DIE DYNAMIKEN VON HATE SPEECH UND COUNTER SPEECH IN SOZIALEN MEDIEN

EINE ZUSAMMENFASSUNG DER WISSENSCHAFTLICHEN FORSCHUNG

Dr. Katarzyna Bojarska[1][2]

EINLEITUNG

Vorurteile, Hass und Hetze gegen Individuen und Gruppen wurden durch die Geschichte hindurch beobachtet. Ihre Dynamiken und Auswirkungen sind bereits Jahrzehnte vor dem digitalen Zeitalter wissenschaftlich erforscht und beschrieben worden. Die feindliche Darstellung und Stereotypisierung von Gruppen und Minderheiten als "anders", "unterschiedlich" oder "gefährlich" befördert einen Prozess der Dehumanisierung. Dieser Effekt kann schnell eskalieren, wenn feindselige Rhetorik mithilfe von Rundfunk, Printmedien oder digitalen Medien ein großes Publikum erreicht und zu gewalttätigen Hassverbrechen, einschließlich Völkermord, führt.

In den letzten Jahren wurden in Europa zunehmend fremdenfeindliche, nationalistische, islamfeindliche, rassistische und antisemitische Einstellungen offen ausgedrückt. Die Auswirkungen beschränken sich nicht auf feindselige Rhetorik, sondern drücken sich in konkreter Gewalt gegen Gruppen und Individuen aus. Der deutsche Justizminister Heiko Maas verdeutlicht: “Häufig bleibt es nicht bei Hassreden, oft sind Worte die Vorstufe von Taten. Dass aus ‚geistiger Brandstiftung’ viel zu oft Gewalt wird, zeigt der sprunghafte Anstieg von Angriffen auf Flüchtlingsunterkünfte: Im Jahr 2014 hat sich die Zahl der Taten im Vergleich zum Vorjahr verdreifacht. "(Maas, 2015, S. 6)

Während sich die Kanäle der medialen Verbreitung im Laufe der Zeit weiterentwickelt haben, wirken weiterhin die gleichen Mechanismen gruppenbasierter Feindseligkeit. Dabei haben soziale Medien wie Facebook oder Twitter neue Formen der sozialen diskursiven Partizipation hervorgebracht. Ihre Nutzer*innen (auch: User) tragen in beispielloser Weise zur Verbreitung von Vorurteilen, Fake News und Feindseligkeit gegenüber Geflüchteten und anderen marginalisierten Gruppen bei. Die Verbreitung von Hass in digitalen Medien stellt daher einen „sozialen Notfall“ mit konkreten individuellen, politischen und sozialen Konsequenzen dar.

Eine kürzlich durchgeführte Internet-Umfrage ergab, dass eine Mehrheit der Internutzer*innen (67 %) bereits auf Hasskommentare im Internet (auch: Hate Speech) gestoßen sind (forsa., 2017). Obwohl in Deutschland das Netzwerkdurchsetzungsgesetz (NetzDG) aus dem Jahr 2017 die Seitenbetreiber sozialer Mediendienste dazu verpflichtet, Hate Speech, Fake News und illegales Material innerhalb kurzer Zeit nach einer Meldung eines solchen Inhalts zu entfernen, deckt das Gesetz nicht jede Form von Hate Speech ab. Trotz der von Land zu Land unterschiedlichen historisch, kulturell und politisch geprägten Regelungen erstreckt sich Hate Speech über nationale Grenzen hinweg. Es werden weiterhin Strategien benötigt, um die Sichtbarkeit derjenigen zu stärken, die sich gegen den Hass im Netz positionieren.

Der vorliegende Bericht bietet einen zusammenfassenden Überblick über den derzeitigen Stand der wissenschaftlichen Forschung, die dazu geeignet sein könnte, Strategien zur wirksamen Bekämpfung von Online-

[1] Centre for Internet and Human Rights, Europa-Universität Viadrina Frankfurt (Oder)

[2] Übersetzung: Sarah Alami, Centre for Internet and Human Rights, Europa-Universität Viadrina Frankfurt (Oder)

Gefördert vom Bundesministerium für Familie, Senioren, Frauen und Jugend im Rahmen des Bundesprogramms Demokratie leben!

Hate-Speech weiterzuentwickeln und demokratische Botschaften im Netz effektiv sichtbar zu machen. Im Fokus steht daher der Vergleich der Dynamiken von Hate Speech und Gegenrede (auch: Counter Speech), im Sinne einer Positionierung gegen Hassposts und Hass im Netz. Bartlett und Krasodomski-Jones definieren Counter Speech als eine gemeinsame „crowd-sourced response“ auf Extremismus oder hasserfüllte Inhalte. (2015, p. 5)

Für ein besseres Verständnis der Herausforderung effektiver Counter Speech wird zunächst Hate Speech als Phänomen definiert, gefolgt von der Diskussion des rechtlichen Status in Deutschland sowie der von Hate Speech ausgehenden sozialen und individuellen Auswirkungen. Anhand einer Untersuchung der verfügbaren Forschungsergebnisse wird der Frage nachgegangen, wie und warum sich Hate Speech in sozialen Medien verbreitet und versucht zu erklären, welche unterschiedlichen Dynamiken die Verbreitung von Hate Speech und Counter Speech beeinflussen. Abschließend kommen wir auf die sich ergebenden Möglichkeiten effektiver Strategien der Gegenrede zurück.

DEFINITION UND JURISTISCHE EINORDNUNG VON HATE SPEECH

Hate Speech kann als Ausdruck einer Feindseligkeit definiert werden, die sich auf die wahrgenommene (zum Beispiel religiöse, nationale, ethnische oder geschlechtliche) Gruppenzugehörigkeit von Individuen oder sozialen Gruppen bezieht. Nach der Definition des Europarates umfasst der Begriff Hate Speech dabei „jegliche Ausdrucksformen, welche Rassenhass, Fremdenfeindlichkeit, Antisemitismus oder andere Formen von Hass, die auf Intoleranz gründen, propagieren, dazu anstiften, sie fördern oder rechtfertigen, unter anderem Intoleranz, die sich in Form eines aggressiven Nationalismus und Ethnozentrismus, einer Diskriminierung und Feindseligkeit gegenüber Minderheiten und Menschen mit Migrationshintergrund ausdrückt." (Europarat Ministerkomitee, 1997)

Während Hassrede oder Hate Speech im deutschen Rechtssystem keine juristischen Begriffe sind, wird der in diesem Kontext relevante Rechtsbegriff der Volksverhetzung seit langem im Strafgesetzbuch (StGB) als Straftat anerkannt, unabhängig davon, ob diese online oder offline begangen wird:

„Wer in einer Weise, die geeignet ist, den öffentlichen Frieden zu stören,

1. gegen eine nationale, rassische, religiöse oder durch ihre ethnische Herkunft bestimmte Gruppe, gegen Teile der Bevölkerung oder gegen einen Einzelnen wegen seiner Zugehörigkeit zu einer vorbezeichneten Gruppe oder zu einem Teil der Bevölkerung zum Hass aufstachelt, zu Gewalt- oder Willkürmaßnahmen auffordert oder
2. die Menschenwürde anderer dadurch angreift, dass er eine vorbezeichnete Gruppe, Teile der Bevölkerung oder einen Einzelnen wegen seiner Zugehörigkeit zu einer vorbezeichneten Gruppe oder zu einem Teil der Bevölkerung beschimpft, böswillig verächtlich macht oder verleumdet,

wird mit Freiheitsstrafe von drei Monaten bis zu fünf Jahren bestraft.“[3]

Seit dem 1. Oktober 2017 verpflichtet das Netzwerk-Durchsetzungsgesetz (NetzDG) die Anbieter sozialer Netzwerke dazu, Hate Speech, Fake News und illegale Inhalte innerhalb von 24 Stunden nach Eingang einer Beschwerde zu entfernen. Das neue Gesetz wurde dafür kritisiert, privaten Unternehmen sozialer Netzwerke, und eben keinen staatlichen Strafverfolgungsbehörden, die Entscheidungsmacht zu geben ohne juristische Kontrolle digitale Inhalte zu löschen. Dies birgt nicht nur die Gefahr einer umfassenden Zensur und untergräbt die Meinungsfreiheit, sondern stellt zugleich ein schlechtes Beispiel für Länder dar, die durch ähnliche Gesetze politische Kritik eindämmen möchten.

Die Kontroverse über das deutsche NetzDG reiht sich in eine umfassende, politische und akademische ‚Redefreiheit vs. Hate Speech’ - Debatte ein. Ein Großteil der wissenschaftlichen

[3] § 130 Abs. 1 StGB

Gefördert vom Bundesministerium für Familie, Senioren, Frauen und Jugend im Rahmen des Bundesprogramms Demokratie leben!

Peer-Review-Publikationen konzentriert sich in diesem Kontext auf den rechtlichen Status von Hate Speech und darauf, ob und in welchem Umfang und mit welchen Instrumenten und Technologien diese reguliert werden sollte. Die weit gefasste Debatte geht über den Rahmen des vorliegenden Berichts hinaus, der sich weniger auf staatlich durchgeführte gesetzliche Maßnahmen, als auf Gegenredestrategien konzentriert, die für einzelne Nutzer*innen sozialer Medien nützlich sein könnten.

GEFÄHRDUNG DURCH HATE SPEECH

Hate Speech stellt sich auf unterschiedlichen Ebenen als schädlich dar. Es birgt das Potenzial, durch die Beeinflussung von Einstellungen und tatsächlichem Verhalten, den sozialen Frieden zu stören (Müller & Schwarz, 2018), dies geht bis hin zu schweren Hassverbrechen wie Völkermord (vgl. Fyfe, 2017; Maravilla, 2008). Online-Hass stellt einen fruchtbaren Boden für das Auftreten von noch mehr Hass dar, da dieser als „sozialer Beweis“, als Erlaubnis (Brodnig, 2016; Clay, 2017) für "angemessene" Einstellungen und Verhaltensweisen gewertet werden kann. (vgl. Anderson, Brossard, Scheufele, Xenos, & Ladwig, 2014) Online-Hass desensibilisiert in diesem Sinne die Öffentlichkeit für verbale Gewalt und verstärkt Vorurteile (Soral, Bilewicz, & Winiewski, 2018), belohnt mit sozialer Akzeptanz, bestraft die Gegenstimmen und lässt diese verstummen (Brodnig, 2016; Coustick-Deal, 2017). Insgesamt gefährdet Hate Speech die körperliche Sicherheit und das psychische Wohlbefinden der betroffenen Gruppenmitglieder. (Baldauf, Banaszczuk, Koreng, Schramm & Stefanowitsch, 2015a; Coustick-Deal, 2017; Gelber & McNamara, 2016) Mehrere der oben genannten Studien bedürfen einer eingehenderen Diskussion.

Das 20. Jahrhundert hat die Bedeutung der Massenmedien (z. B. Rundfunk und Printmedien) bei der Verbreitung von Hass aufgezeigt. Die Resultate eskalierender Entmenschlichung haben mit Blick auf den Holocaust und den Völkermord in Ruanda Gewaltverbrechen in seiner extremsten Form hervorgebracht (Fyfe, 2017; Maravilla, 2008). Eine aktuelle Studie von Müller und Schwarz (2018) legt nahe, dass derselbe Mechanismus auch für die Rolle der digitalen Medien des 21. Jahrhunderts gilt. Die Studie zeigt deutlich den Zusammenhang zwischen der Exposition gegenüber Hate Speech in sozialen Medien und konkreter tatsächlicher Gewalt. Mit einer fundierten Methodik ziehen die Autoren einige ursächliche Rückschlüsse, wie sich hasserfüllte Social-Media-Aktivität gegen Geflüchtete auf der Facebook-Seite der Partei Alternative für Deutschland (AfD) in konkreten Gewalttaten gegen Geflüchtete äußern. Die Autoren fassen zusammen:

> "Mit diesen Maßnahmen stellen wir fest, dass Hassverbrechen gegen Geflüchtete in Gebieten mit höherer Facebook-Nutzung in Zeiten hoher, sich online äußernder Anti-Geflüchteten-Stimmung, überproportional zunehmen. Dieser Effekt ist besonders ausgeprägt bei gewalttätigen Vorfällen gegen Geflüchtete, darunter Brandstiftung und Körperverletzung. Dies deutet auf eine Rolle der sozialen Medien bei der Verbreitung einer deutschlandweiten Anti-Flüchtlingsstimmung hin" (eigene Übersetzung, Müller & Schwarz, 2018, S. 3)[4].

Um mögliche nicht kontrollierte Faktoren auszuschließen, interpretierten die Forscher die Ergebnisse auch mit einem quasi-experimentellen Rückgriff. Sie finden heraus, dass in den Wochen größerer lokaler Internetunterbrechungen der festgestellte Effekt hoher Anti-Flüchtlingsstimmung auf Hassverbrechen (im Vergleich zu den Gemeinden, die nicht von Internetausfällen betroffen waren), deutlich reduziert war. Auch

[4] “Using these measures, we find that anti-refugee hate crimes increase disproportionally in areas with higher Facebook usage during periods of high anti-refugee sentiment online. This effect is especially pronounced for violent incidents against refugees, such as arson and assault. Taken at face value, this suggests a role for social media in the transmission of Germany-wide anti-refugee sentiment.” (Müller & Schwarz, 2018, S. 3)

Gefördert vom Bundesministerium für Familie, Senioren, Frauen und Jugend im Rahmen des Bundesprogramms Demokratie leben!
EUROPA-UNIVERSITÄT VIADRINA FRANKFURT (ODER)

deutschlandweit beobachteten die Autoren, dass "die Wirkung von Flüchtlingsposts auf Hassverbrechen im Wesentlichen in Wochen großer Facebook-Ausfälle verschwindet" (eigene Übersetzung, Müller & Schwarz, 2018, S. 4). Vor diesem Hintergrund scheint die Bedeutung sozialer Medien, zu gewalttätigen Hassverbrechen anzustiften und gewalttätiges Verhalten zu beeinflussen, unbestreitbar.

Es wird weithin anerkannt, dass Hate Speech eine ernsthafte Bedrohung für die physische Sicherheit der von Hate Speech betroffenen Gruppenmitgliedern darstellt. Laut der "Chronik flüchtlingsfeindlichen Vorfälle" (o. J.) der Amadeu Antonio Stiftung wurden im Jahr 2015 insgesamt 1249 gemeldete Übergriffe gegen asylsuchende Personen und ihre Unterkünfte in Deutschland verzeichnet, in 2016 sind es 3769 und in 2017 liegt die Zahl bei 1939.

Während zum Thema Hate Speech, der rechtlichen Lage, den möglichen Ausdrucksformen und Hate Speech verfassenden Personen und Strukturen viel geschrieben wurde, sind die Forschungsergebnisse zu den psychologischen Schäden für die von Online-Hate-Speech betroffenen Personen knapp. Daher müssen wir zumindest teilweise auch von bestehender Forschung zu allgemeineren psychologischen Auswirkungen von Vorurteilen ausgehen. Die besonders tiefgreifende Wirkung gruppenbasierter Vorurteile auf betroffene Personen verdeutlich sich oftmals durch den im Vergleich zu nicht betroffenen Personen höheren Verbrauch emotionaler Ressourcen. Dieser ist mit der ständigen Notwendigkeit verbunden, mit offener Diskriminierung sowie mit Mikroaggressionen umzugehen, die im Alltag erlebt werden und sich in der digitalen Welt durch den *Online-Enthemmungseffekt*, d.h. die fehlende Hemmniswahrnehmung in der Online-Kommunikation im Vergleich zur Face-to-Face-Kommunikation, zusätzlich verstärken.

Die konstant erhöhte Wachsamkeit und mentale Bereitschaft, mit offen gezeigten Vorurteilen oder Mikroaggressionen umzugehen oder darauf zu reagieren, führt zu chronisch erhöhtem Stress, sogenanntem *minority stress*, der zu negativen gesundheitlichen Folgen wie Depressionen oder Angstzuständen führen kann (Meyer, 1995, 2003).

Es ist mit erheblichen emotionalen Belastungen verbunden, Hate Speech als Mitglied einer betroffenen Personengruppe ausgesetzt zu sein. (Coustick-Deal, 2017; Gelber & McNamara, 2016; Mullen & Smyth, 2004). Unter anderem Gefühle von Ungerechtigkeit, Hilflosigkeit, Angst und Bedrohung zählen zu den psychologischen Auswirkungen. Die Situation der Betroffenen stellt bereits an und für sich eine erhebliche Belastung dar, so dass sich womöglich die Entscheidung einer direkten Reaktion der Betroffenen auf hasserfüllte Inhalte als zu große emotionale Anstrengung herausstellt.

Gelber und McNamara haben die außergewöhnliche qualitative Studie "Evidencing the harms of hate speech " (2016) durchgeführt, in der sie die folgenden Formen der durch Hate Speech hervorgerufenen Beeinträchtigungen und Schäden auflisten, die von betroffenen Personen erfahren werden, darunter: die Einstufung der anvisierten Personengruppe oder Person als minderwertig; das zum Schweigen bringen der Opfer; Stress; das Risiko der Zerstörung des eigenen Selbstwertgefühls; die Einschränkung der Bewegungs- und Vereinigungsfreiheit; die Verletzung der Menschenwürde; die Aufrechterhaltung von Machtungleichgewichten; die Verbreitung von negativen Stereotypen und Animierung Außenstehender zu schädlichem Verhalten, die Normalisierung negative Stereotype und diskriminierendes Verhalten auszudrücken und die Ermutigung der Öffentlichkeit hasserfülltes Verhalten zu übernehmen.

DIE DYNAMIK VON HATE SPEECH UND COUNTER SPEECH IN DIGITALEN MEDIEN

DER SOZIALE BEWEIS

Gefördert vom Bundesministerium für Familie, Senioren, Frauen und Jugend
im Rahmen des Bundesprogramms Demokratie leben!
EUROPA-UNIVERSITÄT VIADRINA FRANKFURT (ODER)

Der jüngste Anstieg von Fremdenfeindlichkeit, Islamophobie und Flüchtlingsfeindlichkeit in zahlreichen europäischen Ländern ist kaum zu übersehen. Die Radikalisierung der Einstellungen fiel dabei mit zunehmenden Medienberichten über die steigende Zahl von Geflüchteten und asylsuchenden Personen in der Europäischen Union zusammen. Das Regionale Informationszentrum der Vereinten Nationen für Westeuropa berichtet:

> "In einigen Ländern löste die Flüchtlingskrise eine Welle der Solidarität aus und viele Freiwillige vor Ort waren in Zusammenarbeit mit den zentralen Behörden entschlossen, die Neuankömmlinge in ihren Städten willkommen zu heißen. In anderen Ländern ist das Gegenteil passiert und eine restriktive Grenzpolitik in Verbindung mit giftiger Rhetorik hat ein offen feindseliges Umfeld für Geflüchtete und Migrantinnen und Migranten geschaffen". (Eigene Übersetzung, Regionales Informationszentrum der Vereinten Nationen für Westeuropa, 2016)[5]

Es ist davon auszugehen, dass der vorherrschende Ton der Medienberichterstattung, über anfänglich unbekannte Sachverhalte oder frühe Interpretationen ihrer Bedeutung durch die Öffentlichkeit, die gesellschaftliche Reaktion und das daraus resultierende Verhalten beeinflussen können. Es ist eine mögliche Erklärung für die unterschiedliche Art und Weise, wie sich die Bildung von sozialen Einstellungen zum gleichen Thema in verschiedenen Ländern oder Bevölkerungsgruppen entwickelt hat. Wenn sich die gesellschaftliche Stimmung hin zu einer gefühlten Legitimation von Hass und Hetze entwickelt, besteht die Gefahr einer weiteren Zunahme gruppenspezifischer Menschenfeindlichkeit.

Dies führt uns zur klassischen Erforschung des sozialen Einflusses und individueller Konformität in der Herstellung von Mustern „angemessenen" Verhaltens, d.h. zu normativem sozialen Einfluss und informativem sozialen Einfluss (sozialer Beweis). Wir sprechen von einem *normativen sozialen Einfluss*, wenn wir uns dem Verhalten anderer anpassen, um akzeptiert und gemocht zu werden (Asch & Guetzkow, 1951; Asch, 1956; Aronson, Wilson, & Akert, 2010*). Der soziale Beweis* oder *informative soziale Einfluss* ist wiederum ein psychologisches Phänomen, das in ungewohnten oder mehrdeutigen Situationen auftritt, in denen wir das Verhalten anderer nachahmen, weil wir nicht wissen, was das angemessene Verhalten sein sollte. Wir nehmen dabei an, dass sich andere auf bestimmte Weise verhalten, weil sie mehr Wissen besitzen als wir (Sherif, 1935; Baron, Vandello, & Brunsman, 1996; Aronson ua, 2010). Im Zusammenhang mit der Rolle von Online-Hate Speech scheint der normative und informative soziale Einfluss eine plausible Erklärung für die schnelle Bildung von Einstellungen und Verhaltensmustern als Reaktion auf die Medienberichterstattung zu zunächst unbekannten Themen zu liefern.

HATE SPEECH ERHÖHT DAS MISSTRAUEN, VERTRAUTHEIT IMMUNISIERT GEGEN DIESEN EFFEKT

Ein interessantes Experiment von Anderson, Brossard, Scheufele, Xenos und Ladwig (2014) zeigt, wie inziviles Äußerungen in digitalen Medien dazu beitragen können, negative Einstellungen zu formen und welche individuellen Voraussetzungen diesen Effekte einschränken können. Die Autoren untersuchten die Rolle von „Inzivilität" im Sinne einer offensiven Meinungsäußerung am Beispiel

[5] "In some countries the refugee crisis sparked an outpouring of solidarity and many local volunteers together with central authorities were committed to making the newcomers arriving in their towns feel welcome. In other countries, however, the opposite happened and restrictive border policies combined with a toxic rhetoric have created an openly hostile environment for refugees and migrants." (United Nations Regional Information Centre for Western Europe, 2016)

Gefördert vom Bundesministerium für Familie, Senioren, Frauen und Jugend im Rahmen des Bundesprogramms Demokratie leben!

der Risikowahrnehmung von Nanotechnologie, ein für die meisten Internetnutzer unbekanntes Thema. Sie entwarfen ein Experiment, in dem die Teilnehmer gebeten wurden, einen wissenschaftlichen Blog-Eintrag über die Vorteile und Risiken von Nanosilber zu lesen, gefolgt von Leserkommentaren, die entweder in inziviler oder höflicher Sprache formuliert wurden. Anschließend wurden sie gebeten, das mit der neuen Technologie verbundene Risiko zu bewerten.

Die Forscher machten einige interessante Beobachtungen, darunter die Feststellung, dass Teilnehmende, die bereits mit dem Thema vertraut waren, unabhängig vom Ton der Benutzerkommentare, die Risiken der Nanotechnologie als geringer einschätzten als Teilnehmende, denen das Thema zunächst unbekannt war. Außerdem bewerteten wiederum diejenigen das Risiko auch als höher, die sich in der Selbsteinschätzung dazu fähig sahen, Nanotechnologie sachkundig zu beurteilen. Die Studie zeigte auch, dass Teilnehmende mit bereits bestehenden positiven Einstellungen zu Nanotechnologie nach Lesen des von inzivilen Leserkommentaren gefolgten Blogeintrags die Risiken als geringer einschätzten als Teilnehmende, deren anfängliche Einstellungen nicht positiv waren. Solche Unterschiede waren bei denjenigen Teilnehmenden, die die Version mit höflich formulierten Leserkommentaren erhielten, nicht ersichtlich. Dies deutet daraufhin, dass letztere mehr durch den tatsächlichen Inhalt des Artikels als durch emotionale Bemerkungen anderer Lesender beeinflusst worden sein könnten. Ähnliche Beobachtungen wurden in Bezug auf die Religiosität der Teilnehmenden gemacht. Religiösere Teilnehmende, die die Version mit inzivilen Kommentaren erhielten, schätzten die Risiken der Nanotechnologie höher ein als weniger religiöse Teilnehmende. Dieser Effekt verschwand unter denjenigen Teilnehmenden, deren Kommentarspalte eine höfliche Diskussion enthielt. Die Autoren erwägen, dieser Effekt könne durch eine religiöse, werteorientierte Beurteilung der Nanotechnologie als Störung der "natürlichen Ordnung" erklärt werden, die durch den emotionalen Ton in den Kommentarspalten eher angestoßen würde, als durch eine ruhige, höfliche Diskussion des Themas.

Abschließend lässt sich feststellen, dass die bereits vorhandene Auseinandersetzung mit dem Thema, bestehende positive Einstellungen und ein niedriges Maß religiöser Identifikation, unabhängig voneinander den Effekt des negativen Einflusses inziviler Kommentare auf die Risikowahrnehmung im beschriebenen Experiment eingeschränkt haben und so bestimmte Faktoren eine positive Wahrnehmung der Nanotechnologie hervorbringen. Auf der anderen Seite begünstigen mangelndes Vorwissen, eine fehlende anfängliche positive Einstellung und die hohe Religiosität eine Beeinflussung durch inzivile Kommentare hin zu mehr Misstrauen gegenüber unbekannten Phänomenen.

Soziale Einflüsse erklären möglicherweise die Entwicklung der unterschiedlichen Einstellungen gegenüber Geflüchteten und anderen Minderheiten. Abhängig davon, ob unsere erste Auseinandersetzung mit einem zuvor unbekannten Thema auf informative oder höchst affektive Art und Weise stattfindet, abhängig von den Informationsquellen, aus denen wir unser Wissen beziehen und als zuverlässig einschätzen, abhängig von den Verhaltensweisen, die wir um uns herum beobachten und als einen sozialen Beweis für unsere eigene Reaktion ansehen, abhängig von unserem Willen und der Motivation, über ein bestimmtes Thema zu lernen oder unsere Überzeugungen herauszufordern, können wir positive oder feindliche Einstellungen, Vertrauen oder Angst entwickeln und handeln entsprechend.

ANGST SCHÜRT HASS, FEHLENDE KONSEQUENZEN ERLEICHTERN HASSERFÜLLTE ÄUßERUNGEN

Die Rolle von Angst, Bedrohung und Unsicherheit für den Erwerb und Erhalt konservativer

Gefördert vom Bundesministerium für Familie, Senioren, Frauen und Jugend
im Rahmen des Bundesprogramms Demokratie leben!

Einstellungen wurde in mehreren Studien nachgewiesen. Im Jahr 2003 fassten Jost, Glaser, Kruglanski und Sulloway in ihrer meta-analytischen Überprüfung politischer Einstellungen zusammen:

> "Menschen machen sich politischen Konservatismus (zumindest teilweise) zu eigen, weil er dazu dient, Angst, Sorge und Unsicherheit zu verringern; um Veränderung, Störung und Ambiguität zu vermeiden; und Ungleichheit zwischen Gruppen und Individuen zu erklären, zu ordnen und zu rechtfertigen."(Eigene Übersetzung, Jost, Glaser, Kruglanski, & Sulloway, 2003)[6]

Sowohl bei Konservativen als auch bei Liberalen beobachteten Nail und McGregor zwei Monate nach den Terroranschlägen am 11. September eine im Vergleich zum vorherigen Jahr signifikante Verschiebung in Richtung der politischen Rechten (Nail & McGregor, 2009). In einem erweiterten Team führten dieselben Forscher ein Experiment durch, bei welchem den Teilnehmenden eine Bedrohungssituation dargelegt wurde, die daraufhin eine konservative Einstellungswandlung hervorrief. (Nail, McGregor, Drinkwater, Steele & Thompson, 2009). Im Jahr 2017 verdeutlichten Napier, Huang, Vonasch und Bargh in einem Experiment den umkehrenden Effekt: Die Angstreduktion bei konservativen Teilnehmenden bewirkte eine Wendung hin zu liberalen Einstellungen. (Napier, Huang, Vonasch & Bargh, 2017).

Es ist wahrscheinlich die zugrundeliegende Angst und Unsicherheit im Gegensatz zu Sympathie, Hoffnung oder Verzweiflung, die die unterschiedliche Dynamik von Hate Speech und Counter Speech in digitalen Medien erklärt. Wissenschaftliche Untersuchungen unterstützen den ersten Eindruck, nach dem Gefühle von Wut online eher Verbreitung finden als solche von Freude oder Trauer (Fan, Zhao, Chen & Xu, 2014), so dass auch Hass sich leichter verbreitet als positive Emotionen oder Gegenstimmen (Bartlett & Krasodomski-Jones, 2015). Wie bereits erwähnt, scheinen rechtsgerichtete Einstellungen und daraus resultierende Hate Speech zumindest teilweise durch Angst vor dem Fremden und durch das Bedürfnis nach Sicherheit angetrieben zu werden (Napier u. A., 2017). Wenn Hate Speech durch eine zugrunde liegende Angst und durch Misstrauen motiviert ist, stellt dies eine spontane Handlung dar, die sich aus einem Schutzbedürfnis und dem Drang ergibt, andere zu warnen und zu überzeugen, dasselbe zu tun. Das Äußern hasserfüllter Inhalte trägt dann dazu bei mit emotionalen Belastungen umzugehen und ein zufriedenes Gefühl zu hinterlassen, da einer als vermeintlich guten und gerechten Sache gedient würde. Darüber hinaus wirkt der *Online-Enthemmungseffekt* als das Auftreten verminderter Empathie, welche sich aus der besonderen Bildschirmsituation ergibt. Die verletzende Handlung erscheint anders als in realen Interaktionen als frei von Konsequenzen. (Brodnig, 2016; Terry & Cain, 2016). Terry und Cain beschreiben den ineinandergreifenden Effekt der Anonymität von Verfassenden und Empfangenden, asynchroner Kommunikation und Nicht-Sichtbarkeit:

> "Erstens besteht aufgrund der Anonymität, die mit computervermittelter Kommunikation verbunden ist, die Möglichkeit eine wechselnde Online-Identität anzunehmen und sich im Wesentlichen hinter einem die Identität nicht preisgebenden Pseudonym oder Benutzernamen zu verstecken. Diese Form der dissoziativen Anonymität ermöglicht die persönliche Identität und moralischen Ansprüchen zu separieren und sich frei zu fühlen, Feindseligkeit und Kritik ohne Auswirkung auf die eigene Psyche zu äußern. In ähnlicher Weise können sich Nutzende sozialer Medien am anderen Ende der Kommunikation innerlich

[6] "People embrace political conservatism (at least in part) because it serves to reduce fear, anxiety, and uncertainty; to avoid change, disruption, and ambiguity; and to explain, order, and justify inequality among groups and individuals." (Jost, Glaser, Kruglanski, & Sulloway, 2003)

Gefördert vom Bundesministerium für Familie, Senioren, Frauen und Jugend im Rahmen des Bundesprogramms Demokratie leben!

distanzieren, indem sie diese unbewusst als Avatare oder Benutzernamen anstelle von tatsächlichen Personen ansehen.

Zweitens kann Online-Kommunikation asynchron verlaufen, so dass nicht umgehend auf Online-Konversationen reagiert werden muss und von den Nachwirkungen der Online-Diskussionen Abstand genommen werden kann, bis hin zu einem Abstreiten verfasster feindseliger und einschüchternder Kommentare.

Drittens, sogar in einer in keiner Weise anonymen Umgebung (digitale Krankenakte, E-Mail-Korrespondenz, Blogs) liegt es in der Natur der Online-Kommunikation, dass Individuen für andere unsichtbar sind und ein Ausblenden jeglicher Art von Augenkontakt oder körperlicher Reaktion der anderen Person ermöglicht wird. Ein wesentlicher Teil der traditionellen Face-to-Face-Kommunikation neigt dazu, nonverbal zu sein (z. B. Körpersprache, Tonfall), und ohne diese Signale fehlt Online-Konversationen ein wesentliches Element des Verstehens." (Eigene Übersetzung, Terry & Cain 2016, S. 2).[7]

DIE KOGNITIVE ANSTRENGUNG, HATE SPEECH AUSZUDRÜCKEN ODER DIESER MITHILFE VON COUNTER SPEECH ENTGEGENZUWIRKEN, IST UNAUSGEWOGEN

Eine online ausgedrückte Äußerung (auch: Posting) von Hate Speech oder das Teilen einer solchen Äußerung in den sozialen Medien ist in der Regel impulsiv, sorglos, innerlich motiviert und beinhaltet keine signifikanten kognitiven oder emotionalen Anstrengungen.

Eine größere Anstrengung stellt womöglich die Unterdrückung des hasserfüllten oder wütenden Gefühls dar und nicht, dieses freizulassen. Im Gegensatz zu Hate Speech ist der Akt der Gegenrede nicht spontan, sondern gesteuert, nicht aktiv, sondern reaktiv. Es erfordert eine bewusste Entscheidung und beinhaltet beträchtliche kognitive und emotionale Anstrengungen. Die möglichen Konsequenzen und unangenehmen Folgen der direkten Konfrontation werden bewusst wahrgenommen, so zum Beispiel durch die eigene Reaktion Aufmerksamkeit zu erregen und Beleidigungen ausgesetzt zu sein. Eine Entscheidung zur gezielten Gegenrede erfordert oftmals einen unverhältnismäßig hohen emotionalen Aufwand und Ressourcenverbrauch im Vergleich zu der impulsiven, selbstbelohnenden und affektiven Handlung des Verfassens oder des Teilens eines hasserfüllten Posts in sozialen Medien (vgl. Coustick-Deal, 2017). Dies könnte die Zurückhaltung vieler Internetnutzerinnen und -nutzer erklären, die schweigen, wenn sie Hate Speech ausgesetzt sind.

Laut Meinungsfreiheitsaktivistinnen und -aktivisten und der offiziellen Haltung von Facebook stellt

[7]"First, the anonymity associated with computer-mediated communication may permit people to possess an alternate online identity and essentially hide behind a non-identifying pseudonym or username. This form of dissociative anonymity allows people to separate from in-person identity and moral agency, thereby freeing them to express hostility and criticism without any effect to the psyche. Similarly, online users may dissociate t hose at the other end of the communication by subconsciously viewing them merely as avatars or usernames instead of actual persons. Second, as online communication can be asynchronous, individuals do not have to manage immediate reactions to online conversations and can remove themselves from the repercussions of online discussions, even avoiding ownership for hostile and intimidating comments.Third, even in a completely non-anonymous environment (i.e., computerized medical record, e-mail correspondence, blogs), the nature of online communications is such that individuals are physically invisible to others, permitting them to disregard any type of eye contact or physical reaction of the other person(s). A significant portion of traditional face-to-face communications tends to be nonverbal (e.g., body language, tone of voice), and without these cues, online conversations lack an essential element of understanding." (Terry & Cain 2016, p. 2)

Gefördert vom Bundesministerium für Familie, Senioren, Frauen und Jugend im Rahmen des Bundesprogramms Demokratie leben!

Counter Speech ein wirksameres Mittel gegen Hate Speech dar als das Entfernen solcher Inhalte durch Website-Administratoren (Bartlett & Krasodomski-Jones, 2015, S. 4). Coustick-Deal kritisiert, dass eine solche Beurteilung Machtungleichgewichte und Privilegien nicht berücksichtige (2017):

„Die Art und Weise, wie Counter Speech befürwortet wird, ist so, als gäbe es eine Art Gleichgewicht, das so funktioniere: (...) Rassisten sprechen = Rassisten hören ihren Opfern zu. (...) Gegenreden können sich jedoch nur jene erlauben, die zunächst Stimmen haben. Es sieht eher folgendermaßen aus: Nazi spricht -> Tausende von Unterstützerinnen und Unterstützern stimmen zu -> die Gegner werden angegriffen. Es gibt kein solches Gleichgewicht, wenn jemand als Antwort damit droht ihre Familie zu töten." (Eigene Übersetzung, Coustic-Deal 2017)[8]

Coustic-Deal weist auf die “unsichtbaren Kräfte” (Eigene Übersetzung, ebd., 2017) hin, die eine Person davon abhielten, überhaupt sprechen zu können:

“Zu sehen, dass Menschen belästigt werden, hält Mitglieder derselben Gruppe davon ab, sich zu äußern. Wenn wir über Überwachung sprechen, verwenden wir auch den Ausdruck ‚Chilling-Effekt’ - und Belästigung funktioniert auf die gleiche Art und Weise. Das Wissen ständig überwacht zu werden, hindert uns daran uns frei auszudrücken. Derselbe Zensureffekt tritt durch Belästigung auf, wenn uns die Angst vor missbräuchlichen Handlungen zum Schweigen bringt." (Eigene Übersetzung, ebd., 2017)[9]

FILTERBLASEN SIND SCHWER ZU BREMSEN

In der Tat haben Bartlett und Krasodomski-Jones herausgefunden, dass Counter Speech auf Facebook ein deutlich geringeres Potenzial hat, ein so breites Publikum zu erreichen (2015). Die Forscher analysierten 27.886 Beiträge, die über einen Zeitraum von zwei Monaten auf 150 öffentlichen Hate-Speech- (124) und Counter-Speech-Seiten (26) hochgeladen wurden, hauptsächlich aus Großbritannien, Frankreich und Italien, mit 25.522 bzw. 2.364 Beiträgen. Sie sammelten auch 8,4 Millionen verknüpfte Interaktionen, d.h. Likes, Shares und Kommentare. Ihre Ergebnisse werden in ihrem Bericht mit dem Titel "Counter-speech: Examining content that challenges extremism online” beschrieben. Unter den vier Arten von Posts (Links, Fotos, Statusmeldung, Videos) wurden die meisten Interkationen in der Kategorie Fotos festgestellt. Als „Feierlich“, gefolgt von „wütend“, wird die auf rechten Seiten beliebteste Stimmung von Posts bezeichnet. Auf Counter-Speech-Seiten war dies „lustig“ / „satirisch“. Dieser Befund könnte darauf hindeuten, dass neben Wut - die viral meistverbreitete Emotion in digitalen Medien - Humor der Gewinner innerhalb der Teilmenge der positiven Emotionen sein könnte.

Jede Interaktion auf öffentlichen Facebook-Seiten, also das Posten oder Teilen von Inhalten durch Nutzerinnen und Nutzer der Plattform, spiegelt sich im personalisierten Newsfeed der digitalen Facebook-Freunde wieder (abhängig von den verwendeten Datenschutzeinstellungen). So besteht die Möglichkeit, dass auch andere Lesende, die keine Gruppenmitglieder sind oder einer bestimmten Facebook-Seite nicht folgen, in Interaktion mit von Freunden geteilten Inhalten

[8] “The way ‘counter speech’ is advocated is as though there is some kind of balance which works like this: (…) Racists speak = racists listen to their victims. (…) However, counter speech is actually only afforded to those who have voices to begin with. It’s more like: Nazi speaks -> thousands of his supporters speak with him -> his opponents are attacked. There is no balance when someone replies to your speech by threatening to kill your family.” (Coustic-Deal 2017)

[9] “Seeing people harassed stops members of that same group from speaking out. When we talk about surveillance, we also use the phrase ‘chilling effect’ — and harassment operates in much the same manner. The knowledge that we are under constant surveillance stops us from expressing ourselves freely. This same censoring effect happens through harassment, when the fear of abuse silences us.” (Coustic-Deal 2017)

Gefördert vom Bundesministerium für Familie, Senioren, Frauen und Jugend im Rahmen des Bundesprogramms Demokratie leben!

treten können. Die Nutzung von Social-Media-Inhalten wird jedoch durch Algorithmen geregelt, die jedes Profil individualisieren und im Newsfeed vor allem Inhalte anzeigen, die aufgrund des Algorithmus als am ehesten interessant eingestuft werden (die gleiche Regel gilt auch für Suchmaschinen im Internet). So entstehen *Filterblasen* und *Echokammern*. In ihnen werden wir hauptsächlich mit Inhalten versorgt, die unseren eigenen Überzeugungen ähneln, während wir von Ansichten isoliert werden, die sich von unseren eigenen unterscheiden. Dies erzeugt den Eindruck, dass die meisten anderen Menschen unsere eigenen Überzeugungen teilen.

Bartlett und Krasodomski-Jones (2015) untersuchten die Verbreitung von Hate Speech und Counter-Speech-Inhalten, aus Hate Speech und Counter-Speech-Seiten, im individualisierten Newsfeed. Hierfür berechnete sie den Anteil von Posts, der im Newsfeed von denjenigen Profilen angezeigt wurde, die auf der entsprechenden Facebook-Seite, auf der der Inhalt ursprünglich gepostet wurde, kein „Like“ gesetzt hatten. Sie kamen dabei zu dem Schluss: “Rechtspopulistische Seiten sind wesentlich effektiver beim Posten von Inhalten, die über ihr Netzwerk von ‚Likes’ und ‚Followern’ hinausgehen. Für Counter-Speech-Seiten (und populistische Seiten des rechten Flügels) sind Videos die effektivste Art von Inhalten, die nach Veröffentlichung ein breiteres Publikum erreichen." (Eigene Übersetzung, Bartlell & Krasodomski-Jones 2015)”[10] Populistische rechtsgerichtete Links, Fotos, Videos und Statusmeldungen erhalten durchschnittlich 50%, 52%, 68% und 21% ihrer Likes von Personen, die die ursprüngliche Seite nicht abonniert haben. Im Vergleich dazu sind es für Inhalte von Counter-Speech-Seiten nur 18%, 5%, 26% bzw. 7%. Ähnliches gilt für den Anteil der Kommentare.

Im Gegensatz zu den negativen Affekten, die der Hetzrede zugrunde liegen und Nutzende sozialer Medien motiviert, hasserfüllte Inhalte oder Meinungen zu teilen, erwecken Botschaften zur Unterstützung von Gleichberechtigung, friedlicher Koexistenz und Empathie offenbar weniger Interesse. Abgesehen von der Angst vor einer persönlichen Konfrontation und der persönlichen Belastung durch Beleidigungen, können auch andere Gründe verhindern, dass Nutzerinnen und Nutzer sozialer Medien prodemokratische, vielfältige oder antirassistische Inhalte teilen. Womöglich nehmen viele Facebook-User demokratische Botschaften als so offensichtlich wahr, so dass es als wenig Sinn machend erscheint, diese weiter zu verbreiten. Vor allem der Filterblasen-Effekt könnte dann verstärkend wirken, da der Glaube entstehen kann, demokratische Werte seien für fast jeden offensichtlich.

Während hasserfüllte Inhalte oft als Mittel dienen, um eine Warnung auszusprechen oder um auf vermeintliche soziale Bedrohungen aufmerksam zu machen, und sich folglich schnell verbreiten, könnten pro-demokratische Inhalte als weniger aufsehenerregend und daher weniger affektiv aufgeladen wahrgenommen werden. Insbesondere lösen diese selten Angst aus, die ein viel stärkeres Handlungsmotiv darstellt als positive Gemütserregung oder Sympathie (Fan u. A., 2014; Nail & McGregor, 2009; Nail u. A., 2009).

LEITFADEN FÜR EFFEKTIVE GEGENREDE

Benesch, Ruths, Dillon, Saleem und Wright (2016) beschreiben in ihrem Bericht "Überlegungen für eine erfolgreiche Gegenrede" zwei Möglichkeiten, wie "erfolgreiche Gegenrede" (auf Twitter) verstanden werden kann: "Die erste ist Sprache (Text oder visuelle Medien), die einen positiven Einfluss auf den ursprünglichen (hasserfüllten) Twitter-Nutzer hat und seinen Diskurs, wenn nicht sogar seinen Glauben, verändert. Dies wird in der

[10] "Populist right wing pages are significantly more effective at posting content which goes beyond their network of page fans. For counter-speech pages (and populist right wing pages) videos are the most effective type of content to post to reach a broader audience.” (Bartlell & Krasodomski-Jones 2015)

Regel durch eine Entschuldigung oder einen Widerruf oder die Löschung des ursprünglichen Tweets oder Accounts angezeigt. Der zweite Erfolgstyp besteht darin, die Diskursnormen des ‚Publikums' einer Counter-Speech-Konversation positiv zu beeinflussen: alle anderen Twitter-Nutzer oder ‚Cyberbystanders', die einen oder mehrere der relevanten Tweets lesen. Diese Wirkung ist beim Studium von Counter Speech ‚in freier Wildbahn' schwer abzuschätzen, aber sie kann anhand von längeren Online-Konversationen, die eine zivilisierter Form beibehalten sowie an Counter Speech, die zu weiteren Counter-Speech-Reaktionen führt, erkennbar sein." (Eigene Übersetzung, Benesch et al., S. 2)[11] Je nach Ansatz können sich unterschiedliche Strategien als nützlich erweisen.

Auffallend wenig wissenschaftliche Forschung über erfolgreiche Gegensprech-Strategien wurden bisher veröffentlicht. Die meisten verfügbaren (und wertvollen) Guidelines, DOs und DON'Ts, sind Sammlungen von erfahrungsbasierten Beobachtungen und Schlussfolgerungen von Aktivistinnen und Aktivisten digitaler Medien, die in verschiedenen Anti-Hate-Speech-Projekten tätig sind, und nicht die Ergebnisse von streng kontrollierten wissenschaftlichen Analysen. Sie sind in Form von digital abrufbaren Publikationen oder auf Websites verfügbar.

Ich werde nun versuchen, einige forschungsbasierte Vorschläge für eine erfolgreiche Counter Speech und zur Verhinderung der Online-Verbreitung von Hate Speech zu machen. Abschließend möchte ich auf einige weitere nützliche Materialien und Empfehlungen hinweisen.

DAS VERHALTEN DES „HATERS" BEEINFLUSSEN

ZIVILES ONLINE-ENGAGEMENT FÖRDERN, INDEM AN DIE MENSCHLICHKEIT DER VON HASS BETROFFENEN PERSONEN ERINNERT WIRD: Munger (2017) testete experimentell, wie sich die Online-Sanktionierung von Rassismus auf hasserfüllte weiße männliche Twitter-Nutzer auswirkt. Er benutzte Bots, um eine Erinnerung zu twittern, nachdem ausgewählte Benutzer rassistische Äußerungen verwendeten: "Hey, denk nur daran, dass es echte Menschen gibt, die verletzt werden, wenn du sie mit dieser Art von Sprache belästigst." (Eigene Übersetzung, Munger 2017)[12] Die wahrgenommene Hautfarbe und der soziale Status der Bots wurden verändert, das wahrgenommene männliche Geschlecht der Bots zugleich konstant gehalten. Anschließend analysierte er die rassistische Aktivität der angeschriebenen Nutzer über einen Zeitraum von einem Monat. Bots, die als weiße Männer mit hohem Status auftraten (ablesbar anhand der Anzahl von Followern), erreichten die längste dauerhafte Veränderung, im Sinne einer signifikanten Reduzierung der zukünftigen Verwendung von rassistischen Beleidigungen, bei den angeschriebenen Nutzern. Leider waren die Effekte von Bots, deren Profilbild schwarze männliche Personen darstellten sowie von Bots mit geringem sozialen Status weniger wirksam. Der Autor erklärt den Effekt, indem er ihn auf eine "In-Group"-Identifikation zurückführt, ohne das weiße männliche Privileg als alternative Hypothese zu erwähnen. Der Ausschluss von Frauen vom

[11] "The first is speech (text or visual media) that has a favorable impact on the original (hateful) Twitter user, shifting his or her discourse if not also his or her beliefs. This is usually indicated by an apology or recanting, or the deletion of the original tweet or account. The second type of success is to positively affect the discourse norms of the 'audience' of a counterspeech conversation: all of the other Twitter users or 'cyberbystanders' who read one or more of the relevant exchange of tweets. This impact is difficult to assess when studying counterspeech "in the wild" as we have, but it may be indicated by long conversations that remain civil, and by counterspeech that leads to others counterspeaking." (Benesch et al, S. 2)

[12] "Hey man, just remember that there are real people who are hurt when you harass them with that kind of language." (Munger 2017)

Gefördert vom Bundesministerium für Familie, Senioren, Frauen und Jugend im Rahmen des Bundesprogramms Demokratie leben!

Experiment und die mangelnde Berücksichtigung der spezifischen Frau-zu-Frau, Frau-zu-Mann und Mann-zu-Frau-Voraussetzungen erschweren die genaue Art der Wirkungsweise zu bestimmen.

Eigene Privilegien für guten Zweck nutzen: In Übereinstimmung mit den oben genannten Ergebnissen und unter Berücksichtigung der Argumentation von Coustick-Deal (2017), nach der nicht jeder gleichermaßen in der Lage ist, sich selbst gegen Hate Speech zu verteidigen wird folgendes empfohlen: Verwende dein weißes, männliches, heterosexuell oder Cisgender-Privileg, um Hass entgegenzuwirken, indem du klar die Erwartung zum Ausdruck bringst, dass die soziale Norm, andere mit Respekt zu behandeln, eingehalten wird. Mit anderen Worten, wenn du nicht aus einer der benachteiligten Gruppen stammst, kannst du dich engagieren, indem du im Fall von Hate Speech an die Individualität und Menschlichkeit der von Hate Speech Betroffenen erinnerst. So trägst du dazu bei, den Effekt der Online-Enthemmung zu verringern und eine gesunde Kultur der Online-Debatte zu befördern.

Bekämpfe problematische Online-Aktivität durch private Nachrichten: Dieser Ansatz wurde von Rafael, Dinar und Heyken erwähnt (Dinar&Rafael, 2017; Rafael, Dinar & Heyken, 2017). Das öffentliche Bloßstellen/Beschämen stellt eine Bedrohung des Selbstwertgefühls der bloßgestellten Person dar, und kann zu einer starken Motivation führen, in die öffentliche Konfrontation zu gehen, um das Selbstwertgefühl wiederherzustellen. Das Ziel ist dann die Verteidigung der anfänglichen Position, da ein öffentlicher Meinungswandel als Gesichtsverlust interpretiert werden könnte, insbesondere infolge eines aggressiven Tons der öffentlichen Konfrontation. Die Intervention könnte einfacher werden und nachhaltigere Wirkung zeigen, wenn sie persönlicher, eher besorgt und in privater Form erfolgt.

Eine Möglichkeit, dies zu erreichen, ist, Besorgnis zu äußern, indem dem problematischen User eine private Nachricht geschickt wird und ihr oder ihm das öffentliche Bloßstellen erspart wird. Dies könnte in ähnlicher Weise wie im obigen Experiment geschehen, indem an die Menschlichkeit der von Hate Speech betroffenen Personen erinnert und über die konkrete negative Wirkung von Hass in sozialen Medien informiert wird. Dies kann zum Beispiel in folgender Form erfolgen: "Hi X, ich habe gesehen, dass du ein Meme gepostet hast [beschreibe das Meme]. Ich habe mich gefragt, warum du es geteilt hast. Ich weiß, dass du eine achtsame Person bist und wollte dich nur wissen lassen, dass das Teilen solcher Inhalte konkret Menschen verletzt. Es ist mit wenig Aufwand verbunden, es zu posten, aber es gibt Menschen, die geschlagen und beleidigt werden, weil andere diese Art von Memes verbreiten. Sie wirken sich negativ auf ihren/unseren Alltag aus. Ich würde mich freuen, wenn du diesen Beitrag entfernen könntest und in Zukunft nichts Ähnliches veröffentlichen würdest". Diese Technik sollte für User, die (noch) keine rechtsradikalen Befürworterinnen und Befürworter sind, besser funktionieren.

DRITTE BEEINFLUSSEN

ANGST NEHMEN: Es wurde experimentell nachgewiesen, dass das Wegfallen oder Verringern von Ängsten dazu beiträgt, konservative Einstellungen in Richtung liberalerer Haltungen zu verschieben und umgekehrt. Dies ist als allgemeiner Vorschlag zu verstehen, da die Verringerung von Angst verschiedene Formen annehmen kann. Angstreduktion als Strategie, um die Verbreitung von Hate Speech zu verhindern, dürfte bei Nutzerinnen und Nutzern sozialer Medien, die sich (noch) nicht stark mit den Werten und Überzeugungen der Rechten identifizieren, effektiver sein.

STÄRKE SOZIALE EINFLÜSSE UND ERMUTIGE FREUNDINNEN UND FREUNDE DAZU, COUNTER SPEECH SICHTBAR ZU UNTERSTÜTZEN: Filterblasen sind nicht hilfreich, um Counter-Speech-Bemühungen zu verbreiten. Leider ist die

Gefördert vom Bundesministerium für Familie, Senioren, Frauen und Jugend im Rahmen des Bundesprogramms Demokratie leben!

Verbreitung von Anti-Hate-Speech-Inhalten in sozialen Medien weniger effektiv als die Verbreitung hasserfüllter Inhalte, die das Publikum außerhalb der Filterblase gleichgesinnter Personen eher erreichen (Bartlett & Krasodomski-Jones, 2015). Dadurch wird die Gegenrede weniger sichtbar. Um die Wahrscheinlichkeit zu erhöhen, dass digitale Inhalte die Filterblase auch verlassen, könnte es eine Strategie sein, ein Beispiel zu geben oder einen sozialen Beweis zu liefern, vor allem - aber nicht ausschließlich - auf öffentlichen Facebook-Seiten.

Es wurde argumentiert, dass nur einige Nutzende sozialer Medien genug Mut, Privilegien, Macht oder Motivation haben, um Online-Hass direkt entgegentreten zu können. Es ist daher wichtig, die stille Mehrheit an „Zuschauenden" zu ermutigen, ihre Unterstützung für diejenigen zu zeigen, die aktiv gegen Hate Speech vorgehen.

Counter Speech kann unterstützt werden, indem die Kommentare und Beiträge der Gegenrednerinnen und Gegenredner mit Likes unterstützt werden, dies erhöht zugleich die Sichtbarkeit der Kommentare. Noch besser ist es, wenige Worte der Unterstützung zu schreiben. Die Machtbalance in Diskussionen ist oft unausgeglichen und auf eine Fülle inziviler Kommentare kommen nur einzelne Gegenstimmen. Likes allein stellen sich eventuell für Dritte nicht als ausreichend sichtbare Zeichen der Unterstützung dar.

Abgesehen davon, Likes zu verteilen, könnte es sinnvoll sein, Unterstützung durch kurze Kommentare auszudrücken: "X ist richtig", "Ich stimme X zu", "Ich bin auf deiner Seite, X", "X macht einen guten Punkt", "X, deine Worte überzeugen mich". Solche kurzen Antworten, die eher nicht zum Ausgangspunkt für Beleidigungen werden, könnten anderen stillen Leserinnen und Lesern („Cyber-Zuschauenden") einen *sozialen Beweis* liefern und sie dazu ermutigen, mehr Unterstützung zu verdeutlichen. Eine solche Unterstützung kann zugleich das Selbstverstrauen der *Hater* schmälern, die sich in Auseinandersetzung mit einzelnen Gegenstimmen ansonsten stark fühlen.

Eine andere Möglichkeit, um Unterstützung für pro-demokratische Ansichten nach außen zu tragen, könnte darin bestehen, demokratische Botschaften im Profilfoto mit aufzunehmen, zum Beispiel durch das Hinzufügen eines entsprechenden Fotorahmens (verfügbar auf Facebook) oder durch das Verwenden bestimmter Hashtags. Während die Facebook-Filterblasen viele Aktivitäten unserer Kontakte ausfiltern, insbesondere wenn ihre Überzeugungen sich von unseren unterscheiden, könnte die im Newsfeed der Facebook-Freunde angezeigte Änderung des Profilfotos außerhalb der Filterblase liegen.

POSTE UND ERSTELLE COUNTER-SPEECH-MEMES: Bartlett und Krasodomski-Jones (2015) stellten fest, dass die erfolgreichste Form von Gegenrede die visuelle Form, den satirischen Ton und die konstruktive (und nicht beleidigende) Ausrichtung annimmt.

BETEILIGE DICH AN GRUPPEN UND SEITEN AUS DEM BEREICH COUNTER SPEECH UND ANTIDISKRIMINIERUNG: Diese stellen nützliche Hilfsquellen dar, um selbstbewusst argumentieren zu können. (*Everyday Feminism* bietet zum Beispiel täglich „empowernde" Artikel über Geschlecht, sexuelle Orientierung, Rassismus und Intersektionalität.) Facebook-Gruppen können auch hilfreich sein, um sofort Unterstützung zu erhalten, wenn ein sozialer Beweis nötig ist. Ein Beispiel für eine solche Facebook-Gruppe ist #ichbinhier.

Wer nützlichen Facebook-Seiten folgen möchte, aber nicht weiß, wo diese zu finden sind kann Freund*innen nach Empfehlungen aus den Bereichen Anti-Diskriminierung, Anti-Rassismus, LGBTQ-Unterstützung oder Feminismus fragen. Innerhalb der Gruppen kann erneut nach Empfehlungen gefragt werden, so dass sich ein großes Repertoire ergibt.

WAS FÜR DICH OFFENSICHTLICH IST, KÖNNTE FÜR DEINE FACEBOOK-KONTAKTE UND FÜR DIE MITGLIEDER DER GRUPPEN, DENEN DU ANGEHÖRST,

Gefördert vom Bundesministerium für Familie, Senioren, Frauen und Jugend im Rahmen des Bundesprogramms Demokratie leben!

NICHT OFFENSICHTLICH SEIN: Denk daran, dass die Vertrautheit mit einem bestimmten Thema gegen die negative Wirkung von Online-Hass immunisiert. Dieser Effekt kann womöglich durch die verringerte Angst vor bereits Bekanntem erklärt werden. Steigere innerhalb deiner sozialen Kontakte die Vertrautheit mit bestimmten Themen, indem du hilfreiche Artikel teilst, die dir beim Erkenntniszuwachs geholfen haben: zum Beispiel, wie gegen Rassismus und Sexismus vorgegangen oder wie die Transgender-Gemeinschaft besser unterstützt werden kann. Wenn bereits Wissen über verschiedene Formen von Diskriminierung vorliegen, können Links mit anderen geteilt werden und interessante Auszüge aus dem Text hinzugefügt werden. Ist der Artikel in einer Fremdsprache verfasst? Dann ist es möglich, den Artikel zu verknüpfen und eine kurze Zusammenfassung in der Muttersprache anzufügen. Einige Ihrer Freundinnen und Freunde könnten sich auch für das Thema interessieren.

LESE NÜTZLICHE ERFAHRUNGSBASIERTE LEITFÄDEN UND PROBIERE DIE EMPFOHLENEN INTERVENTIONSMETHODEN AUS: Zu solchen Publikationen mit wertvollen Vorschlägen gehören: „'Geh sterben!' Umgang mit Hate Speech und Kommentaren im Internet" der Amadeu Antonio Stiftung (Baldauf, Banaszczuk, Koreng, Schramm, & Stefanowitsch, 2015b), „Considerations for successful counterspeech" von Benesch, Ruths, Dillon, Saleem und Wright (2016), "Digitale Antidiskriminierungsarbeit" von Rafael, Dinar and Heyken (2017) und "Hass und Hetze im Internet – Analyse und Intervention", auch von Dinar und Rafael. Es gibt außerdem ein informatives und verständliches Buch, "Hass im Netz: Was wir gegen Hetze, Mobbing und Lügen tun können" von Ingrid Brodnig (2016).

ZUSAMMENFASSUNG

Dieser Artikel stellt den Versuch dar, einige aktuelle wissenschaftliche Literatur über die psychologische Dynamik von Hate Speech und Counter Speech im Internet in kurzer Form zusammenzufassen. Während sich viele Internetnutzerinnen und Nutzer spontan in verschiedenen Formen von Counter-Speech-Aktivität engagieren, insbesondere indem hasserfüllte Beiträge kommentiert, aber auch indem Memes erstellt, oder sogar Anti-Hate-Speech-Communities unterstützt werden; während verschiedene Projekte gegen Hass im Netz durchgeführt werden, die Privatpersonen nützliche Werkzeuge, Videos, Artikel und andere Hilfsquellen zur Verfügung stellen: Bisher scheinen nur wenige hilfreiche, wissenschaftliche Bewertungen von Counter-Speech-Strategien veröffentlicht worden sein. Die meisten verfügbaren Quellen und Leitfäden sind wertvolle erfahrungsbasierte Beobachtungen von Aktivistinnen und Aktivisten, die an den genannten Projekten gegen Hass im Netz beteiligt sind.

Die Forschung hat gezeigt, dass Online-Hass ein viel größeres virales Verbreitungspotenzial hat als Freude oder Sympathie. Eine weitere Forschungsgruppe hat den möglichen Mechanismus, welcher der viralen Verbreitung von Hass und Vorurteilen zugrunde liegt, aufgezeigt: Angst und ein psychologisches Bedürfnis nach Sicherheit. Andere Forscher verdeutlichen den immunisierenden Einfluss von Vertrautheit in Bezug auf bestimmte angstbehaftete Themen. Sind wir mit einem Thema bereits vertraut, kann uns dies davon abhalten, eine Bedrohung wahrzunehmen, auch wenn soziale Beweise (zum Beispiel inzivile Kommentare) uns in eine andere Richtung leiten. Angst motiviert dazu, hasserfüllte Inhalte zu teilen und vor einer Bedrohung zu warnen. Dabei werden Hass und Spott als Waffen begriffen, gegen das, wovor wir Angst haben und was uns unbekannt ist.

Die Nichtsichtbarkeit im Online-Diskurs und die "Gesichtslosigkeit" des Gegenübers, kombiniert mit der Asynchronität von Internet-Diskussionen und dem Ausführen einer Handlung, die als folgenlos wahrgenommenen wird, befördern den Effekt einer Online-Enthemmung. Bestimmte Formen von Counter Speech können einen positiven Effekt auf Online-Verhaltensweisen haben, besonders dann, wenn der- oder diejenige einen wahrgenommenen

Gefördert vom Bundesministerium für Familie, Senioren, Frauen und Jugend im Rahmen des Bundesprogramms Demokratie leben!

hohen gesellschaftlichen Status nach außen trägt oder wenn es sich um eine privilegierte weiße und männliche Person handelt.

Am beliebtesten ist die satirische und visuelle Form der Gegenrede, dennoch bleibt es dabei unwahrscheinlich, dass das Mediennutzerinnen und Nutzer außerhalb der Filterblase erreicht werden. Filterblasen und Echokammern sind täuschend, da diese den Eindruck erwecken, die meisten Menschen teilten unsere Werte, Überzeugungen und Ängste.

Vorurteile und Hass im Netz sind eine Bedrohung. Sie verwandeln sich in konkrete Gewalt und gefährden die körperliche Sicherheit und das physische Wohlbefinden der Opfer. Die Angst um die eigene Sicherheit ist einer der Faktoren, der die Opfer zum Schweigen bringt und den aktiven Widerstand unterdrückt. Das Ergebnis ist, dass nur einige Menschen ausreichend privilegiert sind, sich Hass im Netz entgegen zu stellen. Andererseits verringert der privilegierte Status, und damit frei von Unterdrückung und von deren Folgen zu sein, die Motivation, Hass aktiv zu bekämpfen.

Gegenreaktionen auf Hate Speech zeichnen sich durch Anstrengung aus und erfordern Reaktionsschnelligkeit, während das Ausdrücken von Hass impulsiv ist, oftmals spontan und mit wenig Mühe erfolgt. All diese Faktoren drücken eine ausgeprägte Asymmetrie aus, die dazu beiträgt, Hate Speech in sozialen Medien weitgehend zu verbreiten und Gegenrede relativ unsichtbar zu machen.

Der unglaubliche Erfolg der #metoo-Bewegung beweist jedoch, dass eine kollektive Crowd-initiierte Gegenaktion in sozialen Medien virale Verbreitung findet, die weit über die ursprüngliche Filterblase hinausgeht und zu einem konkreten sozialen Wandel auf globaler Ebene führt.

LITERATURVERZEICHNIS

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The “Nasty Effect:” Online Incivility and Risk Perceptions of Emerging Technologies. *Journal of Computer-Mediated Communication*, *19*(3), 373–387. https://doi.org/10.1111/jcc4.12009

Aronson, E., Wilson, T. D., & Akert, R. M. (2010). *Social psychology*. Upper Saddle River, NJ : Prentice Hall. Abgerufen von http://archive.org/details/Social_Psychology_7th_edition_by_Elliot_Aronson_Timothy_D._Wilson_R_M._Akert

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs*, *70*(9).

Asch, S. E., & Guetzkow, H. (1951). Effects of group pressure upon the modification and distortion of judgement. In *Groups, leadership and men* (S. 177–190). Pittsburgh, PA: Carnegie Press.

Baldauf, J., Banaszczuk, Y., Koreng, A., Schramm, J., & Stefanowitsch, A. (2015a). Die direkte Bedrohung durch Hate Speech darf nicht unterschätzt werden! Interview mit Dorothee Scholz, Diplompsychologin. In J. Schramm & A. Lanzke (Hrsg.), *„Geh sterben!“ Umgang mit Hate Speech und Kommentaren im Internet* (S. 25–29). Berlin: Amadeu Antonio Stiftung. Abgerufen von http://www.amadeu-antonio-stiftung.de/w/files/pdfs/hatespeech.pdf

Baldauf, J., Banaszczuk, Y., Koreng, A., Schramm, J., & Stefanowitsch, A. (2015b). *„Geh sterben!“ Umgang mit Hate Speech und Kommentaren im Internet*. (J. Schramm & A. Lanzke, Hrsg.). Berlin: Amadeu Antonio Stiftung. Abgerufen von http://www.amadeu-antonio-stiftung.de/w/files/pdfs/hatespeech.pdf

Baron, R. S., Vandello, J. A., & Brunsman, B. (1996). The forgotten variable in conformity research: Impact of task importance on

social influence. *Journal of Personality and Social Psychology*, *71*(5), 915–927. https://doi.org/10.1037/0022-3514.71.5.915

Bartlett, J., & Krasodomski-Jones, A. (2015). *Counter-speech: Examining content that challenges extremism online* (S. 21). Demos. Abgerufen von https://www.demos.co.uk/wp-content/uploads/2015/10/Counter-speech.pdf

Benesch, S., Ruths, D., Dillon, K. P., & Saleem, H. M. (2016). Considerations for Successful Counterspeech, 1–9.

Brodnig, I. (2016). *Hass im Netz: Was wir gegen Hetze, Mobbing und Lügen tun können*. Wien: Brandstätter Verlag.

Chronik flüchtlingsfeindlicher Vorfälle. (o. J.). Abgerufen 12. Juni 2018, von https://www.mut-gegen-rechte-gewalt.de/service/chronik-vorfaelle

Clay, R. A. (2017). Islamophobia: Psychologists are studying the impact of anti-Muslim sentiment and exploring ways to prevent it. *Monitor on Psychology*, *48*(4), 34. Abgerufen von http://www.apa.org/monitor/2017/04/islamophobia.aspx

Coustick-Deal, R. (2017, Februar 6). What's wrong with counter speech? Abgerufen 14. Mai 2018, von https://medium.com/@ruthcoustickdeal/https-medium-com-whats-wrong-with-counter-speech-f5e972b13e5e

Europarat Ministerkomitee. Empfehlung R (97) 20 des Ministerkomitees an die Mitgliedstaaten über die „Hassrede" (1997).

Fan, R., Zhao, J., Chen, Y., & Xu, K. (2014). Anger Is More Influential than Joy: Sentiment Correlation in Weibo. *PLoS ONE*, *9*(10), e110184. https://doi.org/10.1371/journal.pone.0110184

forsa. (2017). *Hate Speech*. Landesanstalt für Medien Nordrhein-Westfalen. Abgerufen von https://www.lfm-nrw.de/fileadmin/user_upload/lfm-nrw/Service/Pressemitteilungen/Dokumente/2017/Ergebnisbericht_Hate-Speech_forsa-Mai-2017.pdf

Fyfe, S. (2017). Tracking Hate Speech Acts as Incitement to Genocide in International Criminal Law. *Leiden Journal of International Law*, *30*(2), 523–548. https://doi.org/10.1017/S0922156516000753

Gelber, K., & McNamara, L. (2016). Evidencing the harms of hate speech. *Social Identities*, *22*(3), 324–341. https://doi.org/10.1080/13504630.2015.1128810

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political Conservatism as Motivated Social Cognition. *PSYCHOLOGICAL BULLETIN*, *129*(3), 339–375. Abgerufen von https://search.ebscohost.com/login.aspx?direct=true&db=edsbl&AN=RN131422234&lang=pl&site=eds-live

Maas, H. (2015). Geleitwort. In Amadeu Antonio Stiftung (Hrsg.), *„Geh sterben!" Umgang mit Hate Speech und Kommentaren im Internet* (S. 6). Berlin: Amadeu Antonio Stiftung. Abgerufen von http://www.amadeu-antonio-stiftung.de/w/files/pdfs/hatespeech.pdf

Maravilla, C. S. (2008). Hate Speech as a War Crime: Public and Direct Incitement to Genocide in International Law. *Tulane Journal of International & Comparative Law*, *17*(1), 113–144. Abgerufen von https://search.ebscohost.com/login.aspx?direct=true&db=lgs&AN=502065087&lang=pl&site=eds-live

Meyer, I. H. (1995). Minority stress and mental health in gay men. *Journal of Health &*

Social Behavior, *36*(1), 38–56.

Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: Conceptual issues and research evidence. *Psychological Bulletin*, *129*(5), 674–697. https://doi.org/10.1037/0033-2909.129.5.674

Mullen, B., & Smyth, J. M. (2004). Immigrant suicide rates as a function of ethnophaulisms: hate speech predicts death. *Psychosomatic Medicine*, *66*(3), 343–348.

Müller, K., & Schwarz, C. (2018). *Fanning the Flames of Hate: Social Media and Hate Crime* (SSRN Scholarly Paper No. ID 3082972). Rochester, NY: Social Science Research Network. Abgerufen von https://papers.ssrn.com/abstract=3082972

Munger, K. (2017). Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment. *Political Behavior*, *39*(3), 629–649. https://doi.org/10.1007/s11109-016-9373-5

Nail, P. R., & McGregor, I. (2009). Conservative Shift among Liberals and Conservatives Following 9/11/01. *Social Justice Research*, *22*(2–3), 231–240. https://doi.org/10.1007/s11211-009-0098-z

Nail, P. R., McGregor, I., Drinkwater, A. E., Steele, G. M., & Thompson, A. W. (2009). Threat causes liberals to think like conservatives. *Journal of Experimental Social Psychology*, *45*, 901–907. Abgerufen von https://www.academia.edu/23822628/Threat_causes_liberals_to_think_like_conservatives

Napier, J. L., Huang, J., Vonasch, A. J., & Bargh, J. A. (2017). Superheroes for change: Physical safety promotes socially (but not economically) progressive attitudes among conservatives. *European Journal of Social Psychology*, *48*(2), 187–195. https://doi.org/10.1002/ejsp.2315

Rafael, S., Dinar, C., & Heyken, C. (2017). Digitale Antidiskriminierungsarbeit. *Wissen schafft Demokratie, Vol 1, Iss 2, Pp 160-171 (2017)*, (2), 160. https://doi.org/10.19222/201702/15

Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology (Columbia University)*, *187*, 60–60.

Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior*, *44*(2), 136–146. https://doi.org/10.1002/ab.21737

Sue, D. W. (2010). *Microaggressions in Everyday Life: Race, Gender, and Sexual Orientation*. John Wiley & Sons.

Terry, C., & Cain, J. (2016). The Emerging Issue of Digital Empathy. *American Journal Of Pharmaceutical Education*, *80*(4), 58–58. https://doi.org/10.5688/ajpe80458

United Nations Regional Information Centre for Western Europe. (2016). Intolerance and xenophobia on the rise in Europe. Abgerufen 12. Juni 2018, von https://www.unric.org/en/latest-un-buzz/30377-intolerance-and-xenophobia-on-the-rise-in-europe