



Atzmüller, Christiane

Gemeinsam gegen Online-Hass. Aktionsforschung mit Jugendlichen zur Praktikierbarkeit gemeinsamer Gegenreaktionen auf negative Inhalte im Internet

SIAK-Journal – Zeitschrift für Polizeiwissenschaft und polizeiliche Praxis (3/2023), 19-34.

doi: 10.7396/2023_3_B

Um auf diesen Artikel als Quelle zu verweisen, verwenden Sie bitte folgende Angaben:

Atzmüller, Christiane (2023). Gemeinsam gegen Online-Hass. Aktionsforschung mit Jugendlichen zur Praktikierbarkeit gemeinsamer Gegenreaktionen auf negative Inhalte im Internet, SIAK-Journal – Zeitschrift für Polizeiwissenschaft und polizeiliche Praxis (3), 19-34, Online: http://dx.doi.org/10.7396/2023_3_B.

© Bundesministerium für Inneres – Sicherheitsakademie / Verlag Österreich, 2023

Hinweis: Die gedruckte Ausgabe des Artikels ist in der Print-Version des SIAK-Journals im Verlag Österreich (<https://www.verlagoesterreich.at/>) erschienen.

Online publiziert: 11/2023

Gemeinsam gegen Online-Hass

Aktionsforschung mit Jugendlichen zur Praktikierbarkeit gemeinsamer Gegenreaktionen auf negative Inhalte im Internet

Jugendliche werden in Sozialen Medien häufig Zeuginnen und Zeugen von Online-Hass, Cybermobbing und anderen Formen digitaler Gewalt. Maßnahmen wie gezieltes Löschen negativer Inhalte durch Plattformbetreiberinnen und -betreiber wie TikTok, Instagram oder YouTube haben sich bisher als wenig effektiv erwiesen, sodass jugendliche Internet-Nutzerinnen und -Nutzer zunehmend gefordert sind, mit zivilcouragierten Gegenreaktionen (Counter Speech) dagegen aufzutreten. Allerdings scheitert dies häufig nicht nur an Unsicherheiten in der Wahrnehmung und Bewertung von beobachteten Online-Übergriffen und an der Angst, selbst zum Opfer zu werden, sondern auch an mangelnden Kompetenzen, ein solches Auftreten erfolgreich zu gestalten bzw. auch an der antizipierten Wirkungslosigkeit von Online-Interventionen – insbesondere da die Vorstellung herrscht, allein gegen eine Masse an negativen Inhalten vorgehen zu müssen. Dabei würde sich gerade im Internet durch das Vorbild einer Vielzahl an sozialen Bewegungen, wie etwa #fridaysforfuture oder #blacklivesmatter, ein Zusammenschluss gleichgesinnter Counter-Speakerinnen und Counter-Speaker anbieten, um ein gemeinsames Vorgehen gegen Hate Speech zu gestalten. Der vorliegende Beitrag untersucht mittels partizipativer Aktionsforschung mit Jugendlichen die Praktikierbarkeit eines gemeinsamen Vorgehens Jugendlicher gegen negative Inhalte im Internet. Dabei werden nicht Kampagnen in den Mittelpunkt gestellt, sondern der Zusammenschluss und das gemeinsame Vorgehen im konkreten Anlassfall. Die Ergebnisse verweisen auf zahlreiche Hürden: Neben mangelnden Gestaltungskompetenzen sowie fehlenden Erfolgsmomenten sind es auch technische bzw. algorithmusgesteuerte Probleme, die ein gemeinsames Vorgehen für Jugendliche erschweren.

1. EINFÜHRUNG

Jugendliche sind im Internet mit vielfältigen Formen von Online-Hass und Cybermobbing konfrontiert. In einer 2018 von Atzmüller u.a. (vgl. Atzmüller et al. 2019a) durchgeführten, repräsentativen Befragung unter 1.868 14–19-jährigen Schülerinnen und Schülern in Wien gaben 95% der Befragten an, schon einmal Online-Übergriffe gesehen zu haben (z.B.

negative Kommentare/Nachrichten, unangebrachte Fotos/Videos, unautorisierte Veröffentlichungen), knapp 63% waren selbst schon einmal davon betroffen. In der repräsentativen, deutschen JIM-Studie zum Medienumgang von 12- bis 19-Jährigen (vgl. MPFS 2022a) gaben 73% der 1.200 Befragten an, im letzten Monat mit beleidigenden Kommentaren, Hassbotschaften, extremen politischen Ansichten,



CHRISTIANE ATZMÜLLER,
*Sozialwissenschaftlerin am Institut
für Soziologie an der Universität
Wien.*

Verschwörungstheorien, beleidigenden Kommentaren oder Fake News konfrontiert gewesen zu sein; unter den 17- bis 19-Jährigen sind es 87%.

Im Zuge der aktuellen JIMplus-Studie wurde auch ein Fokus darauf gelegt, wie und wo Jugendliche Online-Hass begegnen (vgl. MPFS 2022b). Von den 1.060 befragten deutschen Jugendlichen im Alter von 12 bis 19 Jahren gaben 90% an, negative Inhalte vorwiegend beiläufig bzw. im Zuge ihrer täglichen Internetnutzung in den Kommentarspalten auf den für Jugendliche wichtigsten Apps TikTok, Instagram, YouTube und WhatsApp wahrzunehmen, eine gezielte Suche danach findet in der Regel nicht statt.

Diese Zahlen verdeutlichen, dass sich bisherige Maßnahmen der Plattform-Betreibenden, die durch gezieltes Löschen sowie Blockieren oder Sperren entsprechender User-Accounts versuchen, dem hohen Anteil an negativen Inhalten entgegenzuwirken, als wenig effektiv erweisen. Auch die Forschung zeigt, dass die Identifizierung von Hass im Netz mittels automatisierter Such-Algorithmien aufgrund der Vielfalt möglicher Ausdrucksformen nur mangelhaft gelingt (vgl. Gröndahl et al. 2018), menschliche Content-Moderatorinnen und Content-Moderatoren aus Überforderung oft wenig schlüssige Entscheidungen treffen (vgl. Cobbe 2021) und zudem Verbotsstrategien tendenziell die Verlagerung negativer Inhalte auf unzensurierte Plattformen fördern (vgl. Chandrasekharan et al. 2017).

Als mögliche Alternative bieten sich aktive Gegenreaktionen durch Userinnen und User selbst an, die sich zivilcouragiert den von ihnen wahrgenommenen negativen Inhalten entgegenstellen. Eine solche aktive und öffentlich sichtbare Form des Counter Speech würde ein klares Zeichen setzen, dass Online-Übergriffe nicht toleriert werden. Das betrifft Online-Hass,

Cybermobbing und jede andere Form digitaler Gewalt, deren Intention es ist, andere Personen zu beleidigen, beschimpfen, demütigen, verletzen, bedrohen oder anzufinden. Gerade Jugendliche beobachten Online-Übergriffe vor allem im eigenen Peer-Umfeld. Umso mehr haben Jugendliche gerade hier besonders hohes Potenzial, Konfliktverläufe entscheidend zu beeinflussen (vgl. Weber/Dickter 2015). Dennoch greifen Jugendliche online überwiegend nicht ein (vgl. Atzmüller et al. 2019b; DeSmet et al. 2016; Dillon/Bushman 2015; Olenik-Shemesh et al. 2017; Shultz et al. 2014; Van Cleemput et al. 2014), vor allem dann nicht, wenn Betroffene nicht unmittelbar zum Bekanntenkreis zählen (vgl. MPFS 2022b).

Die bisherige Forschung nennt vielfältige Gründe, warum Jugendliche online sogar noch weniger intervenieren als offline (vgl. Patterson et al. 2016). Dazu zählen Unsicherheiten und fehlende Kontextinformationen für die Bewertung der Ernsthaftigkeit und Schwere von Online-Übergriffen (vgl. Herry et al. 2021), mangelnde Empathie und fehlender Bezug zu Online-Opfern (vgl. Butler et al. 2021), mangelnde Bereitschaft, die Verantwortung für Online-Interventionen zu übernehmen (vgl. Machackova et al. 2018), hohes eigenes Interventionsrisiko (vgl. Dillon/Bushman 2015) sowie mangelnde Selbstwirksamkeit in der Umsetzung (vgl. DeSmet et al. 2016).

Jüngere Erkenntnisse (vgl. Atzmüller et al. 2019b) verweisen im Hinblick auf Online-Zivilcourage allerdings auf drei besonders zentrale Probleme: (1) Jugendliche beurteilen gängige Handlungsoptionen in Online-Umgebungen wie Blockieren, Melden oder Kommentieren schlicht als wirkungslos, (2) es fehlt an Kompetenzen, zivilcouragierte Gegenrede konkret zu gestalten (vgl. Atzmüller/Zartler 2022) und (3) das Gefühl der Wirkungs- und Macht-

losigkeit geht mit der Vorstellung einher, dass sowohl Opfer als auch Beobachterinnen und Beobachter solcher Übergriffe weitgehend alleine damit zurechtkommen müssen. Die Möglichkeit, andere Internet-Nutzerinnen und Internet-Nutzer zur Unterstützung zu mobilisieren, wird von Jugendlichen nicht mitgedacht, was gerade im Hinblick auf die Vielzahl an sozialen Bewegungen und anderen Formen des Online-Aktivismus im Internet überraschend ist. Dabei hätte gerade ein gemeinsames zivilcouragiertes Auftreten in Online-Kontexten hohes Potenzial, die antizipierte Wirkungslosigkeit von Online-Interventionen zu entkräften und ein Gegengewicht zu negativen Inhalten zu schaffen. Die wenigen bisherigen Studien konnten zeigen, dass die zahlenmäßige Überlegenheit an Counter-Speakerinnen und Counter-Speakern die Wirksamkeit beeinflussen kann (vgl. Wright et al. 2017) und dass prosoziales Verhalten auch im Internet zu Nachahmung anregt (vgl. Seering et al. 2017).

Bisherige Initiativen, die die Idee eines Zusammenschlusses gegen Online-Hass forcieren, finden sich eher auf Kampagnen-Ebene, wie z.B. die #NoHateSpeech-Bewegung¹. In der alltäglichen Internetnutzung sind Jugendliche aber in der Regel mit unvorhergesehenen Situationen konfrontiert, die rasche Entscheidungen erfordern und die Mobilisierung von Unterstützung im konkreten Anlassfall verlangen. Gerade unter Jugendlichen ist der Einsatz von spontanem Antwortverhalten wahrscheinlicher als die Initiierung von Counter-Speech-Kampagnen.

Allerdings fehlt es bislang an empirischen Erkenntnissen darüber, inwieweit ein gemeinsames zivilcouragiertes Auftreten im konkreten Anlassfall durch Jugendliche überhaupt praktikierbar ist – besonders dann, wenn die involvierten Personen nicht zum persönlichen Bekanntenkreis

zählen, wo auch Offline-Interventionen möglich und bevorzugt umgesetzt werden (vgl. Atzmüller et al. 2019b).

Der vorliegende Beitrag greift diese Frage auf und untersucht mittels einer partizipativen digitalen Aktionsforschung mit Jugendlichen die Umsetzbarkeit eines gemeinsamen Vorgehens. Im Mittelpunkt steht die Frage, wie es Jugendlichen gelingen kann, sich im konkreten Anlassfall gegen Hate Speech, Cybermobbing und andere Formen digitaler Gewalt zusammenzuschließen bzw. sich gegenseitig zu unterstützen.

2. DATENBASIS UND METHODIK

Die nachfolgenden Ergebnisse stammen aus einer partizipativen Online-Aktionsforschung, durchgeführt von der Universität Wien² in Kooperation mit vier Partnerschulen aus Wien im Schuljahr 2020/21, also während der Corona-Pandemie. Aktionsforschung hat sich insbesondere in der Forschung mit Jugendlichen sehr bewährt (vgl. Wöhrer et al. 2017) und eignet sich für die inhaltliche Auseinandersetzung mit sozialen Problemen und daraus abgeleiteten Interventionen in der Praxis, indem Personen der betroffenen Zielgruppe aktiv am Forschungsprozess mitbeteiligt werden.

Da es u.a. Ziel der Aktionsforschung war, nach kreativen und konstruktiven Lösungen im gemeinsamen Vorgehen gegen Hate Speech zu suchen, sollten möglichst motivierte Jugendliche mit ausreichend kognitiven Fähigkeiten rekrutiert werden. Die Teilnehmerinnen und Teilnehmer stammten schließlich aus höheren Schulen mit unterschiedlicher inhaltlicher Schwerpunktsetzung: zwei Allgemeinbildende Gymnasien (eines aus einem bürgerlichen Bezirk, das andere aus einem traditionellen Arbeiterbezirk), eine Höhere Technische Lehranstalt mit Informatikschwerpunkt sowie eine Höhere

Bundes-Lehr- und Versuchsanstalt mit Graphik- und Multimedia-Schwerpunkt. Die meisten der teilnehmenden 14- bis 18-jährigen Jugendlichen waren an ihren Schulen als Peer-Mediatorinnen und Peer-Mediatoren tätig und daher für die Thematik bereits sensibilisiert. In der Regel sind jugendliche Peer-Mediationsgruppen in Schulen überwiegend weiblich besetzt, was sich auch in den teilnehmenden Aktionsforschungsgruppen widerspiegelte (23 Teilnehmende; 18 weiblich, fünf männlich). Einige Teilnehmende berichteten davon, selbst schon einmal von Hate Speech betroffen gewesen zu sein. Pro Schule arbeitete das Forschungsteam mit einer Gruppe von fünf bis zehn Schülerinnen und Schülern zusammen, und es gab mehrmalige rund zweistündige Aktionsforschungs-Workshops mit unterschiedlichen Zielsetzungen. Der vorliegende Beitrag greift jene Aktionsforschungstreffen auf, in denen die Praktikierbarkeit von gegenseitiger Unterstützung im konkreten Anlassfall erprobt wurde.

Bezeichnend für partizipative Aktionsforschung sind Phasen der Forschung und Aktion, die bedingt durch die Coronapandemie rein digital (mittels MS Teams und Zoom) umgesetzt wurden. Das erste gemeinsame Vorgehen gegen reale Hate-Kommentare wurde in Begleitung durch das Forschungsteam umgesetzt, indem die einzelnen Interventionen auf den jeweiligen Plattformen (z.B. auf TikTok) direkt und „live“ am gemeinsam geteilten Bildschirm mitverfolgt wurden. Gleichzeitig wurden auch sprach- und textbasierte Diskussionsmöglichkeiten (z.B. Posten von Kommentaren im Chat, mündliche Erläuterungen) genutzt, um gemeinsam mit den Jugendlichen ein strategisches Vorgehen zu entwickeln. Um einen geschützten Rahmen für die Jugendlichen zu schaffen, wurden vom Forschungsteam vorab vorbereitete, gesicherte Accounts und Profile

für die Jugendlichen zur Verfügung gestellt, damit diese nicht mit ihren privaten Accounts agieren mussten und so in anonymisierter Form auftreten konnten. Diese gesicherten Accounts und Profile wurden ausschließlich für das Forschungsprojekt verwendet. Gleichzeitig wurden die Jugendlichen vorab eingehend informiert sowie Möglichkeiten geboten, etwaige negative Erfahrungen zu besprechen und zu reflektieren. Das gesamte Prozedere wurde vorab von der Ethikkommission der Universität Wien geprüft, um negative Auswirkungen auf die Teilnehmenden zu vermeiden.

Alle Aktionsforschungs-Workshops wurden aufgezeichnet und transkribiert. Die Transkripte wurden mit Screenshots der entstandenen Chat-Protokolle bereichert. Die Datenauswertung erfolgte inhaltsanalytisch mittels induktiver Kategorienbildung (vgl. Mayring 2010).

3. ERGEBNISSE

Aufgabe der Jugendlichen war es, mit den zur Verfügung gestellten Accounts und Profilen als Gruppe gemeinsam gegen echte Hate-Kommentare vorzugehen. Damit diese Aufgabe umgesetzt werden konnte, musste zunächst nach Hate Speech gesucht werden, auf die reagiert werden könnte.

3.1 Hate Speech finden

Die Aktionsforschungsteilnehmenden wurden dazu vorab gebeten, nach passenden Beiträgen bzw. Hate-Kommentaren z.B. auf Instagram oder TikTok zu suchen, um diese für ein gemeinsames Vorgehen vorzuschlagen.

Dabei zeigte sich, dass allein schon die Auswahl an passenden Beiträgen schwieriger war als erwartet: Jene Teilnehmenden, die im Alltag seltener mit Hate Speech in Berührung kamen, meinten, dass sie Kommentare unter Videos oft grundsätz-

lich nicht lesen würden und folglich auch nicht unbedingt auf Hate Speech stießen. Zudem würde ihnen in der eigenen „Bubble“ kein entsprechender Content in dieser Richtung vorgeschlagen werden, da diese vorwiegend mit positiven Beiträgen aufgeladen sei („auf meiner TikTok Seite gibt’s nicht so wirklich viele Hate-Kommentare, nicht so richtige“ G2³,S2). Auch im weiteren Verlauf der Aktionsforschung verdeutlichte sich, dass der Algorithmus beim Auffinden von negativen Inhalten eine wesentliche Rolle spielt: Mit der Suche nach negativen Inhalten wurde auch der Algorithmus zunehmend in diese Richtung „trainiert“.

Aber auch gezieltes Suchen war herausfordernd: So ist es auf TikTok prinzipiell nicht möglich, nach (negativen) Kommentaren zu suchen, sondern nur nach „Captions“, d.h. Videobeschreibungen, die in der Regel kaum Rückschlüsse über mögliche negative Inhalte zulassen. Zusätzlich sind Filter aktiv, die bestimmte Worte herausfiltern oder deren Sichtbarkeit eingrenzen. Wurde gezielt nach Videoinhalten gesucht, wo anzunehmen war, dass sich hier Hate-Kommentare finden könnten, wie z.B. beim Thema „Body Positivity“, dann waren die Treffer eher „Reaction“-Videos, das sind Videos meist von betroffenen Opfern, die eine öffentliche Reaktion auf die an sie gerichteten Übergriffe zeigen. Allerdings regte dies die Teilnehmenden nicht dazu an, noch weiter auf das gleiche Video zu reagieren, auch wenn zum Teil unter diesen Reaktionsvideos noch weitere Hate-Kommentare zu finden waren.

Ein weiteres Problem war, dass manche vorgeschlagenen Beiträge schon relativ alt waren, da Plattformen Videos nach Beliebtheit und nicht nach Erscheinungsdatum vorschlagen, sodass Interaktionen mit realen Userinnen und Usern eher unwahrscheinlich waren und daher die Aktionsforschungsteilnehmenden weitere Inter-

ventionen als wenig sinnvoll einschätzten. Gezielt nach ganz neu erschienenen Videos zu suchen, wo die Aufmerksamkeit für Kommentare vergleichsweise am höchsten ist, würde die beteiligten Jugendlichen eher unter Zeitdruck versetzen: „Man muss halt ur schnell sein und Videos erwischen, die vor noch nicht langer Zeit gepostet wurden. Sonst muss man da irrsinnig lange suchen“ (G1,S4).

Andere Teilnehmende, die im Alltag bzw. in ihrer „Bubble“ öfter Hate-Kommentaren begegneten, konnten sich nicht entscheiden, ob die vorgefundenen negativen Kommentare tatsächlich als Hate-Kommentare zu werten waren oder nicht. Diskutiert wurde, ob das nicht ein einfaches Sich-lustig-machen wäre, mitunter auch „verdienterweise“, da der gezeigte Content der betroffenen Opfer dies zum Teil auch provozieren würde („das muss man auch irgendwie ein bisschen erwarten, wenn man so ein Video postet“ G1,S8). In der Wahrnehmung der Aktionsforschungsteilnehmenden würden solche Personen „jetzt nicht wirklich gehatet, sondern nur ein bisschen verarscht“ (G2,S4) werden. Nach Ansicht der Aktionsforschungsteilnehmenden würde man in solchen Fällen eher darauf hinweisen, dass jede Person machen kann, was sie oder er will, aber nicht spezifisch auf Kommentare eingehen – ansonsten würde man auch Gefahr laufen, als Spielverderbende und -verderbender dazustehen.

Negative Kommentare seien zudem nicht immer klar als Hate-Kommentare erkennbar: Oft seien diese auch Reaktionen auf den negativen Content von Creatorinnen bzw. Creatoren und daher eigentlich Gegenreaktionen, die aber als Gegenattacken negativ aufgeladen seien. Auch betroffene Opfer selbst würden teils mit aggressiven Antworten reagieren (z.B. „gusch“), sodass nicht klar sei, ob hier einzuschreiten sei oder nicht.

Zusätzlich seien mittlerweile auch viele „Kommentarblocker“ (Filter) aktiv, wodurch bestimmte negative Ausdrücke nicht mehr gepostet werden könnten und so Hate Speech nicht klar als solche erkannt werden könne.

Insgesamt stellte dies die Aktionsforschende vor die Aufgabe, bei jedem negativ wirkenden Kommentar einschätzen zu müssen, inwieweit hier tatsächlich Hate Speech vorliegt oder nicht. Dies war nicht nur teilweise überfordernd, sondern hatte auch Konsequenzen für das eigene Auftreten: das eigene zivilcouragierte Einschreiten sollte nicht dazu führen, letztlich als „Moralapostel“ wahrgenommen zu werden, sodass oft nur sehr zögerlich agiert wurde:

„Also bei gewissen Kommentaren habe ich ihn zwar schon geschrieben, aber nicht abgeschickt, weil ich mir gedacht habe: ach, komm. Wegen sowas muss man halt auch kein Fass aufmachen“ (G1,S3).

Auffallend war auch, dass bereits vorhandene Gegenreaktionen anderer Userinnen und User eher davon abhielten, ebenfalls zu reagieren: „Wo andere Leute auch kommentieren und was gegen diesen Kommentar machen, müsste man jetzt nicht so viel vorgehen. Meiner Meinung nach weil die anderen das schon übernehmen“ (G1,S20). Vorhandene Counter-Speech-Kommentare zusätzlich zu unterstützen, wurde eher nicht mitgedacht.

Insgesamt zeigte sich bereits in der ersten Phase der Aktionsforschung, dass gezieltes, organisiertes gemeinsames Vorgehen gegen Hass-Kommentare schon allein beim Auffinden von Kommentaren auf teils für Jugendliche unerwartet hohe Hürden stößt, wenn sie nicht unmittelbar mit einem klaren Anlassfall konfrontiert sind.

3.2 Strategien des gemeinsamen Vorgehens und Erfolgsbewertung

Für das gemeinsame Vorgehen gegen Hate Speech wurden schließlich in jeder Grup-

pe unterschiedliche TikTok-Videos ausgewählt, bei denen in den Kommentaren Hate Speech zu finden war. Je nach inhaltlicher Gestaltung und Ausrichtung der Kommentare wurden von den Aktionsforschungsteilnehmenden gemeinsam mit dem Forschungsteam unterschiedliche Strategien für ein gemeinsames Vorgehen während der Aktionsforschung entwickelt. Nach dem Aktionsforschungs-Workshop war es Aufgabe der Jugendlichen, noch ca. eine Woche lang mit Hilfe der zur Verfügung gestellten Fake-Accounts weiterhin gegen Hate-Kommentare vorzugehen und dabei auch zu versuchen, sich gegenseitig zu unterstützen bzw. ein gemeinsames Vorgehen zu praktizieren. Diese Erfahrungen werden in den nachfolgenden Darstellungen ebenfalls integriert.

3.2.1 Gemeinsames Verfassen möglichst vieler Gegenkommentare

In einem der ausgewählten Videos für ein gemeinsames Vorgehen gegen Hate Speech präsentierte sich ein männlicher Jugendlicher mit nacktem Oberkörper und erhielt dafür viele negative Kommentare. Als erste Form der Unterstützung wurde durch eine Aktionsforschungsteilnehmerin bzw. einen -teilnehmer ein erster eigener Counter-Speech-Kommentar mit eher humorvoll-sarkastischer Ausrichtung gepostet. Es folgte eine kurze Interaktion mit jener realen Person, die sich negativ geäußert hatte. Als diese gebeten wurde, mit den Übergriffen aufzuhören, folgte lediglich eine Verhöhnung „Ok ich hAbE gRd sOs AnGsT Ich HöRe Sofort AuF“⁴. Die Interaktion wurde schließlich durch ein :) beendet, es fand auch keine weitere Unterstützung statt.

Auch in einem zweiten Video wurde versucht, Gegenkommentare zu posten: Ausgewählt wurde eine bekannte Influencerin, die zwar „Fehler“ gemacht hätte in bisherigen Videos, indem ihr u.a. leicht rassis-

tisches Verhalten vorgeworfen wurde und sie sich nicht entsprechend entschuldigt hätte. Im Verhältnis dazu würde sie aber übermäßig viel Hate erhalten. Die Intention der Aktionsforschungsteilnehmerinnen und -teilnehmer war es, der Community mitzuteilen, dass diese insgesamt aufhören sollte – es wurde dann auch ein entsprechender Kommentar dazu gepostet („Man kanns auch gut sein lassen“). Aber auch hier folgte wieder lediglich Verhöhnung („HAGAHAGAGAHAAAA“), weitere Interaktionen fanden nicht statt.

Die beiden Beispiele verdeutlichen, dass bei den jugendlichen Aktionsforschungsteilnehmenden die Idee einer gegenseitigen Unterstützung grundsätzlich nicht mitgedacht wird und vorsichtige Versuche, Gegenkommentare zu posten, häufig eher erfolglos verliefen.

Im weiteren Vorgehen wurden Videos ausgewählt, bei denen Influencerinnen und Influencer auf TikTok aufgrund ihres Aussehens (Styling/Kleidung/Körper) immer wieder mit Online-Hass konfrontiert wurden. Auch hier wurde versucht, mit Gegenkommentaren zu reagieren. Dieses Mal wurde aber vom Forschungsteam explizit und stärker angeregt, sich auch beim Kommentieren gegenseitig zu unterstützen (siehe Abbildung 1).

Einige der Aktionsforschungsteilnehmenden versuchten, mehrere Kommentare als Gegenreaktionen zu posten, agierten dabei aber individuell und versuchten nicht unbedingt, sich gegenseitig zu stärken. Dies wurde auch damit begründet, dass es oft zu lange dauern würde, bis die Kommentare anderer sichtbar waren bzw. gäbe es aufgrund des Algorithmus Probleme, die

Quelle: Atzmüller [Screenshots aus der Aktionsforschung]

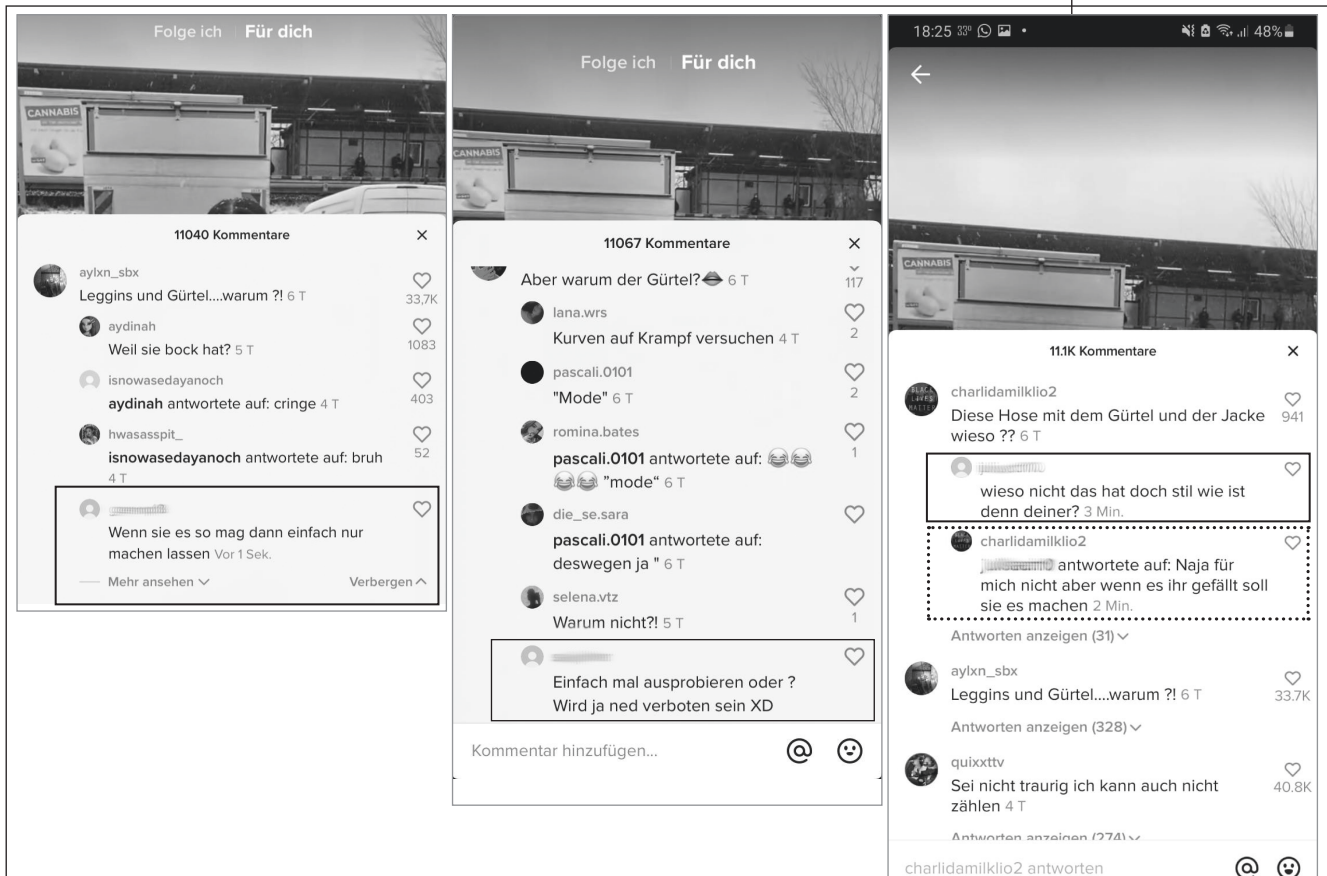


Abb. 1: Reaktionen auf Hate-Kommentare gegen den Kleidungsstil einer Influencerin

Kommentare anderer überhaupt zu finden, da sich diese in der Sichtbarkeit für einzelne Userinnen und User unterschiedlich darstellen, sodass eine gegenseitige Unterstützung im Team schwer möglich war:

„Also wie ich durchgescrollt bin, habe ich von keinem anderen hier in der Gruppe einen Kommentar lesen können“ (G2,S18). „Ja, das Ding ist halt, bis ich den Kommentar gefunden habe, ist eine Ewigkeit vergangen“ (G3,S26).

Andere Aktionsforschungsteilnehmende, die sich nicht als Unterstützung einbrachten, nannten die sich ständig wiederholenden Hate-Kommentare als Herausforderung, die wenig Spielraum für mehrfache Counter-Reaktionen offenlassen würden: „Ich wusste bei manchen nicht, was ich schreiben soll, es hat sich alles nur wiederholt“ (G3,S12).

Gleichzeitig erschöpften sich auch die Möglichkeiten für Counter Speech, da sich auch die Inhalte der Gegenreaktionen ständig wiederholen und aus der Sicht der teilnehmenden Jugendlichen keine weiteren inhaltlich bereichernden Counter Speech-Beiträge mehr möglich waren.

„Ich hab dann gesagt, ja, super. Also da kann man auch nicht viel mehr sagen als stimmt oder sie hat recht oder sowas“ (G3,S13).

Trotzdem zeigten sich vereinzelt auch Erfolgsmomente, in denen Interaktionen mit anderen, realen Userinnen und User stattfanden und diese ihre Aussagen sogar zum Teil etwas relativierten (siehe Abbildung 1, Seite 25, gepunktet umrahmter Kommentar) oder zusätzlich mit Counter Speech unterstützten.

Besser gelang den Aktionsforschungsteilnehmenden das gemeinsame Verfassen von Gegenkommentaren, wenn Online-Hass besonders offensichtlich war bzw. wenn als Gegenreaktion mit geschickter Schlagfertigkeit reagiert werden konnte. So wurde beispielsweise ein TikTok-Video

ausgewählt, wo ein Mädchen zu sehen war, das den gängigen Schönheitsidealen eher wenig entsprach und das aufgrund ihrer sprachlichen Ausdrucksfähigkeit auch intellektuell benachteiligt zu sein schien (siehe Abbildung 2, Seite 27).

Die Aktionsforschungsteilnehmenden gingen hier sehr aktiv gegen Hate-Kommentare vor und versuchten auch, sich gegenseitig zu unterstützen, wobei Unterstützung so aufgefasst wurde, dass die Counter-Speakerinnen und Counter-Speaker auf ähnliche Hate-Kommentare auf unterschiedliche Weise reagierten, was allerdings eine gewisse Schlagfertigkeit verlangte. Häufig wurde der Wortlaut des Hate-Kommentars aufgenommen und als Retourkutsche zurückgespielt.

Schlagfertigkeit bzw. auch ein gelungener humorvolle Konter würden nach Ansicht der Teilnehmenden auch die Chance für entsprechende Aufmerksamkeit erhöhen:

„Wie man bei TikTok bei ganz vielen Topkommentaren eigentlich sieht, ist das Topkommentar eigentlich immer ein Witz oder eine humoristische Antwort, wenn man glaube ich auf dem Level kommentieren würde, könnte man es schaffen, als wenn man sagen würde, nein, böse, lass das“ (G1,S35).

Insgesamt zeigte sich, dass die Strategie, gemeinsam möglichst viele Gegenkommentare zu posten, teilweise erschöpfend sein kann bzw. Jugendliche an die Grenzen ihrer schlagfertigen Reaktionsmöglichkeiten bringt. Die jugendlichen Aktionsforschungsteilnehmenden merkten auch an, dass es oft auch entscheidend sei, auf welcher inhaltlichen Ebene Hate angesiedelt sei: So wären Jugendliche beispielsweise für Themen wie „Body Positivity“ bereits sensibilisiert, wodurch sie bei dieser Thematik eher für Counter Speech zu mobilisieren seien als für andere Themen.

Gleichzeitig wurde auch argumentiert, dass viele einfach keine Lust hätten, noch

Quelle: Atzmüller [Screenshots aus der Aktionsforschung]

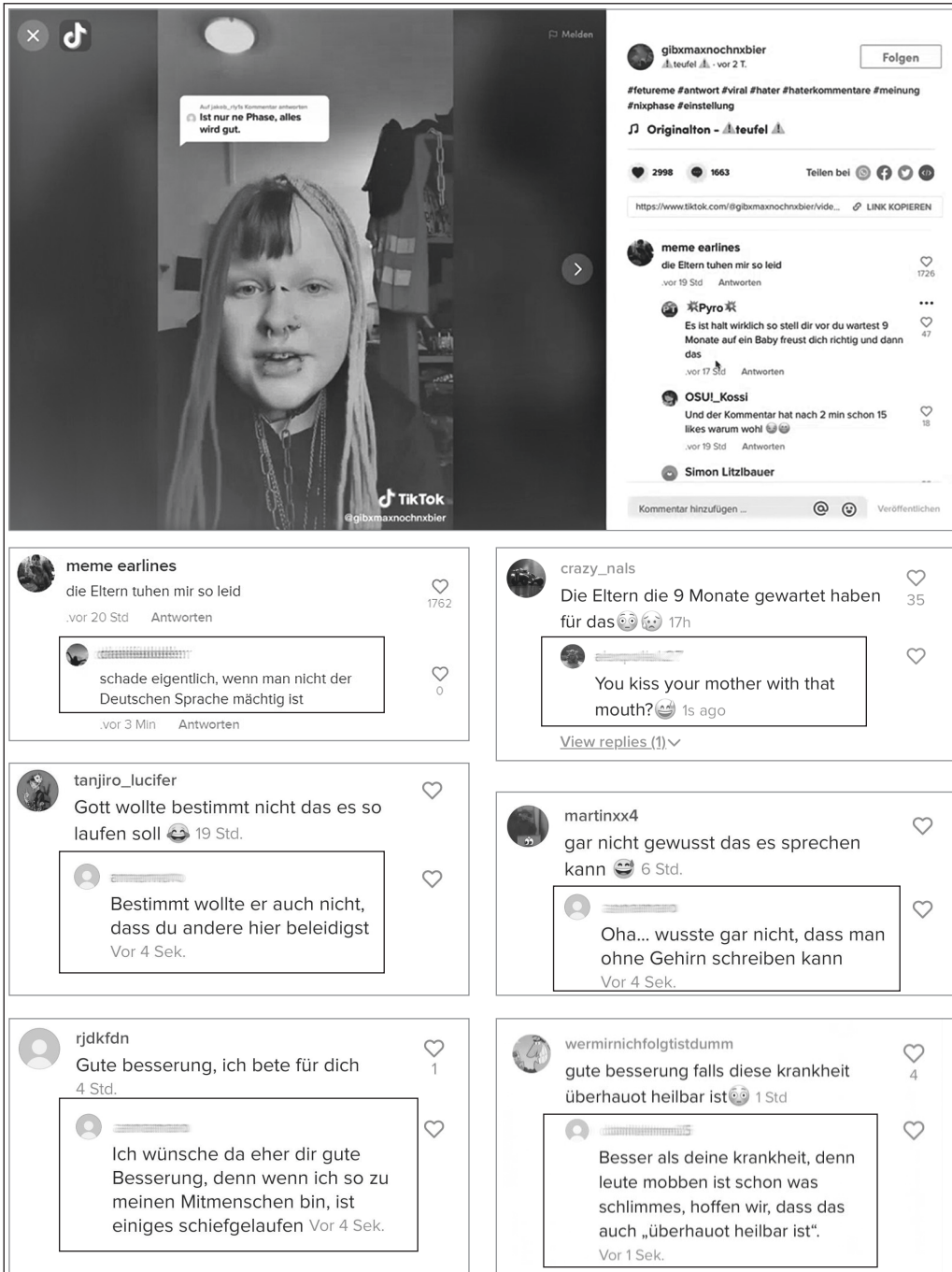


Abb. 2: Reaktionen auf Hate-Kommentare gegen das Aussehen einer Person

etwas anzumerken bzw. sehr viele Kommentare davon abhalten würden, selbst einen weiteren Beitrag zu verfassen:

„Leute [kommen oft] gar nicht auf die Idee [...], wenn viele Kommentare sind, sich die noch anzuschauen, weil sie sehen das Video und bilden sich eine eigene Mei-

nung und sind dann oft einfach schreibfaul“ (G2,S18).

3.2.2 Gegenrede „hochpushen“ durch die Vergabe von Likes

Als weitere Unterstützungs-Strategie wurde überlegt, ob es gelingen könnte, bereits

vorhandene Gegenreaktionen durch die Vergabe von Likes „hochzupushen“, so dass diese in der Liste an Kommentaren weiter oben aufscheinen und damit für andere sichtbarer werden, da viele Gegenreaktionen in der Masse an Kommentaren übersehen werden.

Aber auch hier gelangen die Versuche nur bedingt bzw. wurde der Kommentar aufgrund des Algorithmus nur bei jenen weiter oben angezeigt, die auch Likes vergaben, bei anderen aber nicht. Es wurde auch versucht, andere gefundene Counter Speech-Kommentare von realen Userinnen und Usern durch die Vergabe von Likes nach oben zu reihen, aber auch das gelang nicht bzw. war die Anordnung der Kommentare innerhalb der Gruppe ganz unterschiedlich zu sehen. Um dieses Problem zu umgehen, wurde versucht, die jeweiligen Counter Speech-Kommentare gegenseitig zu taggen, aber dennoch gelang es nicht, die Kommentare weiter nach oben zu reihen, zudem konnten nur maximal fünf Personen markiert werden.

Quelle: Atzmüller (Screenshots aus der Aktionsforschung)

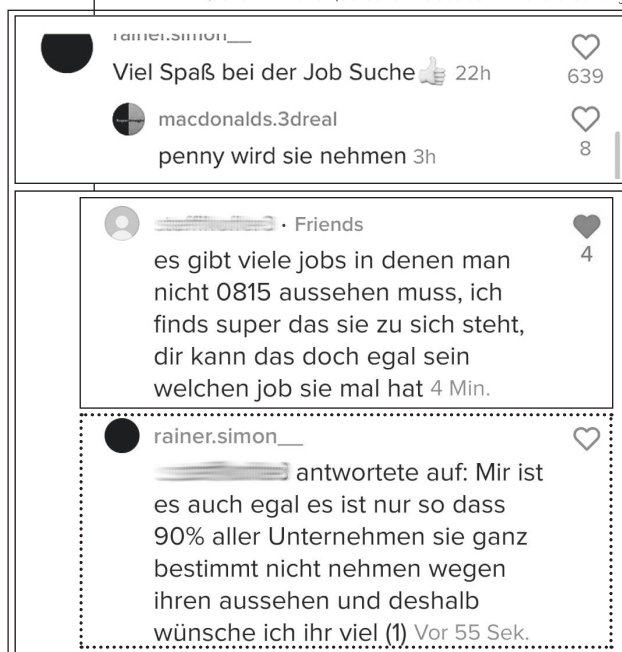


Abb. 3: Unterstützung durch Likes

Als Ursache wurde vermutet, dass die Gruppe an Counter-Speakerinnen und Counter-Speakern einfach zu klein war: „Ich glaube, wir sind jetzt da zu 12., die da dahinter sind und ich würde behaupten, dass es da weit mehr als 12 Leute braucht, [...] die Topkommentare haben über 6.000 Likes und mehr“ (G1,S19).

Trotz der Schwierigkeiten, die Kommentare anderer zu finden, gelang es nach vielen Versuchen hin und wieder, mit Hilfe von Likes Counter Speech-Kommentare anderer Aktionsforschungsteilnehmenden zu unterstützen (siehe Abbildung 3: vier Herzen für einen Counter-Kommentar, das beim Video zum Mädchen, das gängigen Schönheitsidealen wenig entsprach, gepostet wurde). Damit wurden die Kommentare zwar nicht nach oben gereiht, aber für andere sichtbar Unterstützung signalisiert. In diesem Beispiel gelang es auch, eine Interaktion mit jener Person herzustellen, die den negativen Ursprungskommentar gepostet hat (gepunktet markiert).

Schließlich wurde auch versucht, Gegengrede fremder, realer Userinnen und User durch Likes oder durch Kommentare zu unterstützen (siehe Abbildung 4, Seite 29). Bei genauerer Betrachtung ist durch die Lach-Emojis allerdings unklar, ob es sich hier eher um sarkastisch gemeinte Gegengrede handelte, was vermutlich aufgrund der Spontanität der Reaktionen durch die Aktionsforschungsteilnehmerinnen und -teilnehmer leicht übersehen werden konnte.

3.2.3 Bewertung des Erfolgs im gemeinsamen Vorgehen

Insbesondere in den Tagen nach dem Aktionsforschungs-Workshop sollten die Teilnehmenden selbstständig versuchen, Counter Speech zu betreiben und sich gegenseitig zu unterstützen. Diese Aufgabenstellung wurde von den Teilnehmenden allerdings sehr unterschiedlich

umgesetzt: Einige bemühten sich intensiv, als Counter-Speakerinnen und Counter-Speaker aufzutreten, andere weniger bzw. hatten dafür weniger Zeitressourcen zur Verfügung. In der Regel wurde diese Aufgabe alleine durchgeführt, nur vereinzelt schlossen sich Teilnehmende zusammen, um z.B. zu zweit zu überlegen, wie sie kontern könnten.

Die zur Verfügung gestellten gesicherten Accounts wurden von den Aktionsforschungsteilnehmenden sehr unterschiedlich bewertet: Manche betonten, dass sie auch ohne diese reagiert hätten, andere wiederum waren froh, damit einen geschützten Rahmen zu haben, ohne persönliche Informationen preisgeben zu müssen, insbesondere aus Angst vor Gegenreaktionen.

Nach dieser Forschungsphase stellte sich in der nachfolgenden Diskussion die Frage, inwieweit das gemeinsame Vorgehen als erfolgreich bewertet werden konnte oder nicht. Allerdings fehlten den Teilnehmenden entsprechende Rückmeldungen realer Userinnen und User, da letztlich nur in sehr wenigen Fällen tatsächlich Interaktionen stattfanden:

„Also keiner hat meine Beiträge gelikt oder irgendwas dazu geschrieben oder so und ich habe auch jeden Tag geschaut“ (G2,S4).

Keinerlei Reaktionen zu erhalten, sei prinzipiell demotivierend gewesen:

„Es war dann schon ein bisschen komisch, dass keiner – dass man das Gefühl hat, dass keiner das liest“ (G1,S5). „Da jetzt keine Reaktion darauf ist, sehe ich das, was wir gemacht haben, als fast unnötig, für uns jetzt, [...] für die Person war es glaube ich nicht unnötig, wenn da Leute auf ihrer Seite waren“ (G4,S3).

Selbst negative Reaktionen wären erwünscht gewesen, weil man hier noch antworten hätte können und damit Interaktionen insgesamt interessanter geworden

Quelle: Atzmüller [Screenshots aus der Aktionsforschung]



Abb. 4: Unterstützung anderer, realer Kommentare

wären. Es wurde auch bezweifelt, inwieweit Haterinnen und Hater damit erreicht werden konnten, da es im Internet auch einfach möglich sei, sich einer Reaktion zu entziehen und einsichtiges Verhalten eher unwahrscheinlich sei – da wäre es eben einfacher, dann gar nicht mehr zu reagieren.

„Die Person hat erkannt, dass sie falsch gelegen hat, das Ganze zugeben als Kommentar, das ist für die Person schwer, und das ist leichter, nichts zu sagen“ (G4,S3).

Dennoch wurde betont, dass es prinzipiell auch ein Erfolg für sich selbst sei, wenn man den Mut aufbringe, gegen Online-Hass vorzugehen – allerdings nur bei jenen Kommentaren, die auch eindeutig als Hate Speech klassifiziert wurden.

Einige Teilnehmerinnen und Teilnehmer vermuteten auch, dass mit „sanfter“ Counter Speech einfach zu wenig Aufmerksamkeit erzeugt werden könne:

„Ich glaube, dass du die meiste Aufmerksamkeit von den Hatern bekommst, wenn du sie zurückbeleidigst oder was schreibst, was sie triggert. Aber vielleicht sind meine Kommentare nicht die Aufmerksamkeit catchenden, aber dafür habe ich keine anderen Leute verletzt“ (G2,S16).

Dennoch gab es unter den Aktionsforschungsteilnehmenden einzelne Personen, denen es gelang, in Interaktion zu treten:

„Bei mir haben die Leute ganz unterschiedlich drauf reagiert, also manche

haben es eingesehen und geschrieben, ja, ich hätte das netter formulieren können und andere, so, ja, haben sich einfach voll da reingeritten, noch mehr. Und bei einem Kommentar hat sogar die Erstellerin von dem Video, hat den Kommentar gelikt. Das hat sich schon gut angefühlt, weil ich das Gefühl hatte, auch, wenn es gegen den Hate nichts gebracht hat, dass wenigstens sie sich besser gefühlt hat deswegen“ (G2,S1).

Interessant ist aber auch, dass einige Aktionsforschungsteilnehmende gar nicht mitverfolgten, ob es zu Interaktionen kam oder nicht: „Also ich habe jetzt nicht so

viel mich damit beschäftigt, aber es kann schon sein, dass irgendwer das vielleicht gelikt hat“ (G2,S2).

Abbildung 5 zeigt, dass eine Unterstützung von Counter Speech-Kommentaren mittels Likes auch gelingen kann: Der Counter Speech-Kommentar einer fremden, realen Person erhielt mit 646 Likes verhältnismäßig viel Unterstützung, allerdings immer noch deutlich weniger als die Hate-Kommentare weiter oben mit 1.084 bzw. 1.453 Likes.

In der nachfolgenden Diskussion betonten die Aktionsforschungsteilnehmenden, dass das Vorgehen als Gruppe durchaus als bestärkend erlebt wurde:

„Dadurch, dass ich auch wusste, wir sind in so einer Gruppe, wir decken uns ab, dass wir einfach uns auch gegenseitig unterstützen, sozusagen, dass man jemanden hat, der mit einem dagegen vorgeht, habe ich mich in der Sache stärker gefühlt, bei dem, was ich mache und dass es auch so richtig ist und auch so gehört“ (G4,S15).

Einige der Aktionsforschungsteilnehmenden, die auch abseits der Aktionsforschung immer wieder gegen Hate-Kommentare vorgehen, merkten auch an, wie wichtig ihnen positive Unterstützung anderer sei, insbesondere auch im Umgang mit möglichen negativen Reaktionen:

„Ich merke auch selber manchmal, wenn ich was kommentiere, wenn ich gegen Hate vorgehen versuche [...], dass ich dann teilweise schon richtig nervös bin [...] was sagen die Leute jetzt, machen die jetzt auch was auf dem eigenen Post oder schreiben sie Privatnachrichten, wo sie beleidigen. Und sobald nur ein Kommentar kommt, das mich unterstützt, dann bin ich null nervös mehr, dann bin ich so, okay, wurscht, weil egal, was noch für Kommentare kommen, ich bin nicht die einzige, die das so sieht“ (G4,S16).

Obwohl Unterstützung durch andere prinzipiell mit positiven Gefühlen assozii-

Quelle: Atzmüller [Screenshots aus der Aktionsforschung]



Abb. 5: Likes für Hate und Counter Speech

iert wurde, wurde gegenseitige Unterstützung aber im konkreten Vorgehen nicht bewusst mitgedacht und auch nicht unbedingt erwartet. Es wurde zum Teil eher vermutet, dass der Zusammenschluss unter Haterinnen und Hatern besser gelingen würde als unter Counter-Speakerinnen und Counter-Speakern, da es einfacher sei, schlecht über jemanden zu reden als etwas Gutes zu sagen.

Eine mögliche Form der Unterstützung kann auch kollektives Melden sein. Die Aktionsforschungsteilnehmenden meinten, dass es möglich sei, einen konkreten Aufruf z.B. unter den eigenen Abonnetinnen und Abonnenten bzw. im eigenen Bekanntenkreis zu starten mit der Aufforderung, bestimmte Inhalte zu melden, da sich mit der Anzahl an Meldungen die Chance erhöhen würde, dass negative Inhalte tatsächlich gelöscht werden.

4. CONCLUSIO

Im Rahmen einer partizipativen Online-Aktionsforschung mit 23 14–18-jährigen Schülerinnen und Schülern aus vier Partnerschulen in Wien wurde im Schuljahr 2020/21 die Praktikierbarkeit eines gemeinsamen Vorgehens gegen Online-Hass, Cybermobbing und andere Formen digitaler Gewalt erprobt. Die Darstellungen in diesem Bericht beziehen sich auf Erkenntnisse zum gemeinsamen Vorgehen im konkreten Anlassfall und bezogen auf persönlich unbekannte Opfer sowie Täterinnen und Täter bzw. auf typische Szenarien, wie sie in der alltäglichen Social-Media-Nutzung zu finden sind. Die teilnehmenden Jugendlichen waren größtenteils als Peer-Mediatorinnen und Peer-Mediatoren in ihren Schulen tätig und entsprechend motiviert und für das Thema bereits sensibilisiert.

Die Ergebnisse repräsentieren nicht tatsächliches Counter Speech-Verhalten Jugendlicher, sondern verweisen auf mög-

liche Chancen und Grenzen gemeinsamer Gegenreaktionen im konkreten Anlassfall.

Obwohl Jugendliche insgesamt häufig mit negativen Inhalten konfrontiert sind, erwies sich das konkrete Suchen nach Online-Hass, auf den gemeinsam reagiert werden soll, als schwieriger als erwartet: Die Begegnung mit Online-Hass geschieht beiläufig und unvorhergesehen, wird stark vom Algorithmus der jeweiligen Plattformen bzw. von der persönlichen Bubble gesteuert. Insofern ist es für Jugendliche schwer, bewusst und vorbereitet ein gemeinsames Auftreten zu planen. Gegenseitige Unterstützung durch mehrfaches Posten von Gegenreaktionen wird allein schon dadurch erschwert, dass Kommentare anderer Counter-Speakerinnen und Counter-Speaker aus der eigenen Gruppe in den Kommentaren nur schwer gefunden werden, da diese bei verschiedenen Userinnen und Usern oft unterschiedlich angezeigt werden. Gleichzeitig ist auch das Repertoire an möglichen Reaktionen schnell erschöpft, da sich sowohl Hate Speech- als auch Counter Speech-Inhalte oft wiederholen und wenig Spielraum für weitere Interaktionen offenlassen.

Versuche, die Sichtbarkeit von Counter-Kommentaren zu erhöhen, indem diese durch die Vergabe von Likes in der Kommentarliste nach oben gepusht werden sollen, scheitern ebenfalls einerseits am Finden der Kommentare, andererseits auch an der meist zu geringen Zahl an Likes – oft überragt die Anzahl an Likes für Hate Speech bei weitem im Vergleich zu Likes für Counter Speech. Dennoch ist gerade dieser Support eine gute Möglichkeit, sowohl Solidarität dem Opfer gegenüber zu zeigen als auch andere Counter-Speakerinnen und Counter-Speaker in ihrem Vorgehen zu bestärken.

Obwohl in einigen Fällen der eigene Einsatz als demotivierend erlebt wurde, da trotz Bemühungen häufig einfach

keine weiteren Reaktionen folgten und sich damit nicht unbedingt ein Erfolgsgefühl einstellte, gelang es zumindest in einzelnen Fällen, konkrete Interaktionen mit realen Personen herzustellen.

Insgesamt lässt sich schlussfolgern, dass es Jugendlichen vor allem an Erfahrung und Übung fehlt, wie erstens das Auftreten als Counter-Speakerin und Counter-Speaker gestaltet werden kann und wie dies zweitens in einem gemeinsamen, unterstützenden Kontext gelingen kann. Die Jugendlichen, die an der Aktionsforschung teilgenommen haben, meinten, dass schon allein das Ausprobieren, wie sich gemeinsames Kontern anfühlt, sowohl die Auseinandersetzung mit der Thematik als auch das Selbstbewusstsein in ihrer Rolle als Counter-Speakerin und Counter-Speaker enorm gestärkt hätten.

Aus den gewonnenen Erkenntnissen lassen sich zusammenfassend folgende praxisrelevante Empfehlungen ableiten, um gemeinsames zivilcouragiertes Auftreten von Jugendlichen gegen Hate Speech, Cybermobbing und andere Formen digitaler Gewalt im Internet zu fördern:

Besonders wichtig wäre es, Überzeugungsarbeit zu leisten, dass ein gemeinsames Counter Speech-Vorgehen sinnvoll und notwendig ist, z.B. um zu demonstrieren, dass die überwiegende Mehrheit gegen Online-Hass ist, als Vorbildwirkung für Unentschlossene, um Opfern ein Zeichen der Solidarität zu vermitteln und um sich für eine Online-Gesellschaft basierend auf humanistischen und prosozialen Werten einzusetzen. Dabei könnten auch einfache Formen der Unterstützung, wie die Vergabe von Likes einzelner Counter-Speakerinnen und Counter-Speaker eine

Stütze sein, um diesen zu signalisieren, dass sie nicht alleine sind.

Jugendlichen fehlt es nicht nur an praktischer Erfahrung und Übung, wie Counter Speech gestaltet werden kann, sondern auch, wie gegenseitige Unterstützung insbesondere im konkreten Anlassfall gelingen kann bzw. wird die Unterstützung anderer nicht mitgedacht. Dabei wäre auch darauf zu achten, entsprechende „Erfolgsaussichten“ abzuklären, dass also nicht unbedingt ein „Schlagabtausch“ mit Haterinnen und Hatern das Ziel sein muss, sondern die Unterstützung jener mutigen Personen, die mit Counter Speech auf negative Inhalte reagieren bzw. auch, um deren Sichtbarkeit zu erhöhen. Unterstützung kann dabei auch als Gelegenheit gesehen werden, weniger gelungene Counter Speech ins rechte Licht zu rücken oder aber auch um einfach Rückhalt zu signalisieren. In diesem Zusammenhang können auch die unterschiedlichen Möglichkeiten der Unterstützung vermittelt werden, beginnend von einfachen Likes bis hin zur Partizipation an Counter Speech-Kampagnen.

Aktives Praktizieren von Counter Speech kann das Selbstbewusstsein in der Rolle als Counter-Speakerin und Counter-Speaker enorm stärken, vor allem bei jenen, deren sprachliche Versiertheit beim Kontern weniger stark ausgeprägt ist. Aber auch die Vermittlung von praktischem Handlungswissen, wie gemeinsames Vorgehen gelingen kann, kann für Jugendliche ein Impuls sein, sich künftig stärker gegen Online-Hass zu engagieren. Gleichzeitig erfolgt dadurch eine intensive Auseinandersetzung mit der Thematik, die wiederum dazu ermutigt, aktiv als Counter-Speakerin und Counter-Speaker aufzutreten.

¹ „No Hate Speech Movement“ ist eine Initiative des Europarates, die 2013 ins Leben gerufen wurde mit dem Ziel, gegen Hassrede im Internet und für ein respektvolles Miteinander aufzutreten.

² Die Aktionsforschung wurde im Rahmen des Forschungsprojekts „#UnitedAgainstHate“ (2020–2022) durchgeführt, das durch das FTI-Programm KIRAS des Bundesministeriums für Finanzen gefördert bzw. finanziert und von der Österreichischen Forschungsförderungsgesellschaft abgewickelt wurde.

³ Wörtliche Zitate werden mit der Gruppe (G1–G4) gekennzeichnet, aus der sie kommen. Weitere Informationen (z.B. Geschlecht, Alter) werden nicht zugeordnet, da die Jugendlichen im Gruppenkontext agierten und entsprechend Gruppenmeinungen abgebildet wurden.

⁴ Bei allen angeführten Zitaten werden teils bewusst gemachte Tipp- oder Rechtschreibfehler der Jugendlichen übernommen.

Quellenangaben

Atzmüller, Christiane/Zartler, Ulrike (2022). Jugendliche kontern Online-Hass: Partizipative digitale Aktionsforschung zur Gestaltung von Counter Speech, *Sozialwissenschaftliche Rundschau*, 62 (2), 155–178.

Atzmüller, Christiane et al. (2019a). Zivilcourage 2.0 – Zivilcourage von Jugendlichen im Umgang mit wahrgenommener Gewalt im Internet, *Forschungsendbericht, KIRAS-Sicherheitsforschung, FFG*, Online: <https://tinyurl.com/2p8r92zw> (12.06.2023).

Atzmüller, Christiane et al. (2019b). Online-Held*innen gibt es nicht? Was 14- bis 19-jährige Jugendliche an Zivilcourage im Internet hindert, *Sozialwissenschaftliche Rundschau*, 59 (1), 87–109.

Butler, Leah C. et al. (2021). Examining the Effect of Perceived Responsibility on Online Bystander Intervention, Target Hardening, and Inaction, *Journal of Interpersonal Violence*, 37 (21–22), Online: <https://doi.org/10.1177/08862605211055088> (29.06.2023).

Chandrasekharan, Eshwar et al. (2017). You Can't Stay Here. The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech, *Proceed-*

ings of the ACM on Human-Computer Interaction, 1 (CSCW) (31), 1–22.

Cobbe, Jennifer (2021). Algorithmic Censorship by Social Platforms: Power and Resistance, *Philosophy & Technology* (34), 739–766.

DeSmet, Ann et al. (2016). Deciding Whether to Look After them, to Like it, or Leave it: A Multidimensional Analysis of Predictors of Positive and Negative Bystander Behavior in Cyberbullying among Adolescents, *Computers in Human Behavior* (57), 398–415.

Dillon, Kelly P./Bushman, Brad J. (2015). Unresponsive or Unnoticed? Cyberbystander Intervention in an Experimental Cyberbullying Context, *Computers in Human Behavior* (45), 144–150.

Gröndahl, Tommi et al. (2018). All You Need is "Love": Evading Hate-speech Detection, Online: <https://tinyurl.com/b5ezt3x2> (12.06.2023).

Herry, Emily et al. (2021). Digital era bullying: An Examination of Adolescent Judgments about Bystander Intervention Online, *Journal of Applied Developmental Psychology* (7), 1–11.

Machackova, Hana et al. (2018). Bystanders' Supportive and Passive Responses to Cyberaggression, *Journal of School Violence*, 17 (1), 99–110.

Mayring, Philipp (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*, Weinheim/Basel.

MPFS [Medienpädagogischer Forschungsverbund Südwest] (2022a). *JIM-Studie 2022, Jugend, Information, Medien, Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger*, Online: <https://tinyurl.com/yh7rs7vc> (12.06.2023).

MPFS [Medienpädagogischer Forschungsverbund Südwest] (2022b). *JIMplus 2022, Fake News und Hate Speech*, Online: <https://tinyurl.com/5bshvv6d> (12.06.2023).

Olenik-Shemesh, Dorit et al. (2017). Bystanders' Behavior in Cyberbullying Episodes. Active and Passive Patterns in the Context of Personal-Socio-Emotional Factors, *Journal of Interpersonal Violence*, 32 (1), 23–48.

Patterson, Lisa J. et al. (2016). Adolescent Perceptions of Bystanders' Responses to Cyber-

- bullying, *New media & Society*, 19 (3), 366–383.
- Seering, Joseph et al. (2017). *Shaping Pro and Anti-Social Behavior on Twitch Through Moderation and Example-Setting*. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 111–125, Online: <https://tinyurl.com/yckr2nmf> (12.06.2022).
- Shultz, Emily et al. (2014). *Cyberbullying: An Exploration of Bystander Behavior and Motivation*, *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 8 (4), Online: doi: <https://doi.org/10.5817/CP2014-4-3> (20.04.2022).
- Van Cleemput, Katrien et al. (2014). *Personal Characteristics and Contextual Factors that Determine “Helping”, “Joining in”, and “Doing Nothing” when Witnessing Cyberbullying*, *Aggressive Behavior*, 40 (5), 383–396.
- Weber, Danielle M./Dickter, Cheryl L. (2015). *Confronting the “F” Word. The Effects of Gender, Ambiguity, and Individual Difference Variables on Non-Targets’ Confrontation of Heterosexist Comments*, *Journal of Homosexuality*, 62 (10), 1289–1312.
- Wöhrer, Veronika et al. (2017). *Partizipative Aktionsforschung mit Kindern und Jugendlichen: Von Schulsprachen, Liebesorten und anderen Forschungsdingen*, Wiesbaden.
- Wright, Lucas et al. (2017). *Vectors for Counter-speech on Twitter*, *Proceedings of the First Workshop on Abusive Language Online*, 57–62, Online: <https://tinyurl.com/bd zx84vu> (21.06.2023).