making such rhetoric appear more legitimate and acceptable, paving the ground for even more

hate speech. Furthermore, hate speech can in itself have a discouraging effect on the exercise of free speech. One purpose of hate speech is to incite fear in the groups targeted. Hate speech works to guard and reinforce boundaries and hierarchies between groups, and to remind those who are considered 'different' or 'other' of where they belong (cf. Perry, 2001). Experiences with, or fear of, hate speech can shape individuals' propensity to speak their mind, and make targeted individuals or groups more cautious in expressing their views and making themselves visible. A potential consequence of hate speech is that certain groups are silenced, thereby excluding particular voices and viewpoints from public debates. Thus, while legislation against hate speech poses a *legal boundary* on free speech, hate speech in itself can, in effect, also function to limit the individual's exercise of the right to free speech through instilling fear and causing withdrawal from public debate for those targeted. In this sense, hate speech can represent a *social boundary* for free speech.

The aim of this chapter is to study experiences with different forms of hate speech in social media, and whether such experiences discourage people from expressing opinions publicly. So far, discussions about freedom of speech and hate speech have largely been legal and normative, and there has been a remarkable lack of empirical contributions (Bleich, 2011). This chapter takes a sociological and empirical approach to hate speech, and speaks to the overall theoretical framework of this book by analyzing how hate speech can function as a *social boundary* for the individual expression of opinions, and how these boundaries may be different for different groups. The chapter draws on a large-scale population-based survey with more than 5000 respondents, carried out in June 2016 in Norway. The large number of respondents in the survey enables us to scrutinize variations among different subgroups of the population.

The rest of the chapter is structured as follows: We begin by discussing the concept of hate speech, before reviewing previous research on who are targeted by hate speech, and the potential consequences of hate speech for individuals, groups and society. Next, we present our data and empirical analyses, and finally we discuss the implications of our results.

What is hate speech?

Hate speech is a contested term, and there is no shared definition of the concept (Gagliardone, Gal, Alves, & Martinez, 2015; Gelber & McNamara, 2016). Still, definitions of hate speech typically focus on two key features: the tone or style of the message, and what ground(s) the message is directed towards. Hate speech can be defined as persecutory, hateful, or degrading speech that is directed towards an individual or a group on the basis of certain (perceived) group attributes (Boeckmann & Turpin-Petrosino, 2002; Gagliardone et al., 2015 p. 10; Lawrence III, Matsuda, Delgado, & Crenshaw, 1993). Not all groups are included in the concept; it is usually reserved to cover hateful speech directed towards attributes associated with members of historically oppressed (minority) groups (cf. Lawrence III et al., 1993).

Hate speech reflects negative stereotypes, prejudice and stigma, and is based on perceptions of boundaries and hierarchies between groups. It builds on a rhetoric of exclusion, fear and contempt for individuals and groups that are deemed to be different, and can be seen as a way of 'doing difference' (cf. Perry, 2001). The purpose is to guard and highlight the boundaries between groups, and remind groups and individuals who are seen as 'other' of their rightful place in the social hierarchy (Nilsen, 2014; Perry, 2001). Thus, in understanding and

defining what hate speech is, it is central not only to look at the rhetoric and tone of the message, but also at what *grounds* the speech is directed towards.

Historically there has been a high degree of acceptance of racist expressions and discrimination, but after World War II, and in particular since the 1960s, the general trend is that European countries have brought hate speech under increasingly more stringent regulation; the USA remains one of the very few countries to resist the trend to ban hate speech (Bleich, 2011; Parekh, 2006). National and international legislation employ different definitions of hate speech (Gagliardone et al., 2015). The Norwegian Penal Code section 185 protects against hateful or discriminatory speech about persons or groups of persons because of their a) skin colour or national or ethnic origin, b) religion or life stance, c) homosexual orientation, or d) disability. Thus, in Norway, for an utterance to be defined as hate speech in judicial terms it must be directed towards one of these group-based identities (also referred to as *protected grounds*). This does not imply that hateful utterances directed towards members of others groups are necessarily lawful, but that these must rather be tried in relation to other laws, such as laws on threats, discrimination, defamation, etc. (see Wessel-Aas, Fladmoe, & Nadim, 2016).

In popular debates hate speech is often understood in a broader sense than legal definitions, and the concept is used to refer to a wide spectre of phenomena, from online bullying and aggressive and intolerant statements in public, to racism and threats towards individuals (Gagliardone et al., 2015; Sunde, 2013 p. 42; Waldron, 2012 p. 34). For a sociological approach to hate speech that aims to understand and empirically study the phenomenon, it is fruitful to expand the understanding of hate speech from its strict legal understanding. First, it is

methodologically challenging to restrict empirical studies of hate speech to a legal definition. Second, and more substantially, the distinction between criminal and lawful speech is not clear-cut, and expressions that are not covered by the legal definition can also have negative consequences for individuals and society at large.

The definition of hate speech can be extended from the legal version regarding both of the key features in the definition. First, a broader understanding of hate speech can include other grounds than those protected by law. The grounds that are offered legal protection against hate speech are a reflection of historical struggles for group recognition, but they do not necessarily mirror who is most exposed to hate speech or the consequences such speech has for different groups. The creation of group boundaries and hierarchies is an ongoing process, and there are ongoing debates about whether other groups should be included in definitions of hate speech (Jenness, 2003; Maher, McCulloch, & Mason, 2015; McPhail, 2002). In the popular understanding of hate speech, the term is often not restricted to speech directed towards group attributes at all. Second, a legal understanding of hate speech necessarily needs defined criteria distinguishing expressions that are sufficiently harmful in their tone and style to be considered unlawful from those that are not. Such criteria can be difficult to adhere to in empirical investigations. Third, Section 185 in the Norwegian Penal Code limits its definition of hate speech to speech that is expressed in public or in the presence of others. Empirically, it is also relevant to include direct messages to individuals. We wish to emphasize that while we are arguing for employing a broader understanding of hate speech in empirical research, this is not in itself an argument for expanding the regulation or legal definition of hate speech.

In this chapter, we reserve the term hate speech for hateful expressions that are directed towards potentially vulnerable groups. In the empirical analyses we will measure hate speech using different definitions of the phenomenon, distinguishing between a 'protected grounds' definition that is restricted to hateful speech directed at the grounds protected by the Norwegian Penal Code, and an 'expanded definition' also including other characteristics related to people's identities. We also measure experiences with hateful messages directed towards other types of grounds, further from the conventional understanding of hate speech.

Targets of hate speech

As mentioned above, hate speech is understood as persecutory, hateful and degrading speech directed towards historically oppressed groups or individuals, based on their (perceived) group attributes (cf. Lawrence III et al., 1993). Hate speech can be understood as an expression of prejudice, stereotypes and perceptions of differences and hierarchies between groups (cf. Chakraborti & Garland, 2015; Perry, 2001). Thus, the targets of hate speech are first and foremost members of minority groups. But also more general (majority) group attributes, such as gender, may be targeted.

There has been little empirical research that specifically examines experiences and the prevalence of hate speech. One of the few studies that provides information about which group attributes hate speech is directed towards, is Hawdon and colleagues' (2015) international comparison of experiences with hate speech among young adults in the USA, the UK, Germany and Finland. They asked survey respondents whether they had witnessed hate speech online, and if so, what grounds the hate

speech was directed towards. Hatred towards ethnicity and sexual minorities were the most common forms of hate speech observed in all four countries. Ethnicity accounted for between 48 percent (Germany) and 67 percent (Finland) of the hate speech observed in the four countries. Religion was also high on the list in all the countries. Hatebase, a database that gathers instances of hate speech globally, similarly finds that ethnicity and nationality are the most common targets for hate speech, and indicates that there has been a clear increase in hate speech based on religion and class background (Hatebase, 2016). Hawdon and colleagues' comparative study further found large national differences regarding hate speech directed towards gender. Gender was a much more common ground for the observed hate speech in the UK than in the other three countries included in the study (Hawdon et al., 2015 p. 34).

Based on existing research and the insight that hate speech reflects prejudice, stereotypes and group hierarchies in society at large, our expectation is that people with immigrant backgrounds will be especially at risk for receiving hate speech. Thus, our first hypothesis is:

> H1: People of immigrant background are to a larger extent exposed to hate speech than other individuals.

However, there is a difference between the groups of individuals having the most experiences receiving hate speech, and which grounds the hate speech is directed towards. For instance, studies of individuals' experiences with online harassment and with receiving unpleasant and degrading comments indicate that group differences in exposure to these phenomena are not necessarily large, but that the *content* of the comments received by different groups varies considerably (Midtbøen & Steen-Johnsen, 2016; Pew Research Center, 2014).

Hate speech as a silencing mechanism

Hate speech is found to have a range of consequences for individuals, such as fear and other emotional symptoms, lowered self-esteem, loss of dignity, and withdrawal from the public –both physically and in terms of participation in public debate (Boeckmann & Liew, 2002; Boeckmann & Turpin-Petrosino, 2002; Eggebø, Sloan, & Aarbakke, 2016; Gelber & McNamara, 2016; Herek, Cogan, & Gillis, 2002; Leets, 2002; Midtbøen & Steen-Johnsen, 2016; Pew Research Center, 2014). All instances of hate speech will of course not have these consequences, but the empirical studies demonstrate that hate speech *can* produce such outcomes.

In this chapter we examine one possible consequence of receiving hate speech, namely discouraging people from voicing their opinions publicly. One puprose of hate speech is to incite fear in the groups targeted, and to remind those who are considered 'different' or 'other' of where they belong (cf. Perry, 2001). If hate speech works to silence its targets, it can be seen to pose a social boundary on free speech. Furthermore, if certain groups are more likely than others to refrain from voicing opinions publicly due to experiences with hate speech, hate speech is potentially a democratic problem. A precondition for an enlightened democratic debate is that all group-based interests are represented in public discourse (cf. Phillips, 2009).

The idea that hate speech is more harmful than other types of negative and unpleasant expressions, is part of the rationale for passing legislation against this specific type of speech. Also some researchers hold that hate speech can have more adverse consequences than other types of negative speech (Boeckmann & Liew, 2002; Herek et al., 2002). Boeckmann and Liew (2002) find that hate speech produces stronger emotional responses in the recipients than do other forms of degrading speech. Based

on a study of sexual minorities, Herek and colleagues (2002) argue that even less severe expressions of hostility against minorities can be experienced as traumatic because minorities are very aware of the violence and injustice members of their group have been subject to. The argument is that since hate speech triggers the awareness of belonging to a vulnerable group, it incites more fear than other types of negative speech.

Furthermore, because hate speech is not only directed towards individuals, but is also – intentionally or not – targeted against groups, it can have consequences beyond the individuals targeted. Because the content of hate speech is based on certain group attributes of an individual, publicly expressed utterances also send a signal to other individuals with similar attributes (Bell, 1998; Kunst, Sam, & Ulleberg, 2013; Perry, 2014). For members of a minority group, perceptions of other members' experiences – and consequently knowledge about the risk of being subject to the same oneself – can incite fear, even if they themselves have no personal experiences with hate speech (Gelber & McNamara, 2016 p. 327; Perry, 2001).

Does receiving hate speech discourage people from publicly expressing their opinions? A Norwegian study found that, compared to the majority population, ethnic minorities are substantially more prone to become cautious about expressing their opinions after experiencing harassment. While 19 percent of the majority population reported that receiving unpleasant or degrading comments has caused them to be more cautious in expressing their opinions, 36 percent of respondents with immigrant backgrounds reported the same (Midtbøen & Steen-Johnsen, 2016). However, this study did not examine the significance of the *content* of the messages, i.e. whether hate speech directed towards legally protected grounds have more adverse consequences compared to messages directed towards other grounds.

Based on a review of previous research on the consequences of hate speech, our second hypothesis is:

> H2: Hate speech directed towards legally protected grounds (i.e. skin colour, nationality, ethnicity, religion or life stance, homosexual orientation, or disability) has more adverse consequences, in terms of discouraging the expression of opinions publicly, than other types of negative speech.

Data and method

We rely on a web-based survey, carried out in June 2016 as part of the project *Social Media in the Public Sphere* (SMIPS). The sample consists of 5054 respondents, drawn from TNS Gallup's access panel (response rate: 44.6 %). Members of this panel are recruited by means of random sampling through the National Register, and no self-recruitment is allowed.

A limitation of using survey data to study the prevalence of hate speech is that we rely on subjective assessments. Different respondents may understand what constitutes a 'hateful message' differently. In effect, the empirical results must be interpreted precisely as subjective assessments of hate speech. In the following we describe the variables used in the analyses.

Dependent variable 1: Personal experience with hate speech

In order to assess personal experiences with hate speech, respondents were first asked if they themselves had received hate speech via social media, and second towards what grounds these messages were most often directed. In the survey, 'hate speech' (*hatefulle ytringer)* was defined as

statements that are 'degrading, threatening, harassing, or stigmatizing', but the question did not specify any particular grounds. The term hate speech (*hatefulle ytringer*) does not function as a synonym for racist or discriminatory speech in the Norwegian context, as it more commonly does in the American context. Rather it is predominantly understood as containing very negative expressions, without necessarily being related to an individual's group attributes. Thus, the first question captures what respondents themselves perceive as hate speech in general terms, allowing for hateful utterances beyond the legal definition.

In order to be able to distinguish between different definitions of hate speech, the second question asked what the received hateful statements were most often directed towards. It was possible to select one or more attributes from a list of 13. The list included grounds protected by the Norwegian Penal Code (skin colour, nationality, ethnicity, religion, sexual orientation, and disability), in addition to other potentially relevant attributes such as gender, content of one's argument, political views, etc. In effect we can distinguish between hate speech directed towards grounds protected by the Norwegian Penal Code on the one hand, and other types of unpleasant messages perceived as hate speech, on the other.

Dependent variable 2: Reluctance to express opinions

In order to assess reluctance to express personal opinions publicly, we rely on a follow-up question of whether experiences with receiving hate speech have caused the respondents to be more cautious in public debates: 'After experiencing hate speech, have you become more reluctant to express opinions publicly?'

Independent variables

We include the same set of independent variables across different analyses: gender (female=1), age, education (university/college=1), political left-right orientation (1-11), and propensity to share personal opinions on the internet (1: Never – 4: Often). Additionally, initial inspections of the data suggested that political ideology is not linearly related to receiving hate speech, but rather that the relationship is curvilinear – that people on the far left and far right are most likely to have received what they perceive as hate speech. In order to capture this relationship we include squared transformations of the left-right scale. Finally, we include a dummy variable distinguishing between the majority population and respondents with immigrant backgrounds (=1). Following the definition employed by Statistics Norway (see for instance Egge-Hoveid & Sandnes, 2015), this variable includes both people born abroad and people born in Norway of two foreign-born parents. Unfortunately, we have limited information about the country of origin of respondents with immigrant backgrounds. Descriptive statistics for the independent variables are summarized in Table 2.1.

As shown in Table 2.1 women, young people, and low education levels are somewhat underrepresented in the sample. We therefore employ sampling weights in all analyses.

Results

We present our results in two steps. First, we estimate the number of people who have experienced hate speech, how the number varies according to different definitions of the phenomenon, and how the estimates vary among different subgroups. Second, we explore how different types of hate speech may discourage people from expressing opinions publicly.

Table 2.1. Independent variables. Descriptive statistics.

	Obs	Mean	Std.dev.	Min	Max
Gender (female=1)	5054	0.48	-	0	1
Age	5054	51.83	17.74	15	93
Immigrant background	5054	0.06	-	0	1
High school	5054	0.28	-	0	1
Vocational school	5054	0.15	-	0	1
University/College	5054	0.57	-	0	1
Political left-right orientation	5054	6.19	2.25	1	11
Share opinions on the internet	5054	2.04	0.90	1	4

Source: SMIPS (2016).

Experiences with hate speech

Table 2.2 displays the number of respondents who reported that they had experienced what they perceive as hate speech via social media, and what these messages were usually directed towards. The table distinguishes between respondents with immigrant and non-immigrant backgrounds.

The table shows that 7.2 percent of the full sample reported having received what they perceived as hate speech. However, the results suggest that the content of most of these messages falls outside conventional definitions of hate speech. Most of the reported messages are directed towards the content of one's argument, political standpoint and personality. Fewer respondents mentioned any of the legally protected grounds (ethnicity, nationality, skin colour, religion, disability, and sexual orientation). Each of these characteristics is mentioned by less than 1 percent of the total population. The fact that the majority of messages reported are not directed towards protected grounds, shows that the popular comprehension of the concept of 'hate

Table 2.2. Has received what was perceived as hate speech via social media – and what these messages were directed towards. Percent.

	Non-immigrant background	Immigrant background	All
Total	7.0	10.7	7.2
The content of the argument	2.9	3.2	2.9
Political standpoint	2.7	3.1	2.7
Personality	2.7	2.4	2.7
Appearance	1.1	1.3	1.1
Gender	1.0	2.3	1.1
Occupation	0.7	1.0	0.7
Nationality	0.4	3.5	0.6
Religion	0.5	2.3	0.6
Education	0.5	0.0	0.5
Disability	0.4	0.8	0.4
Skin colour	0.3	1.5	0.4
Sexual orientation	0.4	0.0	0.3
Ethnicity	0.1	2.5	0.2
Other	0.6	0.1	0.6
Don't know	0.5	0.0	0.5
n (unweighted)	4767	287	5054

Source: SMIPS (2016). Light blue shading: significant difference ($p<0.05$) between respondents with immigrant and non-immigrant backgrounds.
NOTE: Weighted according to gender, age, and education.

speech' is broader than the legal definition (cf. Gagliardone et al., 2015; Sunde, 2013 p. 42; Waldron, 2012 p. 34). Thus, to fully capture how people experience hate speech (at least in the Norwegian context), it is necessary to employ a rather broad definition of the phenomenon.

With regard to respondents' immigrant background, the table shows that immigrants (10.7 percent) more often than

non-immigrants (7.0 percent) report having experienced what they perceive as hate speech. Furthermore, the grounds the received hate speech is directed towards differs. More immigrants than non-immigrants report hate speech directed towards gender, nationality, religion, skin colour, and ethnicity.

In order to distinguish between different forms of hate speech and other unpleasant messages, we categorized the experiences according to three different definitions of the phenomenon.[1] *Protected grounds* include experiences with hate speech directed towards grounds that are covered by Section 185 of the Norwegian Penal Code, i.e. ethnicity, nationality, skin colour, religion, disability, and sexual orientation. The second group (*expanded definition*) includes the same attributes as the first definition, but adds those who had experienced what they perceived as hate speech directed towards gender, personality, and appearance, which are all characteristics related to people's identities. Finally, in a third 'rest category' (*Other*) we grouped messages directed exclusively towards the content of the argument, political standpoint, occupation, education, other, and 'don't know'. This category thus includes comments that are further from the conventional understanding of hate speech. Table 2.3 sums up the share of non-immigrants and immigrants who reported having received what they perceived as hate speech, grouped by the three definitions.

The table shows that in total about 2 percent of the population have received what they perceive as hate speech directed towards protected grounds. When expanding the definition to include other characteristics related to personal identities, personality, gender and appearance (*Expanded definition*), the share

1 This differentiation between different types of unpleasant expressions is based on which *grounds* or content the expressions are directed towards. We do not have information about the tone or style of the messages.

Table 2.3. Has received what was perceived as hate speech via social media. Different definitions. Percent.

Definition	Non-immigrant background	Immigrant background	All
Protected grounds	1.6	7.0	1.9
Expanded definition	4.2	7.8	4.4
Other	2.8	2.9	2.8
Total	*7.0*	*10.7*	*7.2*
n	4767	287	5054

Source: SMIPS (2016). Light blue shading: significant difference (p<0.05) between respondents with immigrant and non-immigrant backgrounds.
NOTE: 'Protected grounds' include religion, ethnicity, skin colour, nationality, sexual orientation, and disability. 'Expanded definition' includes in addition personality, gender, and appearance. 'Other' includes the content of the argument, political standpoint, occupation, education, other, and don't know. Weighted according to gender, age, and education.

reporting having received hate speech increases to 4.4 percent of the full sample. Finally, 2.8 percent of the full sample reported having received what they perceived as hate speech, but only directed towards other characteristics, such as the content of one's argument and political standpoint.

We see that a relatively large share of the reported experiences with hate speech fall outside a conventional (legal or academic) understanding of hate speech, as they do not refer to speech directed towards group attributes in any sense. Thus, a substantial share of what the respondents report as hate speech should rather be understood as more general unpleasant experiences with online harassment. This underlines the ambiguity of the concept of hate speech in the general public, and illustrates how subjective perceptions of hate speech are broader than the legal definition.

We hypothesised (H1) that people of immigrant background would be more exposed to hate speech than other individuals.

With one exception, the results in Table 2.3 give initial support to this hypothesis. As would be expected, more immigrants (7 percent) than non-immigrants (1.6 percent) have received what they perceive as hate speech directed towards protected grounds. The difference between these two groups is reduced when expanding the definition to also include, gender, appearance and personality (7.8 vs 4.2 percent), but the difference is still statistically significant. However, considering unpleasant messages directed towards other attributes, the data suggests no difference between immigrants (2.9 percent) and non-immigrants (2.8 percent).

Descriptive statistics thus gave initial support to H1. The question is whether or not this relationship holds when controlling for other relevant factors. In order to test this we estimated two logistic regression models for each definition. Model (1) controls for gender, age, immigrant background, education, political ideology and political ideology squared. Model (2) adds propensity to share personal opinions on the internet. The dependent variable is 'has received [what respondents perceive as] hate speech' (1='yes', 0='no'). Odds ratios from the regression models are summarized in Table 2.4. Ratios above '1' indicate a positive relationship, while ratios below '1' indicate a negative relationship.

Controlling for a host of other factors, we see that the hypothesized relationship between experiences with hate speech and immigrant background is less clear-cut. Narrowing the definition of hate speech to *protected grounds*, there is a clear tendency that respondents with immigrant backgrounds are more exposed to hate speech compared to the majority population. Even controlling for propensity to share personal opinions on the internet (model 2), respondents with an immigrant background are almost four times (odds ratio=3.8) as likely as non-immigrants to have received hate speech directed

Table 2.4. Has experienced [what respondents perceive as] hate speech via social media. Logistic regression. Odds.

	Protected grounds		Expanded definition		Other	
	(1)	(2)	(1)	(2)	(1)	(2)
Female (ref=male)	0.47**	0.52*	0.80	0.89	0.55**	0.61*
Age	0.95***	0.95***	0.96***	0.96***	0.99	0.99
Immigrant background	4.44***	3.80***	1.78†	1.45	0.96	0.79
Vocational school (ref=high school)	0.51	0.51	0.52*	0.51*	1.10	1.08
Higher education (ref=high school)	0.79	0.84	0.81	0.84	0.74	0.72
Left-right scale	0.56*	0.83	0.60**	0.85	0.84	1.15
Left-right scale (squared)	1.05*	1.02	1.04**	1.02	1.01	0.98
Share opinions on the Internet		2.99***		2.91***		2.58***
Constant	0.72	0.02	1.05	0.03	0.14	0.005
Pseudo r^2	.125	.205	.072	.165	.019	.093
n	5054	5054	5054	5054	5054	5054

*Source: SMIPS (2016). Sig: †≤0.1 *≤0.05 **≤0.01 ***≤0.001.*
NOTE: 'Protected grounds' include religion, ethnicity, skin colour, nationality, sexual orientation, and disability. 'Expanded definition' includes in addition personality, gender, and appearance. 'Other' includes the content of the argument, political standpoint, occupation, education, other, and don't know. Weighted according to gender, age, and education.

towards protected grounds. This is not surprising considering that several of these grounds – nationality, ethnicity, skin colour – are more relevant to the immigrant population than to the majority population.

If we expand the definition of hate speech to also include personality, gender and appearance, the difference between respondents with immigrant backgrounds and non-immigrant background is reduced. The odds coefficient is still positive,

indicating that immigrants are also somewhat more exposed to hate speech according to the expanded definition. But when we introduce propensity to share personal opinions on the internet (model 2), the difference is no longer statistically significant. This might be due to the fact that few respondents have experienced what they perceive as hate speech. Nevertheless, based on this survey we must conclude that the majority population and respondents with immigrant backgrounds are equally exposed to this expanded definition of hate speech.

Finally, considering unpleasant messages directed towards other attributes, the data suggests that immigrants are *less* exposed to such messages (odds coefficient is below 1). Again, however, the difference is not statistically significant.

In other words, if we only consider hate speech directed towards legally protected grounds the first hypothesis (H1) is clearly supported. However, if we instead employ wider definitions of hate speech, that are perhaps closer to the popular understanding of the concept, H1 is not supported.

The results in Table 2.4 also reveal some other noteworthy findings. Men are more likely than women to have experienced what they perceive as hate speech directed towards protected grounds, and also to have experienced unpleasant messages directed towards other attributes. However, the gender difference is insignificant when messages targeted at gender, appearance and personality are included (expanded definition). The reason is obvious: Women are more likely than men to have received what they perceive as hate speech directed towards gender as an attribute (see Table 2.2). Furthermore, young people are more likely than older people to have received hate speech, but level of education is more or less irrelevant. People who place themselves on the far ends of the political left–right scale are more exposed to hate speech compared to political moderates. This relationship does however disappear when we

introduce propensity to share personal opinions on the internet (model 2). A probable explanation is that radicals, on either side of the political spectrum, are, on average, more politically active than moderates, leading them to engage in more heated discussions in social media. In general, propensity to share personal opinions is a very strong predictor for the likelihood of receiving hate speech. The odds of receiving hate speech directed towards protected grounds increases by about 3 for each increase on the four-point scale of propensity. Thus, the more active you are in debates on the internet, the more exposed you become to hate speech.

To sum up this first empirical part, about 7 percent of the sample reported having received what they perceived as hate speech through social media. These utterances were most often directed towards characteristics other than those covered by Section 185 of the Norwegian Penal Code, but rather directed towards the content of the argument, political standpoint and personality. About 2 percent have received hate speech directed towards legally protected grounds. When we expanded the definition to also include gender, appearance, and personality, 4-5 percent of the sample reported having received hate speech. Finally, we saw that people with immigrant backgrounds are much more exposed to hate speech directed towards legally protected grounds than non-immigrants, but non-immigrants are equally exposed to (what they perceive as) hate speech and unpleasant messages directed towards other attributes.

Discouragement from expressing opinions publicly

Next, we look at one possible consequence of receiving hate speech, namely discouragement from expressing opinions publicly. As in the previous section we distinguish between

three definitions of hate speech: protected grounds, expanded definition, and other. The following analyses are based only on those respondents who had experienced hate speech, and consequently the number of observations is limited and the results must be treated with caution.

Table 2.5 shows the answer distribution on the question of whether respondents who had received hate speech would be more cautious to express their opinions in public.

We hypothesized (H2) that hate speech directed towards legally protected grounds has more adverse consequences, in terms of discouragement from expressing opinions publicly, than other types of negative speech not directed towards minority group characteristics. This hypothesis is *not* supported by the results in Table 2.5. The results rather suggest that on the aggregated level the consequences are the same regardless of what kind of hate speech you measure: Across definitions more than one fourth of the respondents answered that they will indeed be more cautious when expressing their opinions in public. About two thirds said they would not be more cautious,

Table 2.5. Discouragement from expressing opinions publicly after experiencing hateful messages via social media. Percent.

	Protected grounds	Expanded definition	Other
Yes	27.2	26.4	30.0
No	66.3	67.1	59.0
Don't know	6.5	6.5	11.0
n (unweighted)	73	179	127

Source: SMIPS (2016).
NOTE: 'Protected grounds' include religion, ethnicity, skin colour, nationality, sexual orientation, and disability. 'Expanded definition' includes in addition personality, gender, and appearance. 'Other' includes the content of the argument, political standpoint, occupation, education, other, and don't know. Weighted according to gender, age, and education.

whereas the rest answered that they do not know. Differences across definitions of hate speech are not statistically significant. As such, these findings suggest that any experience with what one perceives as hate speech may lead to a retreat from public debates, and that hate speech directed towards protected grounds may not have more negative democratic consequences than other similarly unpleasant messages.

Aggregations may however hide important group variations. One can argue that hate speech and other unpleasant messages only have *democratic* consequences if particular groups are more likely than others to be silenced. We therefore end the empirical investigation by exploring variations in willingness to express opinions among different groups of respondents. As in the previous section, Table 2.6 summarizes results from two logistic regression models for each definition. Model (1) controls for gender, age, immigrant background, education, and political ideology (and political ideology squared), while model (2) adds propensity to share personal opinions on the internet. The dependent variable is 'experience with hate speech will limit willingness to express opinions' (1='yes', 0='No/Don't know').

The results in Table 2.6 suggest that H2 may hold for some segments of the population, most notably women. Across definitions, the regression models clearly suggest that women are more likely than men to state that experiences with hate speech have lead them to be more cautious in expressing personal opinions. However, the magnitude of the gender difference varies: women who have received hate speech directed towards protected grounds are about 5 times more likely than men to state that they will be more cautious. One could have expected that women experiencing hate speech directed towards gender, appearance, or personality, *in addition* to the protected grounds, would be even more affected. This is however not the case. By expanding

Table 2.6. Discouragement from expressing opinions publicly after experiencing hateful messages via social media. Logistic regressions. Odds ratio.

	Protected grounds		Expanded definition		Other	
	(1)	(2)	(1)	(2)	(1)	(2)
Female	5.05*	4.92*	3.50**	3.56**	2.50†	2.46†
Age	1.04†	1.07†	1.02†	1.03**	0.98	0.98
Immigrant background	3.56	2.99	2.42	2.41	1.75	2.04
Vocational school (ref=high school)	0.53	0.38	1.19	1.26	0.41	0.48
Higher education (ref=high school)	1.83	2.06	1.03	1.06	1.07	1.05
Left-right scale	0.38†	0.34*	0.56†	0.47*	1.27	1.09
Left-right scale (squared)	1.06	1.07†	1.04	1.05*	0.97	0.98
Share opinions on the Internet		0.47		0.54***		0.66
Constant	0.65	4.44	0.44	2.74	0.65	2.83
Pseudo r^2	.273	.308	.105	.141	.075	.094
n	73	73	179	179	127	127

*Source: SMIPS (2016). Sig: †≤0.1 *≤0.05 **≤0.01 ***≤0.001.*
NOTE: 'Protected grounds' include religion, ethnicity, skin colour, nationality, sexual orientation, and disability. 'Expanded definition' includes in addition personality, gender, and appearance. 'Other' includes the content of the argument, political standpoint, occupation, education, other, and don't know. Weighted according to gender, age, and education.

the definition to also include these three attributes, the odds coefficient is reduced. The gender difference is however still sizeable: women are about 3 times more likely than men to state that they will be more cautious after having received hate speech according to the expanded definition. Finally, the gender difference is reduced even more when regressing hate speech directed towards other characteristics that are further from

characteristics related to personal identities, and remains significant only at the 0.1 level. In other words, relative to men, the consequences for women seem to be strongest when receiving hate speech directed towards legally protected grounds (religion, ethnicity, skin colour, nationality, sexual orientation, and disability). Thus, for women, it does seem to matter what type of hate speech they receive.

Respondents' immigrant backgrounds are not statistically significant related to reluctance to express personal opinions publicly. Considering hate speech directed towards protected grounds, the size of the odds coefficients are substantial (odds=3), suggesting that immigrants are more affected by these utterances. However, due to few respondents the differences are not statistically significant.[2] Combining these findings with other recent studies (Midtbøen & Steen-Johnsen, 2016; Nadim, Fladmoe, & Wessel-Aas, 2016), we do however see clear indications that people of immigrant background in Norway are more affected by hate speech directed towards protected grounds than the majority population.

Summing up this final empirical section, we have seen that on the aggregate level the consequence of hate speech in terms of discouragement from expressing opinions publicly seems to be similar irrespective of what grounds the hate speech is directed towards. However, there are indications that women and people of immigrant background are more likely than men and the majority population to be affected by hate speech directed towards protected grounds. It is however important to treat the results in this final section with caution, as the number of

2 Of the 73 respondents who had been exposed to hate speech directed towards protected grounds, 14 had immigrant backgrounds. 7 (50 percent) of these said they would be more cautious. 18 (31 percent) of the 41 majority respondents gave a similar answer.

respondents is limited. More research is still needed in order to understand the consequences of experiencing different forms of hate speech and unpleasant comments.

Discussion and conclusion

The empirical analyses in this chapter were motivated by two research questions: (1) Which groups are most exposed to hate speech?, and (2) Are people who have experienced hate speech directed towards legally protected grounds more reluctant to express opinions publicly, compared to people who have experienced other types of negative comments? Survey data from Norway suggested that people with immigrant backgrounds are more exposed to hate speech directed towards grounds that are protected by the Norwegian General Civil Penal Code (skin colour or national or ethnic origin, religion or life stance, homosexual orientation, and disability), but that non-immigrants are equally exposed to hateful messages directed at other grounds, such as gender, appearance, political viewpoints, etc. It is especially people who often share personal opinions on the internet who are vulnerable to hateful and other unpleasant messages.

Several scholars have argued that hate speech directed at protected grounds have more severe consequences compared to other forms of negative speech (e.g. Boeckmann & Liew, 2002; Herek et al., 2002). One explanation for this is that since hate speech triggers the awareness of belonging to a vulnerable group it incites more fear than other types of negative speech (Herek et al., 2002). We tested this claim on one possible consequence, namely discouragement from expressing opinions in the public. We found that a substantial share of those who had received hate speech were indeed discouraged from expressing opinions. However, contrary to what we expected, on the aggregated level

the analyses showed that reluctance to express opinions publicly is more or less on the same numeric level irrespective of what grounds the hate speech is directed towards. This suggests that negative or derogatory speech may function as a social boundary for free speech irrespective of content – as long as people subjectively perceive messages as hateful. Disaggregating the general public, we did however see that women and people of immigrant background seem to be more strongly affected by hate speech directed towards protected grounds, than by other types of negative comments (see also Midtbøen & Steen-Johnsen, 2016; Nadim et al., 2016). A possible interpretation of this finding is that women and immigrants, more than men and the majority population, see themselves as belonging to vulnerable groups, and that they therefore react more negatively to messages directed towards group-based identity characteristics. This suggests that hate speech, more than other types of negative and derogatory speech, can represent a democratic problem in that it might silence *specific* groups and discourage them from voicing their opinions publicly. These findings are based on a relatively small number of respondents, and a task for future research should be to examine more carefully whether hate speech is distinct from other types of speech for minority groups.

Is hate speech in social media an extensive phenomenon in Norway? In the survey analyzed in this chapter, 7 percent said they had experienced what they perceived as hate speech, and 2 percent had experienced hate speech directed towards protected grounds. These are small numbers, and, as such, one may view hate speech as a marginal phenomenon. Such an interpretation is, however, problematic. First, the legal definition of hate speech aims at protecting vulnerable *minorities*. Minorities obviously make up a limited share of the total population, and by

analyzing a national representative sample the number of respondents with any type of minority background will be limited. Indeed, if we only look at respondents with immigrant backgrounds about 7 percent reported that they had experienced hate speech directed towards protected grounds, and the regression models also suggested that – all else equal – this group was almost four times as likely as non-immigrants to have experienced such speech. A second objection is that hate speech can have consequences not only for those who receive messages directly, but also for those who observe the messages (Bell, 1998; Kunst et al., 2013; Perry, 2014). A comparative study of young adults in the US, UK, Germany, and Finland, found that between 30 (Germany) and 50 percent (USA) had during the *past three months* witnessed 'writings or speech online, which inappropriately attacked certain groups of people or individuals' (Hawdon et al., 2015). Thus, although few people have direct experience with receiving hate speech, it appears to be relatively common among young adults to have witnessed it. Observing hate speech can also incite fear among individuals who are not directly targeted, because it highlights the risk of being subjected to it (Gelber & McNamara, 2016 p. 327; Perry, 2001).

Hate speech brings the question of boundaries of freedom of speech to the fore. This chapter has illustrated how hate speech and other unpleasant messages can represent social boundaries to the exercise of free speech. Empirical evidence suggests that a substantial number of individuals who receive hateful messages, become reluctant to express opinions publicly. One purpose of hate speech is to incite fear in the groups targeted, and fear can be an effective silencing mechanism. Hate speech as a response to an individual's public expression of opinions, is an attack on the legitimacy of that person's position as an equal member in public debate (see also Enjolras, ch. 10). If certain groups are

systematically silenced, hate speech ultimately has democratic consequences. Legal regulation of hate speech does, however, also represent (potential) boundaries on freedom of speech. Where to draw the line between freedom of speech and protection against hate speech is a delicate balance, and it is ultimately a political and normative question.

References

Bell, D. (1998). Wealth transfers occasioned by marriage: a comparative reconsideration. In T. Schweizer & D. R. White (Eds.), *Kinship, networks, and exchange*. Cambridge: Cambridge University Press.

Bleich, E. (2011). *The freedom to be racist? How the United States and Europe struggle to preserve freedom and combat racism*. Oxford: Oxford University Press.

Boeckmann, R. J., & Liew, J. (2002). Hate speech: Asian American students' justice judgments and psychological responses. *Journal of Social Issues, 58*(2), 363–381.

Boeckmann, R. J., & Turpin-Petrosino, C. (2002). Understanding the Harm of Hate Crime. *Journal of Social Issues, 58*(2), 207–225.

Chakraborti, N., & Garland, J. (2015). *Hate crime: impact, causes and responses* (2nd ed. ed.). Los Angeles: Sage.

Egge-Hoveid, K., & Sandnes, T. (2015). Innvandrere og norskfødte med innvandrerforeldre i et kjønns- og likestillingsperspektiv: Utdanning, arbeid og samfunnsdeltakelse *Reports*. Oslo/Kongsvinger.

Eggebø, H., Sloan, L., & Aarbakke, M. H. (2016). Erfaringer med digitale krenkelser i Norge: KUN Senter for kunnskap og likestilling.

Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). Countering Online Hate Speech. Paris: UNESCO.

Gelber, K., & McNamara, L. (2015). The effects of civil hate speech laws: Lessons from Australia. *Law & Society Review, 49*(3), 631–664.

Gelber, K., & McNamara, L. (2016). Evidencing the harms of hate speech. *Social Identities, 22*(3), 324–341.

Hatebase. (2016). Most common hate speech. Retrieved 15.08.2016, from https://www.hatebase.org/popular

Hawdon, J., Oksanen, A., & Räsänen, P. (2015). Online Extremism and Online Hate: Exposure among Adolescents and Young Adults in Four Nations. *Nordicom Information: Medie- och kommunikationsforskning i Norden, 37*(3-4), 29–37.

Herek, G. M., Cogan, J. C., & Gillis, J. R. (2002). Victim Experiences in Hate Crimes Based on Sexual Orientation. *Journal of Social Issues, 58*(2), 319–339. doi: 10.1111/1540-4560.00263

Jenness, V. (2003). Engendering Hate Crime Policy: Gender, the 'Dilemma of Difference', and the Creation of Legal Subjects. *Journal of Hate Studies, 2*(1), 73–92.

Kunst, J. R., Sam, D. L., & Ulleberg, P. (2013). Perceived islamophobia: Scale development and validation. *International Journal of Intercultural Relations, 37*(2), 225–237.

Lawrence III, C. R., Matsuda, M. J., Delgado, R., & Crenshaw, K. W. (1993). Introduction. In M. J. Matsuda, C. R. Lawrence III, R. Delgado & K. W. Crenshaw (Eds.), *Words that wound : critical race theory, assaultive speech, and the First Amendment*. Boulder, Colorado: Westview Press.

Leets, L. (2002). Experiencing hate speech: Perceptions and responses to anti-semitism and antigay speech. *Journal of Social Issues, 58*(2), 341–361.

Maher, J., McCulloch, J., & Mason, G. (2015). Punishing Gendered Violence as Hate Crime: Aggravated Sentences as a Means of Recognising Hate as Motivation for Violent Crimes against Women. *Australian Feminist Law Journal, 41*(1), 177–193.

McPhail, B. A. (2002). Gender-bias hate crimes: A review. In B. Perry (Ed.), *Hate and bias crime. A reader* (pp. 261–280). New York: Routhledge.

Midtbøen, A. H., & Steen-Johnsen, K. (2016). Ytringsfrihetens grenser i det flerkulturelle Norge. *Nytt norsk tidsskrift, 33*(1-2), 21–33. doi: 10.18261

Nilsen, A. B. (2014). *Hatprat*. Oslo: Cappelen Damm akademisk.

NOU. (1999). *'Ytringsfrihed bør finde Sted' - Forslag til ny Grunnlov § 100*. (NOU 1999: 27). Oslo: Justis- og politidepartementet Retrieved from https://www.regjeringen.no/no/dokumenter/nou-1999-27/id142119/.

Parekh, B. (2006). Hate speech. *Public policy research, 12*(4), 213–223.

Perry, B. (2001). *In the name of hate: understanding hate crimes*. New York: Routledge.

Perry, B. (2014). Exploring the community impacts of hate crime. In N. Hall, A. Corb, P. Giannasi & J. Grieve (Eds.), *The Routledge international handbook on hate crime*. New York: Routledge.

Pew Research Center. (2014). Online Harassment.

Phillips, A. (2009). *Multiculturalism without Culture*. Princeton: Princeton University Press.

Sunde, I. M. (2013). Forebygging av radikalisering og voldelig ekstremisme på internett. Oslo: Politihøyskolen.

Waldron, J. (2012). *The harm in hate speech*. Cambridge, Mass: Harvard University Press.

Wessel-Aas, J., Fladmoe, A., & Nadim, M. (2016). Hate speech Report 3: The boundary between freedom of speech and criminal law protection against hate speech. Oslo: Institutt for samfunnsforskning.