

# The Anti-Hate Brigade

How a Group of Thousands  
Responds Collectively  
to Online Vitriol

By Cathy Buerger



**DANGEROUS SPEECH PROJECT**

# Introduction

*“Do you think about what you ought to say, or what you wish others would say when you read the comment fields in social media? Would you dare to share your opinion more frequently if you knew that others in the same thread supported you? Are you already taking action to disrupt those who are spreading hate on the Internet, and do you wish you had the support of others who are doing the same?”<sup>1</sup>*

These are the words that greet visitors to the Facebook group #jagärhär, a Sweden-based collective of thousands of people who have made a regular practice of responding en masse to what they regard as hateful comments online. Members of #jagärhär (which means “I am here”) seek out hatred in comment threads of newspaper articles posted on Facebook and then respond together, following a strict set of rules which includes keeping a respectful and non-condescending tone and never spreading prejudice or rumors.

Lena<sup>2</sup> is one of the group’s members. Before joining #jagärhär in 2017, Lena had long been active in online forums, even helping to moderate one for tennis players. But online discourse seemed to her to be deteriorating. She saw comments asserting that immigrants posed a serious threat to Swedish identity and culture, and even some suggesting that more immigration could lead to a ‘white genocide.’ Troubled by what she saw, Lena at first thought that it was just a few extreme voices. When it seemed that such discourse was taking over many of the comment threads she was reading on Facebook, she wondered if more people agreed with these ideas than she imagined.

Then one day she noticed the hashtag #jagärhär in a comment thread. She searched for it on Facebook, found the group, and after reading the page’s greeting and learning more about the group, decided to participate. “It was a big comfort to find them because it is really horrible out there in the comment threads,” she said. “It’s almost that I was thinking, ‘is this the new way?’ But then I found them (#jagärhär), and I thought, ok, I can feel safe. It’s not the new way. I was thinking in the right direction,” she said.

This is the by far the largest and best-organized collective effort to respond directly to hatred online,<sup>3</sup> anywhere in the world, as far as we know. It is also one of only two civil society efforts against hatred online to have been replicated in numerous other countries.

A detailed account of #jagärhär’s efforts, this is the first qualitative study of such a group. In 27 interviews with group members and administrators, I gathered extensive information on how members respond to hateful content, and how doing so affects them. I also focused (as much as possible) on how they seem to be influencing online discourse. Drawing on these interviews as well as observation of the group’s everyday practices, the study asks and attempts to answer these questions:

- What made people want to join such a group?
- Whom are they trying to influence, and what strategies do they use?
- What impact does the group itself have on members’ capacity and willingness to respond to hateful and often hostile comments online?
- What impact does the group have on discourse norms and on non-members?

1. Original Swedish text: “Tänker du vad du borde säga eller vad du önskar att andra sa när du läser kommentarsfälten på nätet? Skulle du våga dela din åsikt oftare om du visste att det fanns andra i samma tråd som stöttade dig? Går du redan in och står upp för andra som blir utsatta och önskar att du hade stöd från fler som sa ifrån?” The text comes from #jagärhär’s Facebook group.

2. To protect the privacy of research participants, all of the names used in this paper are pseudonyms, with the exception of Mina Dennert, who is a public figure.

3. There have been a few other organized collective responses to online hatred including Virtar í Athugasemdum, an Icelandic group that formed in 2015 to counter misogynistic speech (it is no longer active) and Reconquista Internet, a German counterspeech group that was formed in April of 2018. Another notable example is Sleeping Giants, a network of groups active in over a dozen countries that pressures companies to remove their advertising from news sites like Breitbart that spread hateful content. Sleeping Giants and the #iamhere network are the only examples of multinational collective responses to online hatred that we have found.

4. By “undermine” we mean responses that seek to diminish that kind of speech by that person and/or others. For more information, see: The Dangerous Speech Project. “Counterspeech.” <https://dangerousspeech.org/counterspeech/> [<https://perma.cc/9P4A-2CT3>]

Counterspeech, which my colleagues and I define as “any direct response to hateful or harmful speech which seeks to undermine it,”<sup>4</sup> has been touted by internet platforms<sup>5</sup> and civil society<sup>6</sup> as a possible answer to hatred and extremism, but there is only limited empirical evidence of its success<sup>7</sup> and even less research on the individuals who produce counterspeech and what they have learned from practicing their little-known craft. This paper begins to fill that void.

Most studies examining whether counterspeech is an effective remedy to hatred have investigated whether such speech changes the mind or behavior of the person to whom it responds. Not surprisingly, the answer is usually ‘no.’ Miškolci et al. (2018), for example, found that responding directly was not an effective way to change the behavior of a person who had been posting hateful content.<sup>8</sup> Others have found that counterspeech can occasionally work to change someone’s online behavior, but its effectiveness is strongly dependent on specific factors, such as the proportion of counterspeakers to hateful speakers,<sup>9</sup> whether they are counterspeaking as part of a group,<sup>10</sup> the intensity of the beliefs held by those posting hatred,<sup>11</sup> the tone used by a counterspeaker,<sup>12</sup> or even specific characteristics of the people doing the counterspeaking – such as their race or perceived popularity.<sup>13</sup> Because of these factors, success is still rare.

Changing someone’s mind or behavior can be very difficult, depending on the circumstances. There is a wealth of scholarship describing how humans attempt to reduce their cognitive dissonance by bringing their beliefs in line with their actions.<sup>14</sup> People may change their behavior to reflect new information they have received, attempt to justify their behavior within the context of the new information, or deny that the new information exists or is true.<sup>15</sup>

And as psychologist Clifford Lazarus has written, “The bottom line is that when there is a conflict between our attitudes and our behavior, we tend to change our attitudes to make them consistent with our behavior rather than change our behavior to make it consistent with our attitudes.”<sup>16</sup>

Changing the mind or behavior of someone who has posted hateful speech is not the only way counterspeech can be effective, however.<sup>17</sup> Counterspeakers may also try to influence the discourse norms of the audience – the much larger number of people who are witness to exchanges of hateful speech and counterspeech. This is more likely to succeed,<sup>18</sup> and in our larger body of work on counterspeech at the Dangerous Speech Project,<sup>19</sup> we have found that in fact it is usually counterspeakers’ primary goal. Studies on how online speech affects an audience have tracked discourse norm shifts in online spaces, finding that users can influence both so-called “pro-social”<sup>20</sup> and anti-social norms.<sup>21</sup> These findings have important ramifications for counterspeakers, as they demonstrate that the style and tone of responses can influence the behavior of others who encounter the conversation, not just the people with whom they attempt to converse.

As the first qualitative study of online counterspeakers, and perhaps the first study to focus on them, this paper contributes new insight into who some of them are, what they are trying to accomplish, what motivates them (both to counterspeak in general, and to respond to specific posts or comments), and what they have learned from their practices. What conditions are necessary for a person to speak out against hatred when they see it? And how does counterspeaking as a group change the practice? Answers to questions like this offer valuable explanatory context for other counterspeech studies based primarily on quantitative data, which may document an effect but may not explain why it has occurred.

5. Yadron, Danny. 2016. “Facebook’s Sheryl Sandberg: ‘likes’ can help stop Isis recruiters.” *The Guardian*. <https://www.theguardian.com/technology/2016/jan/20/facebook-davos-isis-sheryl-sandberg> [https://perma.cc/X5ZW-5FX4]
6. Anti-Defamation League. 2016. “‘Best Practices’ and Counterspeech Are Key to Combating Online Harassment.” *ADL Blog*. <https://www.adl.org/blog/best-practices-and-counterspeech-are-key-to-combating-online-harassment> [https://perma.cc/GG8Z-3BJM]
7. For a survey of this literature, see Buerger, Catherine and Lucas Wright. 2019. “Counterspeech: A Literature Review.” *The Dangerous Speech Project*. [https://dangerousspeech.org/wp-content/uploads/2019/11/Counterspeech-lit-review\\_complete-11.20.19-2.pdf](https://dangerousspeech.org/wp-content/uploads/2019/11/Counterspeech-lit-review_complete-11.20.19-2.pdf) [https://perma.cc/2LTU-A7RR]
8. Miškolci, Jozef, Lucia Kováčová, and Edita Rigová. 2018. “Countering hate speech on Facebook: The case of the Roma minority in Slovakia.” *Social Science Computer Review*.
9. Schieb, Carla, and Mike Preuss. 2016. “Governing hate speech by means of counterspeech on Facebook.” In 66th ICA annual conference, at Fukuoka, Japan, pp. 1-23.
10. Garland, Joshua, Keyan Ghazi-Zahedi, Jean-Gabriel Young, Laurent Hébert-Dufresne, and Mirta Galesic. 2020. “Impact and dynamics of hate and counter speech online.” *arXiv preprint arXiv:2009.08392*; Friess, Dennis, Marc Ziegele, and Dominique Heinbach. 2020. “Collective Civic Moderation for Deliberation? Exploring the Links between Citizens’ Organized Engagement in Comment Sections and the Deliberative Quality of Online Discussions.” *Political Communication* p. 1-23.11Ibid.Schieb and Preuss, supra note 9.
11. Schieb and Preuss, supra note 9.
12. Bartlett, Jamie, and Alex Krasodomski-Jones. 2015. “Counter speech. Examining content that challenges extremism online.” *Demos; Frenett, Ross, and Moli Dow*. 2015. “One to one online interventions: A Pilot CVE methodology.” *Institute for Strategic Dialogue*.
13. Munger, Kevin. 2017. “Tweentment effects on the tweeted: Experimentally reducing racist harassment.” *Political Behavior* 39, no. 3: 629-649; Seering, Joseph, Robert Kraut, and Laura Dabbish. 2017. “Shaping pro and anti-social behavior on twitch through moderation and example-setting.” In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pp. 111-125. 2017.
14. McGrath, April. 2017. “Dealing with dissonance: A review of cognitive dissonance reduction.” *Social and Personality Psychology Compass* 11, no. 12: e12362.; Elizabeth Kolbert. February 20, 2017. “Why Facts Don’t Change Our Minds. *The New Yorker*. <https://www.newyorker.com/magazine/2017/02/27/why-facts-dont-change-our-minds> [https://perma.cc/T5MC-A6RG]
15. Lazarus, Clifford. 2018. “Why Many People Stubbornly Refuse to Change Their Minds.” *Psychology Today* <https://www.psychologytoday.com/us/blog/think-well/201812/why-many-people-stubbornly-refuse-change-their-minds> [https://perma.cc/G2JL-G8VR]
16. Ibid
17. Benesch et al. note that ‘success’ in counterspeech interactions can be defined in two ways. An interaction can be considered successful if it had some favorable impact on the speech of original hateful speaker - or a favorable impact on the discourse norms of those reading a counterspeech conversation. Those ‘spectators’ are far more numerous than the hateful speakers. Benesch, Susan, Derek Ruths, Kelly P. Dillon, Haji Mohammad Saleem, and Lucas Wright. 2016. “Considerations for successful counterspeech. *Dangerous Speech Project*.” p. 2.
18. Garland et al. (2020) found that organized counterspeech seems to make a difference on overall discourse. See Garland et al., supra note 10.
19. *The Dangerous Speech Project*, supra note 4.
20. Han, Soo-Hye, and LeAnn M. Brazeal. 2015. “Playing Nice: Modeling Civility in Online Political Discussions.” *Communication Research Reports*, 32(1): 20-28
21. Cheng, Justin, Michael Bernstein, Christian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. “Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions.” *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW ’17*, 1217-1230; Kwon, K. Hazel, and Anatoly Gruzd. 2017. “Is offensive commenting contagious online? Examining public vs interpersonal swearing in response to Donald Trump’s YouTube campaign videos.” *Internet Research*. 27(4): 991-1010

# #Jagärhär: Collective Counterspeech

In 2016, a Swedish woman named Mina Dennert had an idea. Dennert had noticed an alarming surge in hateful and xenophobic content online after more than a million Syrians, Afghanis, and Iraqis sought refuge in Europe, mainly in 2015. She wondered: could a group of people working together to respond to hatred online shift the discourse in online spaces, and perhaps therefore in other contexts, toward more civil and fact-based speech?

Dennert emigrated from Iran as a young child, growing up in a small town in southern Sweden. Although she had experienced discrimination and harassment throughout her life, what she saw in 2015 and 2016 felt different. She had become used to seeing racist and anti-immigrant messages posted by people she describes as 'the usual subjects' – extreme right-wing social media users who commented frequently on social issues. But in the spring of 2016, she started seeing the same hateful narratives being repeated by people she knew – people she thought of as 'good people.' She urgently wanted to find a way to counter the spread of such ideas.

Dennert began responding to people on Facebook who were posting what she saw as hatred and misinformation. The work was difficult, and too much for her alone. So she recruited 20 friends to help, and they set up a Facebook group to organize their activity. Dennert called the group '#jagärhär,' Swedish for 'I am here.' She says the name has two different meanings, one practical and one normative. "It has the meaning of 'this is where I am, and I need help here in this comment field,' since we call for each other to help out. It also has the meaning: 'I am present. I can see what you are doing here, and I don't agree. I am here too.'"<sup>22</sup>

As a guide for deciding which speech to counter, the group uses the Swedish legal concept of "agitation against a population group," defined as "a statement or other communication that is disseminated" that "threatens or expresses contempt for a population group by allusion to race, colour, national or ethnic origin, religious belief, sexual orientation or transgender identity or expression[...]."<sup>23</sup> Although this type of speech is outlawed in Sweden, several #jagärhär members told me that the law is rarely enforced. When members find speech that they believe meets this definition in comments below posts on the Facebook pages of newspapers or public groups, they share links to the posts in their group. Members respond to comments, then 'like' each other's comments, pushing them to the top of the comments thread, since Facebook ranks comments on public pages based on interactions ('likes' and replies).<sup>24</sup> This is a vital feature of #jagärhär's model: they make use of Facebook's system to amplify their own civil, fact-based comments and bury hateful or xenophobic comments at the bottom of comment threads, making it less likely that others will see them. Indeed the act of liking other comments is as important to #jagärhär's model of counterspeech as the writing of individual comments. To make it easier to find each other's comments, they also tag them with the hashtag #jagärhär.

In one of their first actions as a group, Dennert and her friends responded to comments posted below an editorial in Aftonbladet, one of the largest newspapers in Scandinavia, accompanied by a widely-published<sup>25</sup> image of a young Syrian child sitting alone and shell-shocked in the back of an ambulance in Aleppo, his body covered with dust from an explosion, and half of his face coated with blood. The headline read "He Opens Our Eyes – To the War."<sup>26</sup> It was August of 2016, and Dennert's group was drawn to this particular article because of comments questioning the veracity of the photo and suggesting that it had been manipulated in order to convince people to support asylum seekers in Sweden who had fled the crisis in Syria."

22. Interview with author, March 12, 2018.

23. Swedish Criminal Code, Chapter 16, Section 8, p. 133. <https://www.government.se/49f780/contentassets/7a2dcae0787e465e9a2431554b5eab03/the-swedish-criminal-code.pdf> [<https://perma.cc/3F45-GZUQ>]

24. Facebook Help Center. 2020. "How do I turn comment ranking on or off for my Facebook Page or profile?" <https://www.facebook.com/help/1494019237530934> [<https://perma.cc/P6SG-T9M6>]

25. The image accompanied stories on the The Guardian, Time, BBC News, and CNN, among many others.

26. <https://www.facebook.com/aftonbladetledare/posts/1185903078127911> [<https://perma.cc/ES22-TCT5>]

Everything is smoke and mirrors, lies and propaganda that serve the elite's agenda"<sup>27</sup> reads one comment. Another links to an article asserting that the CIA engineered the war in Syria. Yet another reads, "The photographer who took the picture hangs out with guys who behead children."<sup>28</sup>

Although these comments are still visible, a reader must put in extra effort – expanding the condensed comment thread and reading down to the bottom – to see these and other comments to which Dennert and others in her group responded. Comments written by members of #jagärhär, including the following comment written by Dennert herself, are the first comments that a reader would encounter:

*This is almost the roughest thing I've seen! I'm trying to understand so I will be able to fall asleep tonight! Would we feel better if these were fake pictures? Would it feel easier to support closed borders and xenophobia if there were no wars that people were forced to flee from? These are still images from filmed material. They're true. This is one child among millions of children at war, among millions of children without clean water, millions of children being exploited, abused. Who falls asleep frozen and hungry. You do not have to worry. Look away if you want. But do not humiliate a child who survived the worst bombings in modern times. Drawn from the ruins. Have more respect, if not for the boy then for yourselves. No human can sink so low. It's not worthy! #jagärhär*

Dennert's comment is responding to the arguments made by those questioning the legitimacy of the picture, but because it now appears at the top of the comments (rather than as a response), it loses some of its context. Members of #jagärhär have become aware of this challenge, and now largely try to craft comments that can be understood even when they are read without the content that inspired them.

Four years later, what began as Dennert's small group of friends has ballooned to about 74,000 members – a substantial number in a country of only 10 million. Of those, about 2,000,000 are active at least once a week.

They all work as volunteers, even though some of them devote more than 10 hours a week to the project.<sup>29</sup> The group has a team of 15-20 moderators who organize counterspeech actions and help manage the day-to-day action on the group's Facebook page (each day there are two moderators on duty), and six administrators (including Dennert) who handle the larger workings of the group. About 70% of members are women and the majority are between 35 and 40 years old.<sup>30</sup> Most members live in Sweden, but some do live in other countries. In the beginning, members were highly concentrated in urban areas; although there are now more members from other areas, urban dwellers continue to predominate.

The group is officially non-political, and members do not necessarily share the same views about many of the topics they include in their actions, such as immigration policy. Its moderators include members of seven of Sweden's eight major political parties.<sup>31</sup> This is something in which Dennert takes great pride, and the group's leaders have intentionally sought political diversity within the group of moderators. Some members volunteer to become moderators, but those on the admin team also sometimes directly reach out to invite members to moderate if they think the person might be a good moderator or if they seem to have an underrepresented perspective to bring to the moderator team. Group members do share some important values, however including that civil and productive political conversation online is possible and worth working for.<sup>32</sup>

In order to create spaces for this type of conversation to occur, when writing counterspeech comments, members follow a set of rules developed by Dennert around the time she founded the group. Dennert said that she decided to write them because "even anti-racists, even people who meant well, were making things worse. It could be very condescending. It was very much like, 'you're wrong, you're right.' And I just thought, this isn't getting us anywhere." Today, the full list of rules appear at the end of each action post.

27. Original Swedish: "Allting är rök och speglar, lögn och propaganda som tjänar elitens agenda."

28. Original Swedish: "Fotografen som tog bilden hänger med killar som halshugger barn."

29. All of the work done for the #jagärhär Facebook group has always been done voluntarily. There is a connected association also called #jagärhär that was started in 2017 that conducts lectures, hosts debates, and provides support to the many #iamhere groups in the international network. In 2017, #jagärhär received some funding when a rock band made a donation to the group. They have also won a few small financial prizes. They have not received any other substantial funding. In 2018, after receiving the donation, Dennert was employed full-time at the association, and a few others were hired on a contract basis. No one has been employed by the association in the past two years.

30. Interview with author, August 20, 2020.

31. The far-right Sweden Democrats are not represented. Interview with author, August 20, 2020.

32. Mina Dennert (2018). "How facing your fears may solve polarization." TEDxAthens. [https://www.ted.com/talks/mina\\_dennert\\_how\\_facing\\_your\\_fears\\_may\\_solve\\_polarisation/transcript?language=en](https://www.ted.com/talks/mina_dennert_how_facing_your_fears_may_solve_polarisation/transcript?language=en) [<https://perma.cc/2PT7-S7F6>]

The rules are as follows:<sup>33</sup>

- Like, react and write supportive responses to good comments to lift them up in the fields and poke down hateful comments. All efforts are important!
- Avoid reacting and writing many answers to hateful comments, as this lifts them higher in the fields. Rather like and react to already existing good answers.
- Write what you think and think yourself. But keep in mind that as members of #iamhere, we never spread hate, prejudice, slander, gossip or rumors. We also do not comment on other people's spelling or writing methods. We always stay factual.
- Keep a good tone! We never express ourselves using condescending, despicable, or scornful language or by insulting other people. Instead, with our choice of words, we show that we stand for transparency, respect and good conversation. This applies both in the comments fields we link to and in here in the group.
- Debate and discussion on the issues of facts should not take place in the group, but in the comments we link to. It is outside the group where we should work to make change and make a difference. It's out there that we're going to stop the hate and nuance the debate. #iamhere is an action group - not a debate group.
- Small talk about the action, such as encouragement, support, tips and advice, however, is OK here in the group. Please also message here if you have commented (K), liked (G) or reacted (R), and in which comments.
- If you want to link here directly to a comment you have written in any field, you can click or right click on the " timestamp " under your comment and copy the link address.

The group's moderators are responsible for ensuring that members follow the rules. If violations occur inside the Facebook group, moderators remove the comment, and write to the person explaining why they took it down. If it happens outside of #jagärhär's Facebook page, during an action, for example, moderators take a screen shot of the comment and send it to the person who posted it, with an explanation of how it violated the rules. Sometimes, when the infractions are serious enough (if a member posted a racist or xenophobic comment, for example) or when they continue over time, moderators remove the member from the group. Some members choose to leave the group soon after joining because they do not agree with the rules. According to Dennert, some people tell her, "what are you doing? You're not making any difference" or "I need to tell people when they are stupid."

Most members with whom I spoke reported being strongly in support of the rules, saying that they believed that following them was the most effective way to change discourse within comment sections. There were a few, however, who admitted that they did not always follow the rules. As one woman said, "The most important is to not get too emotional and attack. It's very easy to do that. I do that and then I erase that comment, and then when I've rephrased it enough, I post it. But no, you can't always follow [the list of rules]. You just get too emotional, too upset, too angry. But I don't ever want to be mean. The most important thing to me is that every person deserves respect. If I do post something nasty, I just don't use the hashtag," she said, laughing.<sup>34</sup> When she does tag her comments with '#jagärhär,' she is sure to follow the group's rules.

Counterspeech groups are very unusual, and #jagärhär's model seems to be unique in that it has been replicated in many other countries - 12 at this writing. All the groups, (in Sweden, Australia, Canada, Czech Republic, France, Finland, Germany, Italy, Norway, Poland, Slovakia, and the UK) are named "I am here" in the relevant languages. (There was briefly a group in the United States, but Dennert struggled to find a team capable of keeping the group going in such a large and complex media landscape.)

33. The following list was taken from a translated version of the rules that appear on the #jagärhär Facebook page.

34. Interview with author, October 25, 2019.

In total, 146,000 people have joined one of the Facebook groups.<sup>35</sup> The Swedish group is the largest, followed by the German group (#ichbinhier), which has about 45,000 members. The other groups vary widely in size, with the smallest (Poland) having just 162 members at the time of writing. Generally, a small proportion of members participate in the groups' activities: for example over one 2-week period, 5,580 out of the nearly 74,000 members of the Swedish group engaged with a post in the #jagärhär Facebook group.

All of the groups follow the same rules for writing comments, but some use slightly different methods. For example, the German group has developed a tool they call the 'Aktions-Bot' (actions bot) that quickly identifies each comment containing the group's hashtag, to help members 'like' them more efficiently. The Slovakian moderators occasionally break with the #iamhere policy of not permitting debate of issues inside the groups, inviting their members to discuss a predetermined topic in the comments of a post on their group's Facebook page. Past topics have included the virtues and pitfalls of socialism and what makes a medium 'credible'. They have even hosted offline debate events. The English speaking groups (#iamhereCanada, #iamhereUK, and #iamhereAustralia) often do joint actions where the members from all three groups work collectively on one article. Although all the groups occasionally take part in so-called 'global actions,' those three groups work together more frequently than others – a feature of both their shared language and smaller sizes. (Small groups need help since it's difficult to amplify their comments in very active comment threads on their own).

## Methodology

I used semi-structured ethnographic interviews to examine why group members got involved with #jagärhär, how they decide when to counterspeak and what to write, what challenges they face in their work, and what keeps them engaged. 25 group members participated in the study. To select them, I made a sampling frame by assembling a list of every member of #jagärhär who had participated (commented or 'liked' a post) on the group's Facebook page over a two-week period (N=5,580). I drew a random sample from that list and invited those individuals to be interviewed. For those who agreed to participate, I conducted verbal interviews in English, over Skype or Facebook Messenger. Interviews were conducted until saturation was reached (the point where no new themes emerged from new interviews). Interviews were done between July and December of 2019 and during July and August of 2020. Primary analysis took place between January and March 2020.

I also used ethnographic observation to better understand the 'culture' of the #iamhere network. I joined all of the #iamhere Facebook groups and regularly visited their pages, reading updates and observing the rhythms of the groups – for example, when moderators posted each day and how many times. Through this observation I was able to learn what kind of content was shared, and how members responded to it.

Digital ethnography is an incipient research technique,<sup>36</sup> although there is a growing body of literature on the methodology.<sup>37</sup> As with 'traditional' offline ethnographies, ethnographic observation helps researchers develop a tacit understanding of the culture that they observe, knowledge that then serves as the basis for a more grounded interpretation of data collected through other methods, such as interviews.<sup>38</sup> In this way, ethnographic observation contributes to both data collection and data analysis.

35. As of August 10, 2020; this does not account for people who have joined more than one group, but given the diversity of languages and each group's focus on its own region, it seems unlikely that a significant number of people are in multiple groups. The #iamhere groups only operate on Facebook, in part because their method relies on utilizing the platform's comment ranking feature.

36. Wilson, Richard Ashby, 2019. "The Digital Ethnography of Law: Studying Online Hate Speech Online and Offline." *Journal of Legal Anthropology* 3(1): 1-20, Available at SSRN: <https://ssrn.com/abstract=3465225>

37. See Boellstorff, Tom. 2012. "Rethinking digital anthropology." *Digital anthropology*, pp. 39-60; Pink, Sarah. 2016. "Digital ethnography." *Innovative methods in media and communication research*, pp. 161-165; Varis, Piia. 2016. "Digital ethnography." *The Routledge handbook of language and digital communication*, pp. 55-68.

38. Dewart, Kathleen M. and Billie R. Dewart. 2002. *Participant Observation: A Guide for Fieldworkers*. Walnut Creek, CA: AltaMira Press.

# Counterspeech in a Changing Sweden

When Dennert founded #jagärhär, she and many others who joined her in the early days felt that Sweden was changing. In 2015, Sweden welcomed over 160,000 asylum seekers – one of the highest per capita rates in Europe<sup>39</sup> – mostly from Syria, Afghanistan, and Iraq. At the time, Sweden had some of the world's most generous immigration policies, significantly exceeding European Union minimum standards for immigration and asylum.<sup>40</sup>

Public support for these policies decreased throughout 2015, however. By the end of the year, there was a growing concern that Sweden would not be able to sustain its social service offerings if the same number of asylum applicants entered in the following year.<sup>41</sup> In public discourse, Swedes started describing the migration situation as a 'crisis,'<sup>42</sup> and the country introduced a temporary highly restrictive law revising the country's asylum and family reunification policies.<sup>43</sup> The law, which applied retroactively to asylum applicants who had entered since 2015, made it much more difficult for recent immigrants to bring their immediate relatives with them to Sweden.<sup>44</sup>

Although the dramatic change in immigration policy may seem sudden, it reflected a gradual movement toward more restrictive, right-wing politics in the country. Four years earlier, the far-right Sweden Democrats won 49 out of 349 seats in Parliament – more than doubling their support from the previous election where they had won 20 seats. In the wake of the migration 'crisis,' the Sweden Democrats crafted a narrative blaming asylum seekers for a perceived increase in violent crime and suggesting that the number of asylum seekers was putting the Swedish state welfare system in jeopardy.<sup>45</sup>

In 2018, support for the party grew again. The Sweden Democrats won 17.9% of the vote, becoming the third largest party in Parliament with 62 seats. The party openly blames immigration for a perceived uptick in crime<sup>46</sup> and supports taking away funding from multicultural initiatives,<sup>47</sup> in line with its slogan "Keep Sweden Swedish."<sup>48</sup>

The rise of extreme-right politics during this time was hardly unique to Sweden. Characterized by their anti-immigration stances and populist rhetoric claiming to defend 'traditional' European culture, parties with values similar to the Sweden Democrats have been on the rise in Europe since the early 2000s.<sup>49</sup> Online, such parties are very active and visible: they were early adopters of digital communication methods.<sup>50</sup> This was especially true in Sweden.

Scholars have noted how political extremists in the country (along with those from the United States) were at the forefront in adopting digital communication technologies<sup>51</sup> and using social media to build communities<sup>52</sup> and spread their ideas to broader, more moderate, audiences. Bartlett et al. (2011), for example, note how extremist groups use social media to reach larger audiences who might be interested in the ideas endorsed by a party enough to 'like' or 'follow' a social media page, but do not yet feel comfortable enough with the rhetoric to become an official party member.<sup>53</sup> The Sweden Democrats, for example, have 30,000 official party members as of 2019, but over 10 times as many followers on their Facebook page (315,473).<sup>54</sup>

Similarly, the Sweden Democrats made modest gains in parliamentary elections, but their online presence soared at the same time. Those who joined #jagärhär in 2016 felt that increase in far-right voices online, and many spoke of it during interviews. Others in Europe also noticed similar increases in xenophobia online, including Hasnain Kazim, a journalist born in Germany to Pakistani parents. After receiving a torrent of anti-Muslim, xenophobic digital messages in 2015, Kazim resolved to respond (often humorously) to each one he received starting on January 1, 2016, and later wrote a book about it.<sup>55</sup>

39. Pew Research Center. 2016. "Number of Refugees to Europe Surges to Record 1.3 Million in 2015." <https://www.pewresearch.org/global/2016/08/02/number-of-refugees-to-europe-surges-to-record-1-3-million-in-2015/> [https://perma.cc/E6BF-CPHP]

40. Emilsson, Henrik. 2020. "Continuity or change? The impact of the refugee crisis on Swedish political parties' migration policy preferences." In *Forced Migration and Resilience*, pp. 99-121. Springer, Wiesbaden, p. 99.

41. Johnson, Simon and Johan Sennero. 2018. "Immigration and Welfare Fears Merge as Sweden Lurches to the Right." Reuters. Available: <https://www.reuters.com/article/us-sweden-election-ljusnarsberg-insight/immigration-and-welfare-fears-merge-as-sweden-lurches-to-the-right-idUSKCN1LL0Z4> [https://perma.cc/JP9D-2TFS]

42. Skodo, Admir. 2018. "Sweden: By Turns Welcoming and Restrictive in its Immigration Policy." Migration Information Source.

43. Lag (2016:752) om tillfälliga begränsningar av möjligheten att få uppehållstillstånd i Sverige [Act on Temporary Limits to the Possibility of Receiving Residency Permits in Sweden]

44. Cerrotti, Rachael. 2017. "Sweden was among the best countries for immigrants. That's changing." *The World*. <https://www.pri.org/stories/2017-09-11/sweden-was-among-best-countries-immigrants-thats-changing> [https://perma.cc/776Y-TPHM]

45. Tomson, Danielle Lee. 2020. "The Rise of Sweden Democrats: Islam, Populism and the End of Swedish Exceptionalism." March 25. Brookings Institute. <https://www.brookings.edu/research/the-rise-of-sweden-democrats-and-the-end-of-swedish-exceptionalism/> [https://perma.cc/FV9K-A8PM]

46. Ibid.

47. Borevi, Karin. 2013. "The political dynamics of multiculturalism in Sweden." *Challenging multiculturalism: European models of diversity*. 138-162.

48. The slogan "Bevara Sverige Svenskt" ("Keep Sweden Swedish") was also the name a neo-Nazi organization that was active in Sweden during the 1980s. The current Sweden Democrats party grew out of this organization. See Peter Vinthagen Simpson. September 13, 2010. "Introducing the Sweden Democrats" *The Local*. <https://www.thelocal.se/20100913/28956>

49. Bartlett, Jamie, Jonathan Birdwell, and Mark Littler. 2011. *The new face of digital populism*. Demos.

50. Barnett, B. A. 2007. *Hate group community-building online: A case study in the visual content of internet hate*. New York: Proceedings of the New York State Communication Association; Ekman, Mattias. 2014. "The dark side of online activism: Swedish right-wing extremist video activism on YouTube." *Mediekultur: Journal of media and communication research* 30, no. 56; Thiesmeyer, L. 1999. "Racism on the Web: Its rhetoric and marketing." *Ethics and Information Technology*. 1(2), 117-125

51. Ekman, Mattias. 2014. "The dark side of online activism: Swedish right-wing extremist video activism on YouTube." *Mediekultur: Journal of media and communication research* 30, no. 56.

52. Barnett, B. A. 2007. *Hate group community-building online: A case study in the visual content of internet hate*. New York: Proceedings of the New York State Communication Association.

53. Bartlett et al., supra note 50

54. As of April 8, 2020.

55. Kazim, Hasnain. 2018. *Post von Karlheinz: Wütende Mails von richtigen Deutschen – und was ich ihnen antworte*. Penguin Verlag



Counterspeakers often find themselves responding to people online they wouldn't meet or listen to in the normal course of their offline lives – people with different backgrounds, interests, and beliefs. Online, counterspeakers and their targets are more likely to encounter each other, and this makes counterspeech possible, painful,<sup>56</sup> and potentially effective. Social media scholars danah boyd and Alice Marwick coined the term “context collapse” to describe how social media allows for the meeting and mixing of social contexts that are generally segmented offline, making our online speech visible to potentially limitless audiences.<sup>57</sup> As they argue, before social media, someone may have had a general sense of who was in the possible audience for their speech. Today, however, people encounter speech that is far more candid and people who share a wider variety of beliefs and speech norms than they would have encountered offline in the past.

Much of #jagärhär's early work focused on countering racist and xenophobic speech. A prime example occurred in the weeks before St. Lucia's Day in 2016. St. Lucia's Day is a festival of light celebrated throughout Scandinavia. Traditionally, young girls dress in long white gowns and wear a crown of candles as they carry a tray of cookies and saffron buns to their families. In 2016, Swedish department store Åhléns posted an ad celebrating the holiday with a photo of a gender nondescript dark-skinned child dressed in the traditional St. Lucia's day costume. The store's social media pages were soon flooded with racist and sexist comments.<sup>58</sup> Many characterized the ad as anti-Swedish. “You are provocative and you are against Swedish culture. You are advocating the death of Swedish culture and complaining about the folks who don't like it”<sup>59</sup> wrote one user. “Looks more like a gingerbread man!” wrote another.<sup>60</sup>

Members of #jagärhär tackled many comments posted below news articles describing the incident. They also directly addressed comments on the Åhléns Facebook page where the ad was posted. “Good picture! Good progress! Love to all and zero tolerance for racism. #jagärhär” wrote one member. Although there were around 200 hateful comments about the ad on the Åhléns Facebook page, there were more than 20,000 ‘likes’ or ‘loves’ on the post, and numerous counterspeech comments, many tagged with #jagärhär.<sup>61</sup> Even though Åhléns eventually pulled the ad at the request of the child's parents, it marked a major turning point for Dennert and her group. #jagärhär received some press attention and the group grew from around 14,000 members to 25,722 in the week following the action.<sup>62</sup>

56. Shin, In-geon, Jin-min Seok, and Youn-kyung Lim. “Too Close and Crowded: Understanding Stress on Mobile Instant Messengers Based on Proxemics.” In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1-12. 2018.

57. Marwick, Alice E., and danah boyd. 2011. “I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience.” *New media & society* 13(1): 114-133.

58. Fiona Keating. Dec 6, 2016. “Swedish department store pulls ad after child model triggers racist abuse by trolls.” *International Business Times*. <https://www.ibtimes.co.uk/swedish-department-store-pulls-ad-after-child-model-triggers-racist-abuse-by-trolls-1595000>

59. “Department store pulls festive ad following racist abuse of dark-skinned child.” 2016. RT. <https://www.rt.com/viral/369237-ahlens-lucia-racist-controversy/> [<https://perma.cc/DHW9-5U3G>]

60. Backlund, Antonia. 2018. “Åhléns hade en icke-vit lucia i sin reklam – då rasade rasisterna.” *NYheter24*. <https://nyheter24.se/nyheter/inrikes/869670-ahlens-hade-en-icke-vit-lucia-i-sin-reklam-da-rasade-rasisterna> [<https://perma.cc/G4U7-GJEX>]

61. Fiona Keating. Dec 6, 2016. “Swedish department store pulls ad after child model triggers racist abuse by trolls.” *International Business Times*. <https://www.ibtimes.co.uk/swedish-department-store-pulls-ad-after-child-model-triggers-racist-abuse-by-trolls-1595000>

62. Interview with author, October 18, 2019. Several of the people interviewed for this research project mentioned learning about the group for the first time because of this incident.

# Research Findings

Instead of focusing their work on preventing future hateful speech (presumably by changing the minds or incentives of those who post it), #jagärhär members fight against its effects – attempting to lessen the impact of the hateful speech by hiding it in the comment threads and by encouraging more counterspeech against it. There is some evidence that this is working. Members flood comment sections with comments that challenge racist, xenophobic, or other hateful narratives, bringing in facts and correcting misinformation where they can. My observation and interviews revealed that they do so in four primary ways:

1. Using Facebook's platform architecture to downrank content they deem to be hateful
2. Influencing the 'movable middle'
3. Activating new counterspeakers
4. Keeping counterspeakers engaged

These methods are discussed in detail below, followed by a section describing the impact that the group has had on other people online.

## 1. Using Facebook's platform architecture

The #iamhere groups from the around the world operate primarily on Facebook,<sup>63</sup> and they do so because their method of responding to hatred has developed around the platform's architecture. Their counterspeech strategy uses the Facebook commenting algorithm as a way to amplify their own comments, while burying what they regard as hatred.

The primary purpose of #jagärhär's Facebook group is to organize their collective counterspeech and direct it into certain online spaces. Group members learn where to counterspeak in two primary ways. The first is through what they call 'actions.' Each day, group members search on Facebook for hateful comments on news articles and other public pages and send those to administrators who confirm that the article's comments meet 'action' requirements. A comment thread would not meet the criteria for an action if there were only a few hateful comments, if it is on a private person's Facebook page or in a closed group, or if the comments contain disagreement rather than hate or slander. For example, the group occasionally gets suggestions for actions from people who simply disagree with the relevant comments or the content of the article. These are politely rejected.

Within the group, at any one time, there is a small set of self-selected people who do most of the work of finding the comment threads for actions. Some people do this for a few weeks or months and then go back to only taking part in actions. Others participate exclusively by suggesting actions. Most never send in suggestions. After receiving suggestions, the moderators choose a few (usually 2-4) to post to the whole group each day, directing members to post and 'like' each other's comments in the relevant comment thread. They try to spread their attention around so as not to be constantly counterspeaking on the same newspaper or group pages.

The second way #jagärhär members get direction is from a daily 'fire extinguisher' post<sup>64</sup> – an open call for group members to share links they find throughout the day to what they see as hateful comments. Each day, a moderator shares a post that begins with the phrase "Dagens Brandsläckare,"

63. At one point, they attempted actions on YouTube and Twitter, but in addition to the problem of not being able to rely on Facebook's commenting algorithm to amplify their comments, they struggled with getting their members to take part in actions on a separate platform from where they organize their actions.

64. Also called a "camp fire" post in some of the groups.

("Today's fire extinguisher") flanked by two fire extinguisher emoji. The fire extinguisher was developed 1-2 years into the project, and it serves several purposes. First, it allows group members to respond quickly, since there is no vetting process as there is for actions. Second, it allows #jagärhär to intervene in cases that are too small for an action post. They use the term 'fire extinguisher' since they are trying to put out small fires quickly with counterspeech, i.e., to stop hatred from spreading and growing large enough for a full group action. Third, a fire extinguisher has become a way for group members to request help if they start counterspeaking on a non-action post, and feel they need backup quickly. Although the fire extinguisher allows members to summon immediate help – directly from others in the group – it is still monitored by the two-person team of daily moderators who remove any links to posts that do not contain hateful or slanderous content.

The group's rules exhort members to "Write what you yourself think and feel."<sup>65</sup> This attitude is evident in the comments they write: not all #jagärhär comments are alike, and they represent a variety of viewpoints and approaches to counterspeaking. In general, however, comments posted by members take one of four forms:

- 1) Comments that seek to correct misinformation or present facts to counter a hateful narrative
- 2) Comments that criticize the tone of hateful comments
- 3) Comments supporting the person or behavior being attacked by the hateful speech
- 4) Comments written in support of other counterspeech comments

Take for example a recent action in which members of the group responded to comments on an article reporting that there had been several confirmed cases of bubonic plague in China. When #jagärhär posted this action, the comment thread was already filled with remarks such as:

*(Original in Swedish)*

Återuppbygg den kinesiska muren utan några öppningar, hela kina borde stängas in från resten av världen, jävla smittohärd hela landet 🇨🇳🇨🇳🇨🇳🇨🇳🇨🇳🇨🇳🇨🇳

Like · Reply · See Translation · 7w



*(Translated to English)*

Rebuild the Chinese wall without any openings, the whole China should be shut down from the rest of the world, fucking contagious country 🇨🇳🇨🇳🇨🇳🇨🇳🇨🇳🇨🇳🇨🇳

Like · Reply · See Original (Swedish) · 7w



And this one, which plays off other comments that derided the diets of Chinese people:

*(Original in Swedish)*

Snart blir det kannibalism där!

Like · Reply · See Translation · 7w

*(Translated to English)*

Soon there will be cannibalism there!

Like · Reply · See Original (Swedish) · 7w

In response, #jagärhär members wrote comments challenging the idea that the diets of Chinese people are uniquely dangerous, correcting misinformation about the plague, and calling many of the comments in the thread racist. They also wrote comments in support of others who were counterspeaking in the thread – both members and non-members. Below are a few examples:

65. Original Swedish: "Skriv vad du själv tänker och tycker"

Människor har blivit sjuka i alla tider av allt möjligt. Just nu är det väldigt uppiskad stämning kring sjukdomar, men böldpest är inte så ovanligt. Det finns dessutom bra bot numera.

Mycket av det vi tycker är delikatesser här i Sverige idag kan också vara sjukdomsframkallande. T.ex. råbiff eller ostron.

Like · Reply · See Translation · 7w · Edited



People have become sick at all times of everything possible. Right now there is a very popular mood around diseases, but bubonic plague is not so unusual. There is also good cure nowadays.

Much of what we think is delicacies here in Sweden today can also be disease-causing. For example. tart or oysters.

Like · Reply · See Original (Swedish) · 7w · Edited



Jag har full förståelse för att detta är oerhört skrämmande. Med rädsla kommer dock inte sällan irrationellt tänkande. Att skriva rent rasistiska saker om ett helt folk är inte ok! Vi har olika matkulturer runt om i världen. Det ena är inte mer rätt än det andra. Bara olika.

Like · Reply · See Translation · 7w



I fully understand that this is extremely scary. With fear, however, does not rarely come irrational thinking. Writing purely racist things about a whole people is not ok! We have different food cultures around the world. One is no more right than the other. Just different.

Like · Reply · See Original (Swedish) · 7w



Jag vet inte om det är coronaeffekten, om reaktionerna i det kommentarsfältet hade varit annorlunda om vi inte just nu befann oss mitt i coronakrisen, men jag blir förskräckt av att läsa många kommentarer. Att man får panik, det kan jag möjligen förstå. Men att man uppmanar till folkmord, att utplåna Kina osv., är det inte lite att överreagera. Tänk om det här hände hos oss. Skulle ni vilja att folka skrev sådana saker om oss?

Like · Reply · See Translation · 7w



I don't know if it's the corona effect, if the reactions in that comment section would have been different if we were not currently in the middle of the corona crisis, but I am horrified by reading many comments. That you panic, I can possibly understand that. But calling for genocide, to wipe out China, etc., it is not a bit to overreact. Imagine if this happened with us. Would you like people to write such things about us?

Like · Reply · See Original (Swedish) · 7w



And this one, replying to another counterspeech comment:

Absolutely right! #jagärhär

Like · Reply · See Original (Swedish) · 7w



Helt rätt! #jagärhär

Like · Reply · See Translation · 7w



Members cite many reasons for choosing to write one type of comment or another. Many members said that they were more likely to write comments that attempt to counter misinformation or challenge hateful narratives if they feel they have some expertise in the topic being discussed. Expertise also informs how some members construct their counterspeech. For example, Fredrik, a 49-year-old academic who lives in Göteborg, Sweden's second largest city, says that he often refers to scientific articles in his counterspeech comments. "I am an academic. I work at the university. So for me, it's very important to follow standards in argument. I see myself as a knowledge producing institution representative."

In addition to (or sometimes in place of) commenting, group members 'like' each other's comments, or other fact-based comments that they see, including those that aren't from #jagärhär members. This trick leverages Facebook's comment algorithm, which rewards engagement. According to Facebook, on pages "with a large number of followers," comments are automatically set to sort by 'relevance' – a ranking determined at least in part by the number of 'likes' or replies that each comment elicits.<sup>66</sup>

66. Facebook Help Center, supra note 24.

When #jagärhär members 'like' each other's comments, it moves them up in the ranking. Ideally, they hope to push their comments into the top section that is visible when one scrolls through one's news feed. Most members with whom I spoke reported both 'liking' comments and writing comments of their own on actions throughout the week. What they choose to do on each action relates to factors such as their level of expertise in the subject being discussed in the article and comments, as well as their daily schedule and energy level. Group administrators note that early in the morning, commuting hours, and late evening are particularly busy times for the group. People tend to 'like' comments early in the day, while riding the bus to work for example, and then engage more deeply by writing and responding to comments later in the evening.

For Elin, a 47-year-old woman who lives just outside of Stockholm, counterspeaking sometimes begins as soon as she wakes up. From bed, she turns on her phone and reads recent comments, adding her 'likes' to ones with which she agrees. In her spare moments throughout the day, or in the evening after everything has quieted down, she again returns to Facebook to check on the actions that were posted during the day. Elin was even more active two years ago, when she went through a painful divorce. During that time, she visited the group's page in nearly every free moment that she had. She enjoyed reading the civil and supportive conversation of the group, and said it served as a kind of 'escape' during a challenging time in her life.<sup>67</sup>

For their efforts to be successful, #jagärhär members must be careful not to amplify the content to which they respond. On Facebook, responding directly to a comment can boost that comment higher in newsfeeds, even if the response is critical. The group's rules therefore state that members should not respond directly to a person or a specific comment. Even if they are writing in response to a specific comment, members write so-called top-level comments<sup>68</sup> (clicking 'reply' on the original post rather than on another comment on that post). By writing their counterspeech as a new comment, rather than as a reply to another comment, others can elevate the counterspeech (by 'liking' it), without also amplifying a hateful comment as they might have done by replying directly to ("engaging with") it. In other words, as one member said, she often doesn't reply directly "because I don't want to give the original post – because of the way the interaction algorithm works – I don't want to give it more views or more power... I don't want to reply to a person, because I don't want to give that person more space."<sup>69</sup>

67. Interview with author, August 7, 2020

68. Also sometimes referred to as a "first level comment"

69. Interview with author, September 6, 2019

70. In many of these studies, it is difficult to distinguish whether the effect on the quality of the conversation is due to changes in who participates or changes in the quality of the contributions. In other words, is the effect of behavioral contagion to encourage more like-minded people to join the conversation or does it actually alter the content of what participants would have otherwise posted?

71. Han, Soo-Hye, and LeAnn M. Brazeal. 2015. "Playing Nice: Modeling Civility in Online Political Discussions." *Communication Research Reports*, 32(1), 20–28.

72. Cheng et al., supra note 21.

73. Friess et al., supra note 10.

74. Ibid p. 15

75. Ibid p. 17

## 2. Addressing the "Movable Middle"

By pushing their own comments up, #jagärhär members try to make their comments the first things (or perhaps the only thing) that people read in reaction to an article. Many studies have documented that the tone of social media comments at the beginning of a comment thread have an impact on the tone of the future discourse within that thread.<sup>70</sup> For example, if a user encounters so-called 'civil' comments, they are more likely to post similarly civil comments.<sup>71</sup> Likewise, several studies have found that exposure to anti-social or uncivil comments make a person more likely to post an anti-social comment.<sup>72</sup> If #jagärhär members are successful in pushing their civil comments to the top and the hateful ones to the bottom, they may well influence at least some of the other users who comment within the same field and may even affect their behavior in other contexts, including offline.

A recent study of #ichbinhier (the German group) suggests that their efforts might be working. Researchers from the Heinrich Heine University Düsseldorf<sup>73</sup> used a dataset of comment threads in which group members engaged between November 01, 2017 and January 31, 2018 to answer two questions: whether comments made by #ichbinhier members were more 'deliberative' than those posted by non-members (researchers coded for rationality, constructiveness, politeness, civility, and reciprocity), and whether deliberative top-level comments were associated with more deliberative second-level comments. Their study found the answer to both questions to be 'yes,' suggesting that discourse norms established or reaffirmed by groups in the #iamhere network can have an impact on the quality of online discourse.<sup>74</sup> The study was somewhat limited in that it only investigated the relationship between top-level comments and direct replies to them<sup>75</sup> – but it is an important step in evaluating the effects of the #jagärhär method.

Pushing their comments to the top of comment threads has other potential impacts as well. Members of #jagärhär said they try to amplify comments that are logically argued, well-written, and fact-based, whether they are written by #jagärhär members or not, because they may be able to reach the larger reading audience – those scrolling through their Facebook feeds who might encounter the article. Some of those people, group members posit, will not have made up their mind yet about the topic being discussed, and therefore could be potentially swayed in different directions by the speech in the comments.

Those working to shift policy or public opinion on social issues such as abortion,<sup>76</sup> migration,<sup>77</sup> or LGBTQ+ issues<sup>78</sup> often call this audience the ‘moveable middle,’ people who do not currently hold strong opinions about a particular topic and are therefore able to be swayed toward one side or the other.<sup>79</sup> Activists generally see the moveable middle as the ideal target of messaging as they are more willing to listen sincerely to an argument than those who openly oppose it.<sup>80</sup> There is support for this strategy in the literature. For example, researchers have found that even a small group of counterspeakers can influence the discourse within an online space if the audience that they are speaking to holds relatively moderate views.<sup>81</sup>

To reach the movable middle, members of #jagärhär generally use two strategies: providing factual information and documenting dissent. Although research indicates that fact-checking is not very effective for changing someone’s mind,<sup>82</sup> some #jagärhär members feel that fact-based arguments are useful in reaching the movable middle. There is a belief, stated by multiple members, that casual readers who encounter these articles and comment sections have a sense that hateful comments often contain misinformation, but are not concerned enough to invest extra time in fact-checking.<sup>83</sup> Members said dispelling myths and making accurate information easily visible would allow the ‘silent readers’ to ‘make up their own minds.’ Noteworthy in these statements is the underlying assumption that, if presented with both accurate and inaccurate information, readers would likely be convinced by #jagärhär’s arguments (which many members described as typically ‘logical’ and ‘well-formulated’), an assumption that presents an inherently optimistic (although possibly misguided) view of the average reader.

A second way in which #jagärhär members try to reach their audience, which many believe is powerful, is to document dissent. One member said he doesn’t counterspeak to make people see that they are wrong, but to show that there are different views. “These comment fields can make the impression that most people are hateful; they’re not,” he stated.<sup>84</sup> Another member shared a similar viewpoint: “Even if you write an answer for that side [those posting hatred], everyone else can read it too. If you go into a place where a lot of bad things are written, then people say, ‘oh, God! That is what everyone thinks!’ But this is not what everyone thinks. A lot of people think differently; and that’s important.”<sup>85</sup>

In her article “Blocking as Counter-Speech,” philosopher Rae Langton describes how readers can ‘block’ the impact of hateful speech by mentally resisting what it is trying to say.<sup>86</sup> The impact of xenophobic speech could be lessened, for example, if people who read the hateful speech understood it to be untrue, and therefore remained unmoved by its message. For this internal blocking to be successful, however, Langton argues that the reader must feel confident that the speech is wrong and must overcome “the fear of being an epistemic outlier—the odd one out, who disagrees not only with the speaker, but also with what everyone supposedly takes for granted.”<sup>87</sup> This is exactly what members of #jagärhär are trying to correct. They want to ensure that when readers encounter hatred online, they see that they are not the only ones who disagree. As one member said, “You are trying to reach every reader to make the reader understand that it seems like everyone in Sweden is against immigration, but that’s not true.”

76. Lane, Victoria. 2019. “Reframing to reach the ‘movable middle’.” <https://ecflive.fairsay.com/reframing-to-reach-the-movable-middle/> [<https://perma.cc/G49Y-6P98>].

77. International Center for Policy Advocacy. n.d. “Step 1: Finding a Focus and Opening.” <http://www.narrativechange.org/toolkit/step-1-finding-focus-and-opening>. [<https://perma.cc/ZB23-CHVJ>].

78. GLAAD and Them. “Why Queer Activists Should Direct Their Message to the ‘Movable Middle’.” Them. <https://www.them.us/story/movable-middle-glaad> [<https://perma.cc/HM2G-35GN>].

79. Ibid.

80. See Chapter 6 of Newlands, Maxine. Environmental activism and the media: the politics of protest. Peter Lang, 2018; GLAAD Media Institute <https://www.youtube.com/watch?v=DWchXfOQrok> [<https://perma.cc/6H7C-TQZD>].

81. Schieb, Carla, and Mike Preuss. 2016 “Governing hate speech by means of counterspeech on Facebook.” In 66th ICA Annual Conference, at Fukuoka, Japan, pp. 1-23.

82. Kolbert, supra note 14.

83. This belief was not backed-up by any specific evidence or anecdotes of the method being effective, and yet multiple group members mentioned this during interviews.

84. Interview with author, August 27, 2019.

85. Interview with author, September 12, 2019.

86. In describing this concept, Langton draws heavily on Austin’s (1962) Speech Act Theory (see Austin, J. L. 1962. How to Do Things with Words. Oxford University Press).

87. Langton, Rae. 2019. “How to undo things with words: Blocking as a way to counter “evil” speech” ABC Religion and Ethics. <https://www.abc.net.au/religion/how-to-undo-things-with-words-blocking-as-counter-evil-speech/10985924> [<https://perma.cc/P6VP-YVDU>].

### 3. Activating new counterspeakers

Convincing others that they are not alone in disagreeing with a hateful comment so that they can more easily 'block'<sup>88</sup> the message is important, but this study identified another mechanism through which #jagärhär's counterspeech fights hatred as well: drawing new counterspeakers into conversations.

During interviews, many members of the group told stories of their own experiences joining #jagärhär – how they had felt alone and hesitant to speak against the hatred they were seeing online. Many also said that they did not counterspeak before joining the group. They were disgusted by the comments that they were reading, but they felt too afraid to say anything.

Mattias has been a member of the group for about three years and a member of the moderator team for over a year. Like many members with whom I spoke, Mattias was not counterspeaking before joining #jagärhär. He was active in environmental justice work but had avoided taking part in online conversations with people he did not know because the commenters were so aggressive, hateful, and quick to spread misinformation. "It was bad to the point where I had decided not to click any comment sections because it led me to so many bad emotions." Mattias is a graduate student studying communication, and during our conversation, he ruminated about how this training has led him to think critically about why he and others choose to participate or not in various types of conversation.

*"I've always been really into communication, so my frustration was even bigger [before joining #jagärhär] because I knew how you can communicate about issues, even emotionally loaded issues, in a way that doesn't devolve into vitriol basically. So it was a period in my life where I said 'I'm just not going to participate in social media at all, it's all so toxic.'"*

Like many members, Mattias first discovered the group by seeing a comment tagged with #jagärhär. He was still hesitant to get involved in the conversations himself, but he supported the group's goals and approach, so he joined.

*"I wasn't really active for a while – I just joined to show my support. Then more and more I started to get involved. So it was more me showing my support in the beginning, then I realized that it actually had an effect and the comments became better and better – and now it's easier for me to have a better conversation. Now more and more the people are commenting in an effort to have a better conversation."*

For Mattias, seeing the discourse improve within comment sections made it easier for him to participate.

One of the most common remarks from those interviewed for this study was that the group made them feel braver. As previously discussed, before #jagärhär, many members stated that they did not feel comfortable entering the comment sections, describing the comments that they used to encounter as predominantly 'toxic,' 'aggressive,' and 'hateful.' A solitary dissenting voice would draw attention and potentially garner attacks. But with the #jagärhär model, members counterspeak as a group, leaving the individual less exposed within the comment thread. Members said this left them "feeling safer" or "more protected."

This is an important finding, as it adds explanatory value to studies like Miškolci et al.<sup>89</sup> Drawing on over 7,500 Facebook comments, Miškolci and his coauthors tested the effectiveness of counterspeech to respond to negative portrayals of the Roma in Slovakia between April 2016 and January 2017 within a select number of comment fields. The study found that counterspeech was not a particularly effective method of changing the behavior of the user who posted anti-Roma comments. It was, however, followed by an increase in the number of pro-Roma comments within a particular comment thread. When combined with the findings from this study, one might posit that counterspeech is able to draw out more counterspeech because each additional counterspeaker is less exposed to potential abuse. As the number of counterspeakers increases within a comment field, the level of risk and related emotional cost decreases for those joining the conversation.

88. Ibid.

89. Miškolci et al. supra note 9.

For comment sections identified by group moderators as official action items, members of #jagärhär can confidently counterspeak knowing that they will not be alone in doing so. For an average action post, about 50 members usually write comments and around 500 'like' various comments. This number fluctuates, depending, at least in part, on the speech to which they are responding. According to Dennert and several group members, when there is a clear 'bad guy' and a sympathetic victim in the content of the article, the actions generally draw more counterspeakers. For example, a post describing a hate crime against a child would likely draw a large number of #jagärhär members.

Many members mentioned that the group's size provides a sense of support that extends beyond simply feeling protected by a herd. Some called it 'support,' others 'trust' or 'camaraderie,' but most mentioned some manner in which being a member of the group allowed them to move more bravely through the online world. "#jagärhär helps in many ways. The effort of many is directed and collected; you can get support, even knowing there are people figuratively behind your cyber back is a comfort" stated one.<sup>90</sup> Another said, "You feel that if I get jumped at, I will have my friends in back to cover me. That is very important."<sup>91</sup> "If you come attack me, there are 10 people who will come to support me. That is incredibly important" echoed a third.<sup>92</sup>

Members of #jagärhär sometimes also provide 'cover' for people outside their group, which in turn recruits some new help for them. Monica, a #jagärhär moderator, was counterspeaking on her own before she joined the group back in December 2016, only a few months after it began. One day, after she responded to a hateful comment about asylum seekers, some of the commenters began attacking her. "I don't remember exactly what they said, but I remember it was aggressive, and that I didn't know exactly what I should do. I thought, should I keep responding? Should I just keep quiet?"<sup>93</sup> Before she had made up her mind, she noticed that others had joined her. People started 'liking' her comment and others started citing statistics about immigration and trying to refute claims that refugees were a danger to Sweden.

The comments included a hashtag: #jagärhär. "I looked it up, and I decided to join. It was just in time. I had started losing some faith that responding was worth it. To see so much hate. That can eat you up at times," she said. But after finding the group, she felt more hopeful. "I thought that I could make a difference with other people. We could do this together."<sup>94</sup>

By including the hashtag in their comments, #jagärhär members demonstrate the fact that they have the group's support and they say that makes them feel braver. When the group first formed, members would tag all of their comments with this hashtag so that other members could easily find and 'like' them. It also helped recruitment, as noted by several members including Monica.

But as the group grew, they changed their strategy. Elin, who found #jagärhär a useful escape during her divorce, said that at some point she began to feel that by using the hashtag, the group was bullying the people to whom they were responding. Elin herself had been bullied as a child, and she was easily able to empathize with readers who disliked seeing hundreds of comments tagged with '#jagärhär' that were written in response to a far smaller number of hateful comments. "I actually wrote to Mina when I was sort of six months in," Elin said. "I said to her that I don't think that it's really clever that everyone use the hashtag because we are too strong. You can't have one person writing something stupid and then have 300 persons just sort of picking at them. Then it turns the other way. Then you have the good guys turning into bad guys because there are too many." Some far-right voices say that the #iamhere groups are de facto censors who muscle their own opinions to the forefront while silencing others.<sup>95</sup>

As time passed, the group norm changed from everyone tagging their comments with #jagärhär to only a few doing so on any given post. Ideally, the message communicated is that many different individuals with unique viewpoints are counterspeaking rather than there being an outpouring of criticism from one unified source. "I was not the only one thinking and talking to Mina I think. I think there were many more than me," said Elin.<sup>96</sup>

90. Interview with author, August 27, 2019.

91. Interview with author, September 12, 2019.

92. Interview with author, September 11, 2019.

93. Interview with author, August 19, 2020.

94. Interview with author, September 18, 2019.

95. Interview with author, February 1, 2018.

96. Interview with author, August 7, 2020.



In my interviews, most members said they generally only reveal their association with the group during their counterspeech in specific circumstances. For example, Lena, the tennis player who had found so much comfort in the group after worrying that Sweden was moving away from her own beliefs, said, "I would use it [the hashtag] in specifically infected commentary fields – if there are a lot of really mean comments and people are attacking each other, then I would put on the hashtag, like armor. It adds a level of protection."

Others, however, felt that the hashtag sometimes made them more vulnerable because it brought them unwanted attention from the aggressive Facebook users whose hateful comments were prompting their counterspeech. Take Monica for example, who joined #jagärhär at the end of 2016 and became a moderator a year or two later. Monica said that, as a moderator, she uses the hashtag in order to show other members that she is there to help if they need it. She is more likely to use the hashtag in the beginning of an action to show group members that they are 'safe,' and not when there are already many counterspeech comments. However, the fact that she is so frequently visible in the comment threads as a member of #jagärhär also means that she has faced more harassment than many other non-moderators. On one occasion, a piece of propaganda for an extremist group was delivered to her home. Even though there was no personal message, and she could not prove that it was targeted, she believes that it was sent to her intentionally to show her that members of the extremist group knew where she lived.

In these examples, members strategically deploy the hashtag, or refrain from using it, because the group's reputation and/or size can offer either benefits or vulnerabilities depending on the circumstances. But considerations about reputation and association with the group can also flow in the other direction. Individual members who have online reputations of their own sometimes avoid using the #jagärhär hashtag to keep their own reputations from tarnishing the group. In 2018, the political editor of a self-declared 'independently liberal' Swedish newspaper (Göteborgsposten), wrote an editorial disagreeing with Swedish Holocaust survivors who compared the ideology and policy priorities of the present-day Sweden Democrats to those held by the Nazis in Germany in the 1930s. In response Fredrik, a #jagärhär member and university professor, wrote a petition calling for her removal from the paper.

His petition received over 800 signatures within the first 24 hours. It also brought a wave of online attacks – social media posts calling him a communist and a traitor, emails to his employer demanding his removal, and even death threats – from those associated with the radical right in Sweden. "It lasted for a week," he said. "I'm pretty used to being in the spotlight, but this was really too much, so since then I've been more cautious about what I post." Fredrik says that although he used to use the #jagärhär hashtag, today he does not:

*"I had a discussion with a friend who said 'your online reputation is already ruined, so they [those posting hatred online] would say that this crazy communist from the university is trying to silence people.' If I use [the hashtag], the discussion isn't about the topic anymore, it's simply about me. And I understand what my friend is talking about. Unless I can counter that image of what is around about me, [people believing he is a communist or a traitor] I understand why it could do more harm than good. So if I would tag my comments, I would provide people ammunition to use my support as an example that the whole initiative is corrupt."<sup>97</sup>*

Both Fredrik and Monica became visible as individuals, distinct from the protective mass of the group. Monica through her work as a moderator (being an early and frequent poster on action threads) and Fredrik by publicly campaigning against a prominent political editor. This visibility brought on attacks that could not be fended off by the support of other group members 'liking' comments or writing their own comments in solidarity with the attacked member. Although the online attacks on Fredrik did not emerge from his work with the group, he now wonders if he should have reached out to Dennert or the other moderators for help after the attacks started. The group's admin team has developed a process for taking reports of online harassment from members and helping them contact the police, if necessary. In the months leading up to the 2018 Swedish general election, the group even developed a separate task force to take reports, as more group members reported harassment. These days, Dennert says that regular members (those not on the team of moderators and administrators) are not harassed often, only occasionally.

97. Interview with author, September 11, 2019.

The possibility of online attacks does not seem to have dampened the confidence of #jagärhär members though. Notably, several group members said that their increased willingness to participate in counterspeech extended beyond #jagärhär group actions, stating that they had changed as individuals, becoming more confident in the value of their own opinions. One said for example, "It has made me stronger, I think. I know that there are a lot of people just like me. I feel stronger, and I think I dare to speak my mind more."<sup>98</sup> Another noted:

*"I speak up more often now online in places where #jagärhär is not involved. I also think it feels a little easier to give my opinion in different situations offline since I became active in the group. It's a good school. You get a lot of practice in patience and methods of dealing with different kind of conversations."*<sup>99</sup>

Elin also described how her participation in the group had changed and empowered her:

*"I have had the words and an interest in writing since I was young, but when I was young, I met a lot of adults who said 'you don't have the language, you don't write well,' sort of pushing me down. So this [joining #jagärhär] was sort of regaining myself saying 'sorry, you're actually wrong. I can use my language. My language is not wrong. I just have to know how to use it, because I know I can touch people by my words. So I sort of reclaimed myself."*<sup>100</sup>

#### 4. Keeping Counterspeakers Engaged

The literature on activism notes that burnout is one of the primary challenges to sustaining participation. Initially coined by psychologist Herbert Freudenberger in 1974, people experience 'burnout' when they lose their original drive for doing their jobs and become physically and mentally exhausted by prolonged stress associated with their work. In research with social justice activists, this loss of drive is most frequently attributed to the toll taken by the intense emotional labor often associated with social justice campaigns.<sup>101</sup> Researchers have also documented the relationship between burnout and a "culture of selflessness" among social justice activists who may feel that, in the context of the huge societal-level challenges they are trying to overcome, taking care of their own mental health would be selfish.<sup>102</sup>

These problems can also exist for online activists. As many studies have demonstrated, offline activists who are able to avoid burnout and remain engaged with causes over many years often do so by developing strong social ties with others in the movement.<sup>103</sup> But strong social ties are not always easy to form in online activism campaigns. In interviews, many #jagärhär members spoke of the emotional energy required to counterspeak and said the work – which they all do as volunteers – can be exhausting. One said, for example, that before deciding whether or not to participate in an action, she asks herself, "How many comments do I have the energy to do?" Another described in more detail how the emotional demands of the work can make it hard to continue. "It's very wearing – one reason why I'm not all that active now. At first, you had to gather all of your courage. Then you feel the support and feel that you are part of the group and it's pretty easy. Then as time passes, it gets harder. You get tired of it. You meet the strangest opinions. There are some [members of #jagärhär] who join other, right-wing groups to start discussions there."<sup>104</sup> I don't do that because I like my peace of mind. You get really tired," she said.<sup>105</sup> Similarly, other group members noted how taxing it can be to continually attempt to counter hateful speech.

98. Interview with author, September 12, 2019.

99. Interview with author, September 19, 2019.

100. Interview with author, August 7, 2020.

101. Goodwin, Jeff, and Steven Pfaff. 2001. "Emotion work in high-risk social movements: Managing fear in the US and East German civil rights movements." *Passionate politics: Emotions and social movements* p282-302. Maslach, Christina, and Mary E. Gomes. 2006. "Overcoming burnout." *Working for peace: A handbook of practical psychology*. p43-49.

102. Rodgers, K. 2010. "'Anger is why we're All Here': Mobilizing and Managing Emotions in a Professional Activist Organization." *Social Movement Studies* 9(3): 273-91.

103. Gladwell, Malcolm. "Why the revolution will not be tweeted." *The New Yorker* 4, no. 2010 (2010): 42-49. Plyler, J. 2006. "How to Keep On Keeping On." *Upping the Anti* 3: 123-34

104. The few members who do this do not do it as representatives of #jagärhär, and the group does not organize these actions.

105. Interview with author, September 18, 2019.

Despite the emotional toll of the work, most #jagärhär members have managed to avoid burnout and have continued counterspeaking for years, building up experience and helping the group to remain sustainable. There are several reasons for #jagärhär's relative longevity, they said. One is the salutary effects of working together in a large group. Research has shown that activists – even those working offline – often face a feeling of isolation stemming from the fact that they deal directly with societal problems that others in their communities seem “unable or unwilling to face.”<sup>106</sup> By coming together as a group and participating jointly in actions, #jagärhär members stave off feelings of isolation. This was evident in the 73% of members interviewed for this study who spontaneously explicitly stated that joining #jagärhär has made them feel less alone.<sup>107</sup> “You don't feel that you are the only one who thinks some way,” said moderator Mattias. “When it comes to a certain topic, you can feel that the only way that people react is with hate. But you can always bring it up in #jagärhär and get a totally different reaction – it's much closer to what you want to see. It's a group with similar values. It's very powerful. It makes me more secure in my own values.” Another member stated, “#iamhere is somewhere where you can charge your batteries somehow.”<sup>108</sup> She later said that it felt like “being with friends, even though you don't know them. But you know they want the same things, so they feel like friends.” Other members said similar things: that despite not actually meeting other members in person,<sup>109</sup> or even really forming strong individual connections online, there was a general feeling of friendship and familiarity because members felt that they had a shared value system with others in the group. Having this feeling of solidarity may prevent people from feeling isolated.

Elin, who was already bruised by her personal life, echoed this sentiment. She described how her then-husband often used very harsh language with her and her two children. The contrast between this language and the civil discourse of #jagärhär members was strong, and reinforced for her by her own memories of being bullied as a child. Language, and the way people can harness it to either help or hurt, became a real passion for Elin. For her, the #jagärhär Facebook page became a sort of refuge, a place where she could go and know she would not encounter the kind of aggressive language she did in her offline life.

The various #iamhere groups also all have their own collective rituals, developed to bolster the mental and emotional well-being of their members, and to fend off burnout by sharing encouraging stories.

The Canadian group, for example, does something called ‘well-wishing Wednesdays’ where group members celebrate individuals<sup>110</sup> who go out of their way to help people. Each week, group administrators select one person, post a bit about their story, and then encourage members to send messages of support to the person being honored. #iamhereCanada supports a wide range of people through their well-wishing Wednesdays; recipients have included Autumn Peltier, a teenager from the Wiikwemkoong Unceded Territory who advocates for clean drinking water for First Nation communities, and Toronto Raptors basketball player Serge Ibaka whose foundation has provided thousands of meals to those in need.<sup>111</sup> These well-wishing posts often draw even more engagement than action posts. “Some people have told me that this [fighting against hatred] is just too hard. They need to see that they aren't struggling by themselves” said one of the group's administrators. The Swedish group does something similar, calling its practice ‘love bombing.’ As with the Canadian group, these posts are quite popular, with one – a tribute to a Swedish man who sewed over 6,000 face masks in his home to give away during the COVID-19 pandemic – receiving over 1,500 ‘likes.’

#Ichbinhier, the German group, has created perhaps the most extensive infrastructure to take care of members. Every evening, a group administrator posts an “Absacker” (‘nightcap’), a post inviting discussion around a topic unrelated to counterspeech, like being stuck at home during the COVID-19 lockdown, or favorite first sentences from books. These posts give members a break from the emotional labor of fighting hatred, while providing a space for group discussion and bonding. Several members of #ichbinhier also created a sister Facebook group called ‘Happy Place für #ichbinhier,’ where members post a steady stream of light-hearted video clips and memes, animal pictures, and feel-good stories. The group is open to all members of the larger #ichbinhier group, and has 851 members at the time of writing.<sup>112</sup>

The specific practices and structures designed to promote self-care, and the general feeling of belonging cited by members, largely attributed to sharing a common moral language and goal, are both likely reasons why members view the #jagärhär Facebook group as a rejuvenating place. They are surely part of the reason why thousands of people have continued to do the unpaid work of responding to online hatred, week in and week out, for years.

106. Maslach and Gomes, *supra* note 101.

107. 16 out of the 22 #jagärhär members interviewed for this study (73%) mentioned no longer feeling alone.

108. Interview with author, October 25, 2019.

109. Even many of the leaders of the various international groups had not met before 2018 when the Dangerous Speech Project brought a few of them together, with other counterspeakers from around the world, for a two-day workshop in Berlin.

110. Generally people in Canada, but occasionally they send support to people living in other countries as well such as the Japanese wheelchair dancer Kenta Kambara.

111. Ricci, Talia. 2019. “The Serge Ibaka Foundation's donation will provide thousands of meals through a drop-in meal program.” CBC News. [https://www.cbc.ca/news/canada/toronto/serge-ibaka-regent-park-partnership-1.5232667?fbclid=IwAR0hSmAeucQmIHNCf3qdBYs11OrrorjeipZR71R0eGxu11J\\_1SvlvdTPlo](https://www.cbc.ca/news/canada/toronto/serge-ibaka-regent-park-partnership-1.5232667?fbclid=IwAR0hSmAeucQmIHNCf3qdBYs11OrrorjeipZR71R0eGxu11J_1SvlvdTPlo) [<https://perma.cc/R883-UFU9>]

112. #ichbinhier had 44,894 members as of May 4, 2020.

# Impact of #jagärhär on non-members and on discourse norms

It is clear that #jagärhär has affected the mental states and behavior of its members. There is also evidence that the group has had a wider impact, affecting the way that individuals outside of the group behave. This can be observed on multiple levels. The first is the effect that the group has on speech in particular comment threads. As discussed above, pushing their comments to the top of comment sections allows #jagärhär to influence the tone of comments that are posted by non-members who may read the thread and write comments. It also provides a pathway to reach those in the 'moveable middle' and convince them that the hateful comments they may have seen are not the predominant views in Sweden.

On a second level, #jagärhär may have helped to make discourse norms on the Facebook pages of newspapers and public groups in Sweden more civil and less xenophobic. There seems little doubt that there has been a notable shift, observed by #jagärhär members and others such as journalists, in the years since the large influx of migrants to Sweden and other European countries. The shift has coincided with #jagärhär's four years of efforts, though we (so far) lack evidence of a causal relationship between the two.

There is anecdotal evidence, however, including among people outside #jagärhär. Eva Burman, editor-in-chief for a network of self-identified 'liberal' regional newspapers in Sweden, told me that from her observations, she believes #jagärhär's counterspeech has had a large impact on the discourse within Swedish comment threads. She said that several years ago, there was not much discussion on her network's newspapers' Facebook pages – and the comments that appeared did not seem to provide a representative sampling of readers' opinions.

"It was kind of weird. It didn't matter what kind of story it was, all of the comments were hatred and right-wing comments. It didn't matter what we were writing about."<sup>113</sup> Now, the comments are more balanced. Burman believes #jagärhär made it feel safe for the paper's readers (who are not members of the group) to comment on articles. The group's actions seem to have diluted hateful comments significantly, though they have not eliminated and may have not even decreased them.

Many members with whom I spoke mentioned that they felt that the discourse in the comment threads had improved over the time that they had been members of #jagärhär. "Let's dial back five years," one member said. "Whenever there was a debate about immigration, there was an absolute majority of people throwing up hate, just drowning everything, and [there was] very little counterspeech because there were so few people doing it, you got attacked." He continued, "Five years ago, the amount of personal attacks you got was enough to deter quite a few, I'm sure. But with #jagärhär, it wasn't just me standing up. They got so many others on the bandwagon that they [those posting hatred] started to disappear in the threads. From being 90% of threads, they went to 20% of the threads."<sup>114</sup>

Graduate student Mattias shared a similar opinion when I asked whether he felt the group had had an effect:

*"I don't really understand why or how it works, but I definitely notice that it does work. I didn't even feel like I could, I mean, I never made comments on public Facebook pages a few years ago. Almost every comment was toxic. What #jagärhär has done, somehow, I don't know how, people can now make comments expressing their opinions and they don't have to be toxic. I mean sure, there still are toxic people, but there is always someone there to back you up."*

*I asked Mattias if he felt the change had come from there being fewer people posting hatred online. "No," he said, stopping to think for a moment. "It's just that there are more reasonable people. Now the status quo is more balanced, so people who go along with status quo are less toxic."*

113. Interview with author, November 8, 2019.

114. Interview with the author, September 11, 2019.

The hateful commenters didn't go down in number, but because the number of counterspeakers has increased dramatically, the proportion of hatred has changed. This means that those who encounter a comment thread are less likely to reach the conclusion that the opinions expressed in the hateful comments are the prevailing view in Sweden. They are also less likely to refrain from adding their own counterspeech. The model of counterspeech used by #jagärhär documents dissent to hatred and supports other counterspeech comments, two actions that make online spaces in which #jagärhär members are present feel less risky for others who may be contemplating adding their own counterspeech comments. Thus although the actual hatred may not have changed, the impact of that speech likely has.

The perceived changes in discourse norms also motivate #jagärhär group members to comment more since they feel they are having a real impact. As one member said:

*"The thing is that since the group has grown so much, you actually find that the tone at least in normal media has changed a lot. There are many more people who are contradicting racist things, so it's easier. You don't have to go by #iamhere, you can just go and start commenting, and you will always have people supporting you."*<sup>115</sup>

Another agreed:

*"People aren't as afraid to give their opinions. For me personally, it means I am more prone to comment now than ever before when I just avoided any commenting at all. I realize that if I set the tone with the first comment when my local paper publishes something it makes a difference."*<sup>116</sup>

The growth and influence of #iamhere groups has brought them criticism as well. For example, some people have come to see them as a coordinated effort to silence voices with which they disagree. One #jagärhär member told me, "I have seen that quite a lot of people believe it's like a sect. So if you have 10 comments, and they all have the hashtag, then one guy will comment, 'oh the sect is here!'"<sup>117</sup> Others (generally those promoting far-right political ideas) have described the group as censors. The German group has faced similar criticism. Critics have called them the 'Stasi 2.0.' and a Facebook user once called the group 'opinion gorillas' in a comment thread.<sup>118</sup> As discussed above, criticisms such as these have led to some members feeling that using the group's hashtag can cause more harm than good within a comment thread. Members worry that when it's apparent that they are highly coordinated, their counterspeech will not be trusted or deemed 'authentic.' "I did it (used the hashtag) in the beginning," said one #jagärhär member who has been with the group for four years. "For me, it felt like it was protecting me and helping me say what I wanted to say. After a while, it wasn't working because when some people saw the hashtag, people reacted right away. They weren't reading what I was writing – they just saw the hashtag." As the groups around the world continue to grow and become more well-known, this criticism is likely to continue.

115. Interview with author, September 18, 2019.

116. Interview with author, September 30, 2019.

117. Interview with author, July 24, 2019.

118. Some in the group believe this user intended to write 'opinion guerillas,' as she was criticizing the left-wing nature of the group. Interview with author, February 1, 2018.

## Conclusion

This study suggests that there is value in collective action against hateful speech online. Although previous research has documented that counterspeaking as a group may have an impact on discourse in certain cases,<sup>119</sup> this study is the first to consider the various consequences of doing collective counterspeech for members of the group. Group members report feeling braver and more willing to enter difficult conversations. Additionally, they spoke of many aspects of the #jagärhär model that may prevent burnout, a major obstacle to sustainability for many social change initiatives.

Are they succeeding? The findings from this research suggest that they likely are. 'Success' for #jagärhär members isn't measured by how many hateful comments exist in a conversation, but by how much space has been created for alternative viewpoints. As one member said:

*"In the end, it's about democracy, it's about debate, it's about freedom of speech that people will have the courage to say what they think. If you have lots of hate comments, maybe you are afraid, and you don't want to say what you think. But if we are 10-20 people arguing against the hate then I imagine that others will also want to do so, so that not only the people screaming the highest can say their opinion."*<sup>120</sup>

This new way of conceptualizing effectiveness poses challenges for measurement and calls for further inquiry. But the findings of this study demonstrate many kinds of consequences and potential audiences of counterspeech. And in doing so, it helps researchers begin to design studies to measure various impacts of counterspeech.

## Acknowledgements

I am grateful to Dangerous Speech Project Executive Director Susan Benesch for her thought-provoking questions, meticulous editing, and encouragement, and for guiding this project from start to finish. I owe so much to Mina Dennert and all the members of the #iamhere network, especially those who so generously shared their time and experiences with me; this paper wouldn't exist without you. I am also grateful to Tonei Glavinic for their very thoughtful editing and general support over the years, and to Richard Wilson and Joshua Garland, who both provided insightful comments on the paper.

The Dangerous Speech Project's work is supported by the John D. and Catherine T. MacArthur Foundation, the Ford Foundation, and the Open Society Foundations.

119. Garland et al, supra note 17; Friess, supra note 118.

120. Interview with author, October 18, 2019.