

CYBERHEROES

Mobilisierung Jugendlicher zu Counter Speech im Umgang mit Online Hass, Cybermobbing und anderen Formen digitaler Gewalt

Endbericht 2022

Christiane Atzmüller, Lina Kröncke und Ulrike Zartler



Institut für Soziologie an der Universität Wien in Kooperation mit:

- Österreichisches Institut für angewandte Telekommunikation / saferinternet.at
- Mauthausen Komitee Österreich
- Bundeskanzleramt: Referat VI/5/a – Kompetenzzentrum Jugend
- BM.I / Bundeskriminalamt / Büro 1.6 Kriminalprävention und Opferhilfe

Unter Mitarbeit von: Ingrid Kromer, Sarah Kerschhagl, Gregor Rettenegger

Finanziert im Sicherheitsforschungs-Förderprogramm KIRAS des Bundesministeriums für Landwirtschaft, Regionen und Tourismus

Inhaltsverzeichnis

Zusammenfassung und praxisrelevante Schlussfolgerungen	4
Problemstellung und Relevanz	13
Theoretische Fundierung.....	14
Zielsetzung und Forschungsfragen	21
Überblick über das methodische Gesamtdesign	22
1. Online Rollenspiele zur Erfassung des typischen Interaktionsverhaltens Jugendlicher	24
1.1 Methodisches Design	24
1.1.1 Smartphones als Erhebungsinstrumente	24
1.1.2 Konzeption der Umsetzung: Impulsgebende Szenarien und Rollenvorgaben	26
1.1.3 Prüfung auf ethische Angemessenheit.....	30
1.1.4 Umsetzung der Rollenspiel-Workshops, Sampleübersicht und Analyse­methode	32
1.2 Ergebnisse	34
1.2.1 Verwendete Counter Speech Strategien	35
1.2.2 Unterschiede im Bystander*innenverhalten nach Geschlecht, Bildung und Migrationshintergrund	43
1.2.3 Hinweis auf Erfolg und Misserfolg von Counter Speech Strategien	64
1.2.4 Wahrnehmung der Online-Rollenspiele aus der Sicht der Teilnehmer*innen.....	67
1.3 Zusammenfassung und Schlussfolgerungen	85
2. Inhaltsanalyse von realen User*innen-Reaktionen auf YouTube.....	90
2.1 Methodisches Design	90
2.1.1 Datensammlung und Stichprobendesign	90
2.1.2 Ausarbeitung des Codebooks und Datenanalyse	96
2.2 Ergebnisse	101
2.2.1 Identifikation von Hate Speech	101
2.2.2 Identifikation von Counter Speech.....	105
2.2.3 Rollenzuordnung von Opfern und Counter Speaker*innen	109
2.2.4 Darstellung der Komplexität von Hate Speech – Counter Speech Interaktionen	111
2.2.5 Bewertung der Schwere von Hate Speech	112
2.2.6 Kategorisierung und Bewertung von Counter Speech Strategien	115
2.2.7 Erfolg von Counter Speech	121
2.3 Zusammenfassung und Schlussfolgerungen	124
3. Online Aktionsforschung zur Erarbeitung von Counter Speech Vorschlägen.....	128
3.1 Methodisches Design	129
3.1.1 Datenbasis und Methodik der Digitalen Aktionsforschung.....	129
3.1.2 Konzeption der partizipativen, digitalen Aktionsforschung	130
3.2 Ergebnisse	131
3.2.1 Erfahrungen Jugendlicher mit Counter Speech im Internet-Alltag	131

3.2.2 Jugendliche Gestaltungskompetenzen und Interventions-Strategien in privaten Kontexten (z.B. WhatsApp-Gruppen)	133
3.2.3 Jugendliche Gestaltungskompetenzen und Interventions-Strategien in öffentlichen Kontexten (z.B. TikTok, Instagram, YouTube)	142
3.2.4 Counter Speech in der Praxis: Vorgehen gegen reale Hate-Kommentare	153
3.2.5 Wahrnehmung und Bewertung des Erfolgs von Counter Speech	168
3.3 Zusammenfassung und Schlussfolgerungen	170
4. Online Feld-Experiment zur Prüfung von Counter Speech Wirkungspotenzialen	177
4.1 Methodisches Design	177
4.2 Durchführung des Online Feld-Experiments	181
4.2.1 Erste Intervention Hate Speech OHNE Counter Speech (Experiment-Phase 1)	182
4.2.2 Erste Intervention Hate Speech MIT Counter Speech (Experiment-Phase 2)	184
4.2.3 Zweite Intervention Hate Speech OHNE Counter Speech (Experiment-Phase 3)	189
4.2.4 Zweite Intervention Hate Speech MIT Counter Speech (Experiment-Phase 4).....	192
4.2.5 Dritte Intervention Hate Speech OHNE Counter Speech (Experiment-Phase 5).....	195
4.2.6 Dritte Intervention Hate Speech MIT Counter Speech (Experiment-Phase 6)	197
4.3 Zusammenfassung und Schlussfolgerungen	201
<i>Unterstützungsmaßnahmen der Praxispartner*innen zur Förderung von Counter Speech</i>	<i>204</i>
Entwickelte Maßnahmen	206
ÖIAT/Saferinternet.at.....	206
Mauthausen Komitee Österreich / MKÖ.....	211
BM.I / Bundeskriminalamt / Büro 1.6 Kriminalprävention und Opferhilfe	213
Bundeskanzleramt: Referat VI/5/a – Kompetenzzentrum Jugend.....	214
Innovationsgehalt des Projekts aus Sicht der Praxispartner*innen	214
Nachhaltige Nutzungsstrategien und Visionen für die Zukunft	216
Praxisrelevante Verwertung und Verbreitung der Studienergebnisse	218
<i>Literaturverzeichnis</i>	<i>219</i>

Zusammenfassung und praxisrelevante Schlussfolgerungen

Als intensive Internetnutzer*innen sind Jugendliche verstärkt auch mit problematischen Entwicklungen und Risiken konfrontiert. Dazu gehört insbesondere der Umgang mit digitaler Gewalt: Jugendliche erleben eine enorme Bandbreite unterschiedlichster Formen von Online Hass und Cybermobbing (z.B. Hasspostings, öffentliche Bloßstellungen oder Belästigungen, ungewollte Konfrontationen mit schockierenden, pornographischen, diskriminierenden oder illegalen Inhalten, unautorisierte Veröffentlichungen von persönlichen oder manipulierten Bildern oder Videos, missbräuchliche Verwendung von Bildern/Videos oder Falschdarstellungen von persönlichen Inhalten, Online-Stalking bis hin zu digital übermittelten physischen Gewalt- oder Tötungsandrohungen).

Trotz Bemühungen vieler Plattformbetreiber*innen wie YouTube, Instagram oder TikTok, die besonders häufig von Jugendlichen genutzt werden, hat sich bisher weder das Blockieren noch das Melden oder das gezielte Löschen solcher Inhalte auf Basis von komplexen Such-Algorithmen als besonders effektiv erwiesen. Daher entstand die Idee, nicht das Löschen in den Mittelpunkt zu rücken, sondern diesen negativen Inhalten gezielt zivilcouragierte Gegenreaktionen, d.h. Counter Speech, entgegenzusetzen. Counter Speech ist auf vielfache Weise gestaltbar und umfasst jegliche Form der Kommunikation (z.B. Videobotschaften, humoristische oder Widersprüche aufzeigende Memes bzw. GIFs, faktenbasierte Gegendarstellungen, aktive Gegenrede durch Einzelpersonen).

Jugendliche Online Counter Speaker*innen, die spontan und öffentlich sichtbar gegen negative Inhalte im Internet auftreten, haben besonders hohes Potenzial, jene jugendlichen Peers zu erreichen, deren Diskurs-Normen verändert werden sollen. Bislang vorliegende Studien zeigen allerdings, dass Jugendliche überwiegend nicht handeln und Online-Übergriffe in der Regel ignorieren. Ein wesentliches Problem dabei ist, dass es Jugendlichen aus allen Bildungsschichten scheinbar massiv an Kompetenzen fehlt, wie Counter Speech erfolgreich gestaltet werden kann. Gleichzeitig fehlt es an gezielter Forschung zur Frage, wie Jugendliche in ihrem Auftreten als Counter Speaker*innen unterstützt werden können.

Die vorliegende Studie greift diesen Forschungsbedarf auf, mit dem Ziel, jugendliche Reaktionen und Interaktionen in der Begegnung mit negativen Inhalten im Internet zu analysieren, jugendgenerierte Gegenrede zu verstehen, gemeinsam mit Jugendlichen Möglichkeiten für praktikable und wirksame jugendgerechte Counter Speech zu entwickeln und damit eine Grundlage zu schaffen, um zivilcouragierte Gegenrede von Jugendlichen zu unterstützen und zu fördern. Mit der Mobilisierung Jugendlicher zu Counter Speech soll auch ein klares und nachhaltiges humanistisches Zeichen gesetzt werden, dass Online Hass bzw. jede andere Form hasserfüllter Rhetorik und digitaler Gewalt nicht toleriert wird.

In der vorliegenden Studie wird notwendige Grundlagenforschung für die Altersgruppe der 14- bis 18-jährigen Jugendlichen geleistet: Untersucht wurde (1) wie sich das Interaktionsverhalten von Jugendlichen typischerweise gestaltet, wenn diese auf negative Inhalte im Internet reagieren, (2) welche Kommunikations-Elemente sich aus Sicht von Mädchen und Jungen als praktikabel und wirkungsvoll erweisen und (3) wie diese bei der konkreten Gestaltung von Counter Speech optimiert werden können. Die gewonnenen Erkenntnisse werden von den am Projekt beteiligten Praxispartner*innen genutzt, um zielgruppengerechte Counter Speech Unterstützungsangebote in Form von Informationsmaterialien, Schulungs- und Trainingsmaßnahmen auszuarbeiten.

Mit dieser Studie wurde erstmals eine umfassende empirische Datenbasis zum Counter Speech Verhalten Jugendlicher geschaffen, die tiefgehende und differenzierte Einblicke in den Umgang Jugendlicher mit negativen Inhalten im Internet ermöglicht und dabei auch geschlechts- und bildungsspezifische Besonderheiten berücksichtigt. Dabei wurde ein breites Set an innovativen Methoden eingesetzt:

- **Online Rollenspiele:** Ziel war es, das spontane Antwortverhalten von Jugendlichen in der konkreten Online Interaktion zu erfassen. Dabei wurden nicht nur öffentlich sichtbare, sondern auch über die Privat-Chat Funktion stattfindende Interaktionen berücksichtigt. Zu diesem Zweck wurde ein aufwändiges Umsetzungsverfahren unter Berücksichtigung sicherheitsrelevanter und forschungsethischer Aspekte entwickelt (Ausarbeitung impulsgebender Szenarien und Rollenbeschreibungen, Einsatz von Social Media Fake-Profilen, Nutzung projekteigener Endgeräte). Durchgeführt wurden insgesamt 21 Online-Rollenspiel Workshops mit rund 170 14- bis 18-jährigen Jugendlichen, die sich nach Geschlecht, Bildung und Migrationshintergrund unterschiedlich zusammensetzten.
- **Inhaltsanalyse:** Untersucht wurde das spontane, „alltägliche“ und öffentlich sichtbare Antwortverhalten von jugendlichen User*innen-Kommentaren auf ausgewählten YouTube Kanälen. Dabei wurden auch wichtige Erkenntnisse in Bezug auf die Identifizierung und Klassifizierung von Hate- und Counter Speech Interaktionen gewonnen. Die Umsetzung gestaltete sich als besonders herausfordernd, sowohl hinsichtlich der Datensammlung (z.B. lückenhafte Daten aufgrund automatisierter Filter, passendes Stichprobendesign, Selektion relevanter Interaktionen mittels „Wörterbuch“ [=Schlagwortliste mit im jugendlichen Netzjargon gebräuchlichen Schimpfwörtern und gewalthaltigen Begriffen] aus insgesamt 32.852 Kommentarblöcken von 517 Videos, die mittels eigenem Scraper heruntergeladen wurden), als auch die Datenanalyse betreffend (Codebook zur Identifizierung von Hate- und Counter Speech Interaktionen und zur Dokumentation jugendtypischer Counter Speech Strategien). Als besonders problematisch erwies sich die große Uneinigkeit der einzelnen Kodierer*innen in der Wahrnehmung und Bewertungen von Hate- und Counter Speech Interaktionen, trotz mehrfacher Schulungen und detaillierter Codebook-Definitionen mit Fallbeispielen. Da gerade die Identifizierung von Hate Speech bei automatisierten Detection Tools oft sehr problematisch ist und auch Jugendliche aufgrund fehlender Kontextinformationen, subtiler

sprachlicher Darstellungs- und Ausdrucksweisen oder auch aufgrund von Sensibilisierungs- und Wissensdefiziten auf große Schwierigkeiten stoßen, wurden schließlich diese Wahrnehmungs- und Bewertungsunterschiede in den Analysefokus gerückt. Als Analysegrundlage wurden 500 ausgewählte Kommentarblöcke herangezogen, die insgesamt 4 Mal durch 4 Kodierer*innen mittels Codebook kategorisiert und bewertet wurden (insgesamt 2000 Bewertungen).

- **Online Aktionsforschung zur Erarbeitung von Counter Speech Vorschlägen:** In enger Zusammenarbeit zwischen Jugendlichen, Forscher*innen und Praxispartner*innen wurden unterschiedliche Gestaltungsmöglichkeiten jugendgerechter Counter Speech erarbeitet und diskutiert und im konkreten Einsatz mit Jugendlichen erprobt. Gearbeitet wurde mit insgesamt 31 Jugendlichen im Alter von 14 bis 22 Jahren, diese wurden in vier Gruppen à 5-11 Teilnehmer*innen organisiert, die sich jeweils nach Geschlechtszusammensetzung und Bildungshintergrund unterschieden. Jede Gruppe traf sich je vier Mal (= insgesamt 16 Workshops) in einem Workshop-Format mit dem Forschungsteam. Dabei wurden verschiedene Phasen der Forschung und Aktion gemeinsam gestaltet, z.B. um auf Basis von Online-Rollenspiel-Auszügen Counter Speech Optimierungspotenziale zu erarbeiten. Weiters wurden Counter Speech Simulationsübungen in Form von „Speed Counter Speech“ durchgeführt, mit Hilfe von Fake-Profilen gegen reale Hate-Kommentare vorgegangen, oder Quizfragen zum Erfolg von Counter Speech entwickelt.
- **Online Feld-Experiment:** In Zusammenarbeit mit einem bekannten YouTuber wurde Counter Speech durch Jugendliche in einem kontrollierten Setting (gezielt inszenierte Online-Übergriffe und Counter Speech Interventionen) konkret eingesetzt und hinsichtlich möglicher Wirkungseffekte geprüft. Dabei wurde mit einem bekannten YouTuber mit relativ hoher Reichweite kooperiert und eine „Counter Speech Gruppe“ von zehn 14- bis 19-jährigen Jugendlichen (geschlechtergemischt) eingerichtet, die vom Forschungsteam begleitet wurde und für ihren Einsatz vorbereitete Fake-Profile nutzte. Insgesamt wurde das Online Feld-Experiment über einen Zeitraum von 5 Monaten mit insgesamt 6 Interventionsphasen umgesetzt (abwechselnde Inszenierung von Hate Speech einmal mit und einmal ohne zusätzlich inszenierter Counter Speech).
- **Entwicklung von Unterstützungsangeboten zur Förderung von Counter Speech:** Aufbauend auf den Erkenntnissen wurden durch die Praxispartner*innen ÖIAT/Saferinternet.at, Mauthausen Komitee Österreich, das Büro 1.6 Kriminalprävention und Opferhilfe des Bundeskriminalamts/BM.I und das Bundeskanzleramt / Sektion VI: Familie und Jugend /REFERAT VI/5/a: Kompetenzzentrum Jugend Informationsmaterialien, Schulungs- und Trainingsmaßnahmen entwickelt, um Counter Speech durch Jugendliche gezielt zu fördern. Die beteiligten Projektpartner*innen nehmen zentrale Positionen für (Online) Zivilcourage bzw. für den sicheren Umgang von Jugendlichen mit digitalen Medien in einem österreichweiten und internationalen Netzwerk ein, daher ist ein hohes und breites Nutzungs- und Einsatzpotenzial vorhanden.

Die **zentralen Erkenntnisse** aus dem Gesamtprojekt werden im Folgenden komprimiert zusammengefasst. Zusätzlich finden sich ausführliche Zusammenfassungen jeweils am Ende der einzelnen Berichtsteile im Gesamtbericht.

Die Ergebnisse lassen sich zunächst bezugnehmend auf die **Forschungsfrage, wie sich das Interaktionsverhalten von Jugendlichen typischerweise gestaltet, wenn diese auf negative Inhalte im Internet reagieren bzw. welche unterschiedlichen Counter Strategien auffindbar sind**, wie folgt zusammenfassen: Wie insbesondere aus den durchgeführten Online-Rollenspielen ersichtlich wurde, verfügen Jugendliche prinzipiell über ein breites Repertoire an unterschiedlichen Counter Speech Strategien, deren Großteil aber auf Täter*innen fokussiert ist (z.B. Warnungen, Abwertungen, Gegenattacken, Aufforderungen aufzuhören, sachliche Gegenargumente und Fakten, moralische Appelle). Deeskalierende Strategien, z.B. auch in Form von Ablenkungen, werden eher an alle Beteiligten gerichtet. Counter Speech Strategien, die sich gezielt an das Opfer wenden (z.B. um zu trösten oder zu beraten), kommen deutlich seltener vor und finden vorwiegend im Privat-Chat statt. Sehr selten richten sich Counter Speech Strategien an andere Bystander*innen, um nach Unterstützung für das eigene Eintreten zu suchen (z.B. in Form von Likes oder kurzen Statements).

Die Bandbreite an möglichen Counter Speech Strategien wird von Jugendlichen sehr unterschiedlich eingesetzt, abhängig von der Gestaltung der Übergriffe selbst (z.B. gemäßigte vs. sehr ausfällige Übergriffe). Darüber hinaus zeigen sich deutliche geschlechts- und bildungsbezogene Unterschiede sowie Unterschiede zwischen Jugendlichen mit und ohne Migrationshintergrund. So kontern männliche Jugendliche mit höherem Bildungshintergrund eher überheblich, herablassend und genervt und heben gegenüber Täter*innen deren Unreife hervor. Burschen mit geringerer formaler Bildung versuchen eher, durch aggressive Gegenattacken ernst genommen zu werden. Weibliche Jugendliche mit höherer Bildung mahnen stärker die Beibehaltung einer zivilisierten Gesprächskultur ein und weisen klar auf Regelverstöße hin, Mädchen mit geringerer Bildung argumentieren eher moralisierend und zeigen sich gegenüber Täter*innen konfrontativer. Jugendliche mit Migrationshintergrund betonen moralische Aspekte oft stärker und drücken dies teilweise auch mit einer sehr expliziten Sprache aus, die aufgrund ihrer Emotionalität häufig angriffig wirkt. Mädchen verfügen insgesamt über ein vielfältigeres Interventionsrepertoire als Burschen und agieren auch gegenüber Opfern stärker unterstützend. Burschen setzen hingegen eher auf Ablenkung und Irritation anstatt auf konstruktive Lösungsversuche.

Häufig sind jugendliche Counter Speaker*innen zunächst damit beschäftigt, sich als Counter Speaker*innen überhaupt Gehör zu verschaffen: Dies geschieht teils mit aggressiven Kraftausdrücken, teils wird versucht, die Kommunikation mit Ablenkungstaktiken zu durchbrechen oder es werden deeskalierende Aufrufe getätigt. In vielen Fällen zeigen sich Jugendliche insbesondere bei sehr ausfälligen Übergriffen aber auch überfordert und reagieren darauf nur wenig oder kaum. Insbesondere bei

Jugendlichen aus Schulen mit formell geringerer Bildung fällt auf, dass diese oft relativ lange inaktiv bleiben und insgesamt der Eindruck von Hilflosigkeit und Ohnmacht entsteht.

Der starke Fokus auf Täter*innen-bezogene Strategien hinterlässt bei Jugendlichen das Gefühl, mit den eigenen Interventionen letztlich erfolglos zu sein. Das eigene Auftreten als Counter Speaker*in wird als wirkungslos und anstrengend empfunden, mit höchstens punktuellen und kurzfristig Erfolgen, etwa durch Ablenkung mit lustigen GIFs. Entsprechend werden gängige Strategien wie Blockieren, Melden, der Ausschluss aus der Gruppe, oder das weit verbreitete Ignorieren der Konflikte vergleichsweise als am effizientesten betrachtet.

In diesem Zusammenhang müssen auch die großen Kontextunsicherheiten sowie Differenzen in der Wahrnehmung und Bewertung von Online-Übergriffen berücksichtigt werden, denen Jugendliche in ihrer Rolle als Counter Speaker*innen ausgesetzt sind, wie insbesondere die Inhaltsanalyse der YouTube-Daten zeigt. Was als Online-Übergriff bzw. Hate Speech gilt und was nicht, bzw. als wie schwerwiegend solche negativen Inhalte einzuordnen sind, ist schwer zu klassifizieren. Dementsprechend bleibt auch unklar, ob der eigene Counter Speech Einsatz gefragt ist oder nicht. Dies ist umso schwieriger, je subtiler zum Beispiel Übergriffe formuliert werden (z.B. ohne Einsatz von „gängigen“ Schimpfwörtern), je unklarer bestimmte Diskriminierungsformen im gesellschaftlichen Wertgefüge eingeordnet sind (z.B. Transfeindlichkeit) und je komplexer sich Interaktionsverläufe darstellen, wo zum Teil die Trennung zwischen Hate- und Counter Speech unklar bleibt, was es erschwert, den Überblick zu behalten.

Die Schwierigkeiten in der Identifizierung von Hate Speech bedingen auch häufig Probleme mit der Identifizierung von Counter Speech, die potenziell unterstützt werden könnte, zumal diese in zumindest der Hälfte der analysierten YouTube-Daten als aggressive Gegenattacken formuliert waren. Counter Speech wird auch auf öffentlichen Plattformen als insgesamt erfolglos wahrgenommen: Hate Speech / Counter Speech Interaktionen enden häufig mit einem negativen Hate Kommentar bzw. wird Hate Speech im Laufe von Interaktionen in der Regel verstärkt.

Insofern lässt sich auch die im Projekt gestellte **Forschungsfrage, inwieweit sich bestimmte Counter Speech Strategien als produktiv oder kontraproduktiv klassifizieren lassen**, nur schwer beantworten: Im Rahmen der Online Rollenspiele sind keine Erfolgshinweise im Hinblick auf Absender*innen negativer Inhalte (z.B. Einsicht/Entschuldigung, positive Verhaltensänderungen, positive Entwicklung von Diskursnormen) auffindbar, zudem blieben Täter*innen gemäß ihren Vorgaben auch in ihrer Rolle. Die Analyse der YouTube-Daten gibt nur bedingt Hinweise: so scheinen Hate Speech Folgeaktionen bei humorvollen Counter Speech Strategien tendenziell etwas weniger übergriffig auszufallen, weitere positive Entwicklungen fanden sich insgesamt am ehesten dort, wo Counter Speech Strategien eine entsprechend positive Gesprächsatmosphäre zulassen (z.B. Humor, Verständnis zeigen, deeskalieren), der

Anteil an festgestellten positiven Entwicklungen in Bezug auf Täter*innen ist aber sowohl bei den analysierten YouTube-Daten, als auch im gesamten empirischen Datenmaterial der Studie äußerst gering.

Als produktiv könnten Counter Speech Strategien auch dann eingestuft werden, wenn es gelingt, andere Bystander*innen als Unterstützung für das eigene Counter Speech Verhalten zu mobilisieren. Obwohl in den Rollenspiel-Daten Bystander*innen-bezogene Strategien nur in geringem Umfang zu finden sind, lassen sich erste Hinweise für eine Mobilisierung anderer ableiten: Es werden eher solche Counter Speech Strategien durch andere unterstützt, (a) die die Verbesserung der eigenen Befindlichkeit (es ist unangenehm, mit Konflikten konfrontiert zu sein, dritte Personen fühlen sich davon genervt) in Aussicht stellen (z.B. schnelle Lösungen zur Wiederherstellung einer angenehmen Gesprächsatmosphäre), (b) die eine Orientierung anbieten, wo ein Eingreifen tatsächlich angebracht ist und wo nicht, indem Übergriffe klar als Normverstöße oder als moralische Verwerflichkeit deklariert werden (z.B. „das ist strafbar“, „das ist unterste Schublade“) und (c) die eine klare Anweisung enthalten (z.B. meldet alle) oder vorbildhafte Interventionen zur Nachahmung anbieten (z.B. öffentliche Zeichen der Solidarität: alle schreiben Nachrichten mit Namen des Opfers und positiven Emojis).

Schließlich kann auch für die letzte **Forschungsfrage, welche Formen der aktiven Gegenrede aus Sicht von Jugendlichen als praktikabel und wirkungsvoll erscheinen bzw. welche Optimierungspotenziale sich aus dem Counter Speech Verhalten durch Jugendliche ableiten lassen**, folgendes als Antwort zusammengefasst werden: Spätestens im Zuge der Online Aktionsforschung wurde klar, dass Jugendliche unter Counter Speech nicht unbedingt eine zivilcouragierte Handlung verstehen, sondern beispielsweise auch mit Formen der Zensur in Verbindung bringen, wie etwa die „Cancel Culture“, d.h. der systematische Ausschluss von bzw. Angriff auf Personen (was prinzipiell auch als wirkungsvoll eingeschätzt wird). Auch Reaktionen von Fan-Gruppen, die Celebrities unterstützen und Reaktionen des Opfers selbst werden als Counter Speech zugeordnet. Wahrgenommene Interventionen durch Dritte werden eher als unterhaltsam zu lesen wahrgenommen, aber nicht als mutiges Einschreiten betrachtet.

Da Jugendliche prinzipiell wenig Erfahrung im Praktizieren von Counter Speech haben, war auch die Entwicklung von Optimierungspotenzialen im Rahmen dieser Studie mitunter herausfordernd. Zudem ist das eigene Auftreten als Counter Speaker*in prinzipiell eher negativ konnotiert und wird als mühsam, langwierig, überfordernd, als unnötiges Einmischen und letztlich als Zeitverschwendung gesehen. In gemeinsamen Phasen der Forschung und Aktion konnten aber dennoch konstruktive und praxistaugliche Counter Speech Vorschlägen entwickelt werden: Die Ergebnisse zeigen, dass Jugendliche für öffentliche und private Kontexte unterschiedliche Counter Speech Strategien vorschlagen und dass dabei auch bildungsspezifische Differenzen sichtbar werden: in privaten Kontexten wie z.B. auf WhatsApp, spielen Jugendliche mit höherer Bildung ihre intellektuellen Vorteile aus, um in ihrem Auftreten als Counter Speaker*in ernst genommen zu werden (z.B. durch Posten recherchierter Gesetzestexte); Jugendliche mit niedrigerer Bildung versuchen hingegen eher, durch Gegenattacken ernst genommen zu werden. Die

jeweiligen Erfolgsaussichten wurden aber von den Jugendlichen der jeweils anderen Bildungsebene durchaus skeptisch eingeschätzt, indem argumentiert wurde, dass z.B. Gesetzestexte entweder gar nicht gelesen oder nicht ernst genommen würden. Andererseits wurde befürchtet, dass Gegenattacken zu einer Eskalation der Gesamtsituation führen würden. Jugendliche mit geringerer Bildung schlugen auch häufiger moralisierende Zugänge aus der Opferperspektive vor; aus Sicht der Jugendlichen mit höherer Bildung wurden solche Strategien aber genau dadurch als eher schwach bewertet. In größeren öffentlichen Kontexten, wie z.B. auf YouTube sahen Jugendliche mit höherer Bildung eher schlagfertige, überhebliche, sich lustig machende Counter Speech Strategien als erfolgversprechend, jene mit geringerer Bildung eher aggressiv-emotionale Reaktionen. Auffallend ist auch, dass Jugendliche mit höherer Bildung Counter Speech häufiger mit z.B. inhaltlich passenden Memes kombinieren, um damit ihren Botschaften mehr Aufmerksamkeit zu verleihen und insgesamt in der sprachlichen Versiertheit (z.B. Formulierung von schlagfertigen Retourkutschen) und in der argumentativen Darlegung der eigenen Positionierung bei kontrovers diskutierten Themen klar im Vorteil sind. Insgesamt lässt sich zusammenfassen, dass Counter Speech im Sinne von Online Zivilcourage durch Jugendliche nur im Maßstab vorhandener Basiskompetenzen gestaltbar ist und bildungsbezogene Unterschiede berücksichtigt werden müssen.

Im Zuge der Aktionsforschung wurde auch deutlich, dass viele Counter Speech Vorschläge insbesondere in öffentlichen Kontexten nicht praktikierbar sind: So argumentieren die teilnehmenden Jugendlichen, dass aufgrund der Anonymität z.B. die Konfrontation mit Fakten zu möglichen strafrechtlichen Folgen nicht ernst genommen wird. Gleichzeitig schränken auch technische Gegebenheiten Counter Speaker*innen häufig ein, beispielsweise ist das Posten von Memes auf YouTube und TikTok nicht möglich.

Die von Jugendlichen entwickelten Gestaltungsvorschläge wurden schließlich im Rahmen der Aktionsforschung hinsichtlich ihrer Praktikabilität im Vorgehen gegen reale Hate-Kommentare auf YouTube und TikTok konkret eingesetzt – allerdings endete dies schließlich mit dem resignierenden Gefühl der beteiligten Jugendlichen, letztlich in ihren Counter Speech Bemühungen nicht erfolgreich gewesen zu sein. Das lag vor allem daran, dass die Wirkung und der Erfolg der eingesetzten Counter Speech Strategien für Jugendliche schlicht nicht erkennbar war, da positive und unterstützende Reaktionen auf ihr Counter Speech Verhalten, z.B. in Form von Likes oder Kommentaren, in der Regel ausblieben.

Ähnlich zeigte sich dies auch im durchgeführten Online Feld-Experiment, wo versucht wurde, durch intensive Counter Speech auch andere reale User*innen zu mobilisieren. Entsprechende Effekte zeigten sich aber nur in einem sehr geringen Ausmaß und nur eher kurzfristig bzw. verlagerte sich Counter Speech eher auf eine indirekte Ebene (z.B. Nutzung der Meldefunktion; indirekte Counter Speech Kommentare z.B. „was ist los mit den Leuten“) und nicht auf aktives Vorgehen.

Schlussfolgerungen für die Praxis

Aus den gewonnenen Erkenntnissen lassen sich zusammenfassend folgende praxisrelevante Empfehlungen ableiten, um Counter Speech durch Jugendliche nachhaltig zu fördern:

- **Aufklärung, was Counter Speech ist und was sie bewirkt**

Counter Speech wird durch Jugendliche nicht unbedingt mit zivilcouragierten Handlungen in Verbindung gebracht und ist tendenziell sogar eher negativ besetzt oder wird zumindest als unwirksam betrachtet. Besonders wichtig wäre es daher, Jugendliche darüber aufzuklären, was unter Counter Speech zu verstehen ist, aber auch Überzeugungsarbeit zu leisten, dass Counter Speech 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.

- **Erklärung, wann und warum Counter Speech erfolgreich ist**

In diesem Zusammenhang sind auch Erklärungen erforderlich, was unter erfolgreicher Counter Speech verstanden werden kann bzw. wie dieser Erfolg wahrgenommen werden kann (z.B. durch unterstützende Reaktionen der Community in Form von Likes und Kommentaren, nicht unbedingt als positive Reaktion jener Personen, die negative Inhalte verbreiten). Als zentral erscheint hierbei die hohe Bedeutung des eigenen guten Gefühls bzw. guten Gewissens, wenn gehandelt und nicht einfach ignoriert wird. Zur Bekräftigung und Illustration könnten auch gelungene Counter Speech Beispiele als Vorbildwirkung eingesetzt werden.

- **Sensibilisierung für das Erkennen und Benennen von Hate Speech**

Die Ergebnisse verdeutlichen, dass es für eine Mobilisierung zu Counter Speech einer hohen Sensibilisierung in der Wahrnehmung von Hate Speech bedarf, auch in Hinblick auf die Bewertung ihrer gravierenden (auch strafrechtlichen) Bedeutung und ihrer Auswirkung auf Betroffene. Eine entsprechende Bewusstseinsbildung und Sensibilisierung könnte am ehesten über das Bildungssystem erreicht werden. Aufgrund der großen Unsicherheiten in der Wahrnehmung und Bewertung von Hate Speech könnte es auch Teil von Counter Speech Strategien sein, Hate Speech klar als solche zu deklarieren – auch mit dem Gedanken, dies für andere Bystander*innen erkenntlich zu machen und ihnen eine Orientierung zu bieten. Gleichzeitig wäre es wichtig, Jugendliche zu mobilisieren, möglichst häufig als Counter Speaker*innen gegen Hate Speech aufzutreten, um das wahrgenommene Ungleichgewicht zwischen Hate- und Counter Speech auszugleichen – und dies möglichst ohne aggressive Gegenattacken, damit für andere Bystander*innen klar trennbar bleibt, wer als Hate- und wer als Counter Speaker*in einzuordnen ist.

- **Übung von Counter Speech**

Vielen Jugendlichen fehlt es schlicht an Übung, Counter Speech umzusetzen. Um die Umsetzungs- und Gestaltungskompetenzen insbesondere von Jugendlichen mit geringerer Bildung zu stärken, ist eine Vermittlung von Gestaltungsoptionen unter Berücksichtigung vorhandener Basiskompetenzen (z.B. sprachlich-argumentative Fähigkeiten) erforderlich. Gleichzeitig wird auch ein entsprechender Rahmen benötigt, damit die Umsetzung von Counter Speech nicht zwangsläufig als unangenehm, sondern vielleicht sogar als attraktiv erlebt wird. In diesem Zusammenhang wären auch zukünftige technologische Entwicklungen gefordert, Counter Speaker*innen in ihrem Handeln stärker zu unterstützen, indem z.B. Jugendlichen mit geringeren Kompetenzen automatisierte Counter Speech Vorschläge zur Verfügung gestellt werden. Zusätzlich könnte auch der Erfolg von Counter Speech durch gezielte, automatisierte Unterstützung z.B. durch Likes und positive Kommentare sichtbar gestaltet werden.

- **Fokussierung auf Counter Speech Strategien, die sich an andere Bystander*innen richten**

Das Repertoire an vorhandenen Counter Speech Strategien unter Jugendlichen ist stark auf Täter*innen fokussiert. Die Möglichkeit, andere Bystander*innen zur Unterstützung des eigenen Auftretens als Counter Speaker*in zu mobilisieren, wird von Jugendlichen zumeist nicht mitgedacht. Es wäre daher wichtig, den Fokus bewusst auch auf andere Bystander*innen zu lenken, wodurch Counter Speaker*innen auch entlastet werden würden, da sich diese gegenüber Täter*innen in der Regel als erfolglos sehen. Mitzudenken wären in diesem Zusammenhang auch Möglichkeiten der Vernetzung und des Zusammenschlusses zwischen Bystander*innen, um nicht alleine gegen Hate vorgehen zu müssen.

- **Mobilisierung jugendlicher Bystander*innen zur Unterstützung anderer Counter Speaker*innen**

Um die Sichtbarkeit von Counter Speech zu erhöhen und um sich gegenseitig zu unterstützen, könnten Jugendliche stärker dazu ermutigt werden, bestehende Counter Speech Kommentare zu unterstützen (z.B. auch durch einfache Likes). Dies hätte auch das Potential, gut gemeinte, aber aggressiv formulierte Gegenattacken zu relativieren. Unterstützend kann es auch sein, Interaktionsketten bewusst mit positiven Counter Speech Kommentaren abzuschließen. Solche Signale des Erfolgs könnten die Wahrnehmung der Sinnhaftigkeit von Counter Speech insgesamt verbessern.

Problemstellung und Relevanz

Der hohe Anteil an Hate Speech, Cybermobbing und anderen Formen digitaler Gewalt im Internet betrifft vor allem junge Menschen: In einer repräsentativen Umfrage unter 1.200 Jugendlichen in Deutschland gaben 74% der Befragten an, im letzten Monat mit Hassbotschaften, extremen politischen Ansichten, Verschwörungstheorien, beleidigenden Kommentaren oder Fake News konfrontiert gewesen zu sein (MPFS 2020). In einer repräsentativen österreichischen Studie unter 1.868 14- bis 19-jährigen Schüler*innen 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 (Atzmüller et al. 2019a).

Viele Plattformbetreiber*innen, wie z.B. YouTube, Instagram oder TikTok, die besonders häufig von Jugendlichen genutzt werden, begegnen diesem Problem mit dem gezielten Löschen negativer Inhalte sowie der Blockierung oder Sperrung entsprechender User*innen-Accounts. Damit sind aber auch viele Probleme verbunden: (i) Automatisierte Such-Algorithmen haben aufgrund der großen Bandbreite an text- und bildbasierten Ausdrucksmöglichkeiten große Probleme, Hate Speech umfassend zu identifizieren (Gröndahl et al. 2018). (ii) Die Bewältigung der riesigen Mengen an negativen Inhalten, die durch User*innen gemeldet werden (Nutzung verfügbarer Meldefunktionen) erfordert eine große Anzahl menschlicher Content-Moderator*innen, um die gemeldeten Inhalte zu überprüfen; die damit einhergehende hohe psychische Belastung für die Moderator*innen führt zu oft wenig schlüssigen Entscheidungen (Cobbe 2020). (iii) Verbotsstrategien implizieren Einschränkungen der Meinungsfreiheit und bergen die Gefahr, dass negative Inhalte auf unzensurierte Plattformen verlagert werden, deren Nutzer*innen sich in ihren negativen Einstellungen gegenseitig verstärken (Chandrasekharan et al. 2017).

Daher entstand die Idee, nicht das Löschen als Strategie gegen Hate Speech in den Mittelpunkt zu rücken, sondern diesen negativen Inhalten gezielt Gegenreaktionen, d.h. Counter Speech, entgegenzusetzen. Die Mobilisierung Jugendlicher zu aktiver und öffentlich sichtbarer Gegenrede, die wir im vorliegenden Bericht vorwiegend auf spontanes Antwortverhalten als unmittelbare Reaktion auf negative Inhalte im Internet beziehen, ist ein besonders vielversprechender Zugang, um ein klares, humanistisches Zeichen zu setzen, dass Online Hass bzw. jede andere Form hasserfüllter Rhetorik und digitaler Gewalt nicht toleriert wird. Gerade Jugendliche haben hohes normveränderndes und deeskalierendes Potenzial und können Konfliktverläufe entscheidend beeinflussen (Weber & Dickter 2015).

Counter Speech ist auf vielfache Weise gestaltbar und umfasst jegliche Form der Kommunikation (z.B. Videobotschaften, humoristische oder Widersprüche aufzeigende Memes / GIFs, faktenbasierte Gegendarstellungen, aktive Gegenrede durch Einzelpersonen), die sich solchen negativen Inhalten entgegenstellt oder diesen widerspricht. Counter Speech ist auch als eine zivilcouragierte Handlung zu

verstehen, die den Grundstein für eine aktive demokratische Online-Gesellschaft basierend auf humanistischen und prosozialen Werten legt.

Allerdings ist bislang gerade unter Jugendlichen aktive Counter Speech kaum Thema: Das zeigt sich sowohl in einer ersten österreichischen empirischen Studie dazu (Atzmüller et al. 2019 a,b) als auch in internationalen wissenschaftlichen Arbeiten (DeSmet et al. 2016; Dillon & Bushman 2015; Olenik-Shemesh et al. 2017; Shultz et al. 2014; Van Cleemput et al. 2014). Die Studienergebnisse zeigen, dass Jugendliche überwiegend nicht handeln oder entsprechende Vorfälle ignorieren. Selbst bei über 20-Jährigen mit höherem Bildungshintergrund steigt der Anteil aktiver Counter Speech Aktivitäten zum Teil nur mäßig an, wie eine jüngere Studie unter Studierenden zeigt (Henson et al. 2020).

Eine wichtige Ursache für die geringe Verbreitung von Counter Speech unter Jugendlichen scheint darin zu liegen, dass es Jugendlichen aus allen Bildungsschichten offenbar massiv an Kompetenzen fehlt, wie zivilcouragierte Gegenrede als aktive und sichtbare Reaktion auf Cybermobbing, Online Hass und andere Formen digitaler Gewalt erfolgreich gestaltet werden kann. Die vorliegende Studie greift diesen Forschungsbedarf mit dem Ziel auf, jugendliches Counter Speech Verhalten zu verstehen und gemeinsam mit Jugendlichen Möglichkeiten für Counter Speech zu entwickeln. Damit soll eine Grundlage geschaffen werden, um jugendgenerierte Gegenrede zu unterstützen und zu optimieren und als langfristigen und nachhaltigen Weg zur Bekämpfung von Online Hass, Cybermobbing und anderen Formen digitaler Gewalt zu fördern.

Theoretische Fundierung

Counter Speech gilt dann als erfolgreich, wenn es gelingt, ideologische oder normative Einstellungen und das Verhalten (1) von Absender*innen negativer Inhalte und jenen, die negative Inhalte unterstützen, zu verändern, aber auch, wenn sich ein solcher positiver Effekt (2) in der Online-Öffentlichkeit zeigt, indem zum Beispiel bestehende Diskurs-Normen verändert oder Counter Speech Inhalte unterstützt und weiterverbreitet werden. Aufgrund der Größe der Online-Community kann dieser Effekt bedeutsamer sein als die gezielte Beeinflussung einzelner Personen, die Hassrede verbreiten (Benesch et al. 2016a). Eine solche norm- und verhaltensverändernde Funktion wurde bisher vor allem dem Einsatz von breit angelegten Counter Speech Initiativen zugeschrieben, z.B. in Form von gezielten Kampagnen¹. Inwieweit auch aktive, spontane Gegenrede durch einzelne Internetnutzer*innen (z.B. User*innen-Kommentare) zu Normveränderungen beitragen kann, wurde bislang noch kaum untersucht.

¹ z.B. No Hate Speech Movement des Europarats (www.nohatespeechmovement.org) mit nationalen Kampagnen in über 40 Ländern, darunter auch in Österreich (www.bmfj.gv.at/jugend/lebensqualitaet-miteinander/nohatespeech/komitee_nohatespeech)

Die Bedeutung von User*innen-Kommentaren darf gerade bei Counter Speech Maßnahmen nicht unterschätzt werden: So zeigten Lee und Jang (2010), dass Rezipient*innen von Online-Nachrichtenbeiträgen ihre Meinung ändern, wenn den gezeigten Inhalten in User-Kommentaren widersprochen wird. Vor diesem Hintergrund ist es auch wahrscheinlich, dass öffentliche User*innen-Kommentare die Wirkung von Counter Speech Kampagnen beeinflussen können (Ernst et al. 2017): z.B., wenn direkt unter einem Counter Speech Video hasserfüllte Kommentare zu lesen sind, die von jenen stammen, deren negative Einstellungen und Verhaltensweisen verändert werden sollen (Frischlich et al. 2017). Die Massivität an negativen Ausdrucksweisen in User*innen-Kommentaren wird außerdem durch die Anonymität des Internets zusätzlich gefördert (Halpern & Gibbs 2013) und eine aggressive Wortwahl von User*innen führt auch zu einem Anstieg an aggressiver Wortwahl anderer (Rösner & Krämer 2016). Zur Deeskalation können positive User*innen-Kommentare durch Internetnutzer*innen in ihrer Funktion als Counter Speaker*innen somit von entscheidender Bedeutung sein.

Gerade unter Jugendlichen ist der Einsatz von spontanem Antwortverhalten wahrscheinlicher als die Initiierung von Counter Speech Kampagnen. Diese werden zudem vorwiegend als breit angelegte Gegenstrategien zu Hass, Extremismus oder Propaganda betrachtet. Jugendliche sind in digitalen Medien aber mit vielfältigeren Übergriffen konfrontiert wie z.B. Beleidigungen, Belästigungen, Hasspostings auf Social Media Profilen, Erpressungen durch die angedrohte Weitergabe von intimen Bildern oder Videos, Einschüchterungen oder öffentliche Bloßstellung durch Falschmeldungen oder missbräuchliche Verwendung von persönlichen Social Media Profilen, ungewollte Konfrontationen mit Schockvideos oder pornographischen Inhalten bis hin zu physischen Gewalt- oder Tötungsandrohungen (Gradinger & Strohmeier 2013). Während es für Opfer solcher Übergriffe besonders belastend ist, wenn diese vor unbeteiligten Dritten (=Online-Bystander*in) stattfinden (Thomas et al. 2012), haben gerade diese Online-Bystander*innen ein enorm hohes Potenzial, als Counter Speaker*innen aufzutreten.

Forschungen zum Konfrontationsverhalten von Beobachter*innen rassistischer oder sexistischer Übergriffe in Offline-Kontexten zeigen, dass Bystander*innen eine hohe Wahrscheinlichkeit haben, bei Täter*innen Schuldgefühle hervorzurufen (Fazio & Hilden 2001; Czopp et al. 2006) und dass sie auch bei anderen Beobachter*innen die Veränderung von Vorurteilen bewirken können (Blanchard et al. 1994). Bystander*innen werden in Offline-Kontexten eher positiv wahrgenommen, wenn sie konfrontativ für andere eintreten (Dickter et al. 2012; Schultz & Maddox 2013). Zudem können besonders jene Bystander*innen, die nicht der kompromittierten Zielgruppe angehören, besonders effektiv in der Konfrontation sein, weil sie kein „berechtigtes“ Interesse haben, das Vorurteil zu reduzieren, sondern durch ihre Intervention für Täter*innen eher überraschend sind (Czopp & Monteith 2003). Bystander*innen in ihrer Funktion als unbeteiligte Beobachter*innen scheinen auch effektiver zu sein, um vorurteilsbehaftete Fehlannahmen von Täter*innen zu korrigieren (Rasinski & Czopp 2010). Obwohl diese Erkenntnisse das hohe normverändernde und deeskalierende Potenzial von Bystander*innen

unterstreichen (Weber und Dickter 2015), ist noch weitgehend unklar, inwieweit solche Effekte auch dann zu finden sind, wenn Bystander*innen als Online Counter Speaker*innen in Online-Kontexten auftreten.

Die bisher vorliegende Forschung verweist auf eine Reihe an Gründen, warum Jugendliche in Online-Kontexten sogar noch seltener intervenieren als offline (Patterson et al. 2016):

- 1) Unsicherheiten in der Bewertung der Ernsthaftigkeit und Schwere von Online-Übergriffen: Gerade unter Jugendlichen werden Übergriffe in Online-Kontexten oft auch aus Spaß inszeniert, wobei Mädchen diese meist als ernster und verletzender wahrnehmen als Jungen und daher auch öfter intervenieren (Herry et al. 2021, Patterson et al. 2016a). Männliche Jugendliche intervenieren eher dann, wenn Kommentare besonders offensiv sind (Dickter & Newton 2013). Zusätzlich sind auch Altersunterschiede bemerkbar: Jüngere Jugendliche empfinden Online-Übergriffe schlimmer als ältere und haben entsprechend auch ein größeres Bedürfnis zu intervenieren (Herry et al. 2021).
- 2) Fehlende Kontextinformationen zur Bewertung der Situation: Gerade bei Online-Übergriffen ist die Kontextunsicherheit besonders hoch, meist ist schwer nachvollziehbar, was zu einem Übergriff geführt hat und ob sich das Opfer überhaupt betroffen fühlt (Thomas et al. 2012). Hinzu kommt die Asynchronität des Internet – es ist unklar, ob sich im Konfliktverlauf vielleicht schon etwas geändert hat bzw. herrscht das Gefühl, bei jedem Eingreifen „zu spät dran zu sein“, zumal online verbreitete Übergriffe nicht rückgängig gemacht werden können (Atzmüller et al. 2019 a,b).
- 3) Mangelnde Empathie gegenüber Online-Opfern: Empathiefähigkeit gilt als besonders wichtiger Faktor für zivilcouragiertes Handeln (Van Cleemput et al. 2014). Mädchen sind als Bystanderinnen offenbar empathischer als Jungen (Patterson, et al. 2016; Thornberg & Jungert 2013) und scheinen auf affektive Emotionen stärker aktiv zu reagieren (Dickter & Newton 2013). Die Empathiefähigkeit wird auch von eigenen Opfer- (Butler et al. 2021; Van Cleemput et al. 2014) und Täter*innenerfahrungen (Barlińska et al. 2013) beeinflusst. Personen werden durch solche Erfahrungen nicht nur sensibilisiert, sondern verfügen auch über ein spezielles Erfahrungswissen, was das Einnehmen einer Rolle als Counter Speaker*innen begünstigt (Butler & Fisher 2020).
- 4) Mangel an Bezügen zu Opfern: Mit dem Opfer befreundet zu sein erhöht die Interventionsbereitschaft deutlich (Patterson et al. 2016; Brody & Vangelisti 2015). In der Regel setzen sich jugendliche Bystander*innen für fremde Personen im Internet nur sehr selten und wenn möglich nicht öffentlich, sondern über die Privat-Chat-Funktion ein (Atzmüller et al. 2019 a,b). Sind Freund*innen betroffen, werden Offline-Interventionen bevorzugt. Damit fehlt allerdings ein wesentliches Fundament, um öffentlich sichtbar als Counter Speaker*innen aufzutreten. Entscheidend ist gerade unter jugendlichen Peers auch die Beziehung zu Täter*innen (Machackova et al. 2013), da ein Naheverhältnis zivilcouragiertes Eingreifen unwahrscheinlicher macht.
- 5) Unzureichende normative Werthaltungen gegenüber Online-Interventionen: Das Eintreten für andere erfordert auch eine entsprechende Überzeugung, dass ein solches Verhalten richtig ist (Machackova et al. 2013). Neben persönlichen Wertvorstellungen und moralischen Prinzipien kann der „übliche“ Umgang z.B. in der Schule oder Klasse (Gini et al. 2008) bzw. im jeweiligen sozialen

Netzwerk (Wächter et al. 2016) die Wahrscheinlichkeit des Eingreifens beeinflussen. Studien aus dem Offline-Bereich zeigen, dass Personen, die ihren Moralvorstellungen entsprechend stark intervenieren, zufriedener sind (Dickter & Newton 2013) – ein solcher Effekt tritt im Online-Kontext offenbar aber nicht ein: In einer bereits durchgeführten Studie des Forschungsteams (Atzmüller et al. 2019 a,b) bewerten die befragten Jugendlichen Online-Interventionen im Vergleich zu Offline-Interventionen als deutlich emotionsloser („es stellt sich kein gutes Gefühl ein“). Dieser Umstand erschwert die Mobilisierung zu aktivem Counter Speech Verhalten zusätzlich.

- 6) Mangelnde Unterstützung durch andere: Zivilcourage durch Jugendliche ist wahrscheinlicher, wenn sie das Gefühl haben, von der Peer Group unterstützt zu werden (Bastiaensens et al. 2014, 2016; Tarabulus et al. 2015). Je gefestigter und sicherer die eigene Position innerhalb des Freundeskreises wahrgenommen wird, desto leichter fällt es zu intervenieren (Levine & Crowther 2008). Unterstützungsleistungen werden aber insbesondere dann unwahrscheinlicher, wenn Übergriffe außerhalb des persönlichen Freundes-/Bekannteskreises stattfinden.
- 7) Mangelnde Bereitschaft zur Verantwortungsübernahme: Der große Anteil an potenziellen Unterstützer*innen in öffentlichen Kontexten kann sich sogar gegenteilig auswirken: Aus der Forschung im Bereich Zivilcourage ist bekannt, dass umso weniger eingegriffen wird, je höher die Anzahl an anderen Zuschauer*innen ist (Verantwortungsdiffusion). Nach Obermaier et al. (2016) gilt dieser Effekt auch im Internet. In diesem Zusammenhang ist auch die prinzipielle Zuschreibung der Handlungsverantwortung relevant: Butler et al. (2021) zeigen in ihrer Studie, dass jene Personen, die User*innen eine zentrale Rolle beim Eingreifen zuschreiben, auch selbst signifikant öfter Counter Speech betreiben, während ein Eingreifen weniger wahrscheinlich ist, wenn andere Instanzen wie die Polizei oder die Politik als Verantwortliche gesehen werden.
- 8) Hohes eigenes Interventionsrisiko: Jugendliche, die intervenieren, laufen Gefahr, als „Spielverderber*innen“ zu gelten, wenn z.B. Übergriffe als Spaß gemeint waren (Wagner et al. 2012). Aber auch bei ernstesten Situationen besteht die Gefahr, selbst zum Opfer zu werden. Herausfordernd ist dabei besonders, dass jede Art der öffentlichen Intervention permanent verfügbar bleibt (Slonje & Smith 2007). Dadurch werden Online-Interventionen auch Teil viral verbreiteter Narrative. Die Intervention kann zudem zeitunabhängig aktualisiert werden, sobald jemand die Kommunikation bemerkt – damit besteht das soziale Risiko nahezu unbegrenzt (Dillon & Bushman 2015).
- 9) Antizipierte Wirkungslosigkeit von Online-Interventionen: In einer mehrdimensionalen Studie (DeSmet et al. 2016) wurde als einer der bedeutendsten Prädiktoren für positives Online-Bystander*innenverhalten die positive Absicht von Bystander*innen eruiert, die wiederum am stärksten durch positive Ausgangsbewertungen der eigenen Handlung beeinflusst wird. In einer Studie des Forschungsteams (Atzmüller et al. 2019 a,b) gaben die befragten Jugendlichen an, dass zur Verfügung stehende Online-Optionen im Vergleich zu Offline-Handlungen einerseits als limitierend („ist ja nur Schreiben“) und andererseits aufgrund erforderlicher längerer Interaktionen und Diskussionen als anstrengend empfunden werden („da muss man so viel schreiben, dazu bin ich zu faul“). Auffallend ist auch, dass selbst einfache, indirekte Möglichkeiten ohne Risiko, wie etwa die

Nutzung der Meldefunktion in Sozialen Netzwerken, offenbar nur wenig genutzt werden. Insbesondere bei unbekanntenen Personen werden Online-Interventionen aber als wirkungslos wahrgenommen („*Ich würde mir auch nichts von Fremden sagen lassen*“) – als besser wird die Strategie erachtet, den Angreifer*innen keine Aufmerksamkeit schenken („*don't feed the troll*“-Taktik).

- 10) Die Vorstellung von Zivilcourage ist nicht auf Online Umgebungen übertragbar: Erkenntnisse des Projektteams aus bisherigen Studien (Atzmüller et al. 2019 a,b) verweisen auf ein weiteres zentrales Problem: Während Zivilcourage im alltäglichen Verständnis mit mutigem oder sogar heldenhaftem Einsatz in Verbindung gebracht wird, wird Online Zivilcourage nicht nur als wirkungslos, sondern auch als nicht besonders couragiert betrachtet. Entsprechend wird auch keine besondere Anerkennung als “Online Held*in” erwartet.
- 11) Darüber hinaus wird deutlich, dass es Jugendlichen aus allen Bildungsschichten massiv an Kompetenzen fehlt, wie Online Zivilcourage als aktive Form des sichtbaren Widerstandes gegen Cybermobbing, Online Hass und andere Formen digitaler Gewalt erfolgreich gestaltet werden kann.

Die Ausführungen zeigen, dass die Mobilisierung Jugendlicher zu aktiven Counter Speaker*innen vor großen Herausforderungen steht. Damit das gelingen kann, ist ein Zusammenspiel zwischen Sensibilisierungsmaßnahmen und Unterstützungsleistungen zur aktiven Umsetzung erforderlich: Sensibilisierung kann in Form von breit angelegten Online- und Offline-Kampagnen erfolgen. Ziel kann es z.B. sein, (1) das Normbewusstsein und die Empathiefähigkeit beim Beobachten von Online-Übergriffen zu stärken und zu betonen, dass es auch bei fremden Personen wichtig ist, aktiv aufzutreten. Wichtig erscheint es aber auch zu vermitteln, dass (2) Counter Speech tatsächlich wirkungsvoll sein kann. Ein solches, unter Jugendlichen noch nicht verankertes Bewusstsein sollte nicht nur bei Täter*innen, sondern auch bei anderen Internet-Nutzer*innen gefördert werden.

Trotz Bemühungen, Counter Speech Vorschläge für Jugendliche zu sammeln (z.B. konkrete Counter Speech Angebote in Form von Memes, Videos und Sprüchen, die in Deutschland im Rahmen der No Hate Speech Kampagne gesammelt wurden: <https://no-hate-speech.de/de/kontern/>), fehlt es an gezielter Forschung, wie Jugendliche darin unterstützt werden können, ihr Auftreten als Counter Speaker*innen konkret zu gestalten – und zwar mit Berücksichtigung der bekannten hemmenden Faktoren.

Um das Wirkungspotenzial von kampagnenorientierten Counter Speech Maßnahmen zu optimieren, gibt es zahlreiche Initiativen (z.B. Institute for Strategic Dialogue², Anti-Defamation League³, Demos⁴), die sich mit der Erarbeitung von Gestaltungsvorschlägen beschäftigen (z.B. Brown 2016; Frenett & Dow 2015), um zum Beispiel Personen zu erreichen, die negative Inhalte verbreiten. Weitgehend unberücksichtigt bleibt

² <https://www.isdglobal.org/>

³ <https://www.adl.org/>

⁴ <https://www.demos.co.uk/>

jedoch die zielgruppengerechte Gestaltung von Counter Speech Angeboten für jene jungen Menschen, die zur aktiven Gegenrede oder zumindest zur Verbreitung von Counter Speech Botschaften mobilisiert werden sollen, wobei hier sowohl geschlechterspezifische Unterschiede, aber auch Unterschiede aufgrund unterschiedlicher Bildung und sozialer Herkunft in der Gestaltung des spontanen Antwortverhaltens berücksichtigt werden müssen. Dazu ist es erforderlich zu wissen, (1) wie sich das Interaktionsverhalten von Jugendlichen in solchen Fällen typischerweise gestaltet, (2) welche Kommunikations-Elemente sich aus Sicht von Mädchen und Jungen als praktikabel und wirkungsvoll erweisen und (3) wie diese bei der konkreten Gestaltung von Counter Speech optimiert werden können. Entsprechende Studien fehlen bislang.

Bisher gibt es nur wenige Studien, die gezielt das spontane Counter Speech Verhalten von Internet-Nutzer*innen als Reaktion auf hasserfüllte und extremistische Inhalte untersucht haben; diese sind allerdings nicht auf Jugendliche bezogen: So haben Benesch et al. (2016a, 2016b) eine umfangreiche Untersuchung unter Twitter-User*innen durchgeführt und identifizierten folgende Counter Speech Strategien im spontanen Antwortverhalten:

- Warnung vor Konsequenzen: zum Beispiel Sanktionierung durch (zukünftige) Arbeitgeber*innen, durch die Familie oder Netzwerkbetreiber*innen (Sperrung des Accounts)
- Shaming und Labeling: eine klare und öffentliche Deklaration als z.B. rassistisch, frauenfeindlich etc. kann dazu führen, dass sich Absender*innen nicht dadurch stigmatisieren lassen möchten. Gleichzeitig kann eine solche Deklaration andere Beobachter*innen dabei unterstützen, solche Inhalte entsprechend zu interpretieren und darauf zu reagieren.
- Mitgefühl und Zugehörigkeit: ein z.B. gezielt freundlicher, empathischer und friedlicher Tonfall kann weitere Eskalation vermeiden. Aber auch der Ausdruck eines Zugehörigkeitsgefühls (z.B. „*Ich bin auch Anhänger von x, aber...*“) kann Empathie erzeugen und dabei unterstützen, hasserfüllte Interaktionen in eine zivilisierte Form zu transformieren.
- Humor: Hier scheint ein besonderes Potenzial verankert zu sein, da Humor in vielfältigsten Variationen (z.B. Karikatur, Sarkasmus) die Dynamik einer Kommunikation verändern und damit deeskalierend wirken kann und gleichzeitig die Aufmerksamkeit anderer deutlich erhöht.
- Bilder: Memes, Grafiken, Fotos, animierte GIFs oder Videos sind oft überzeugender als Text alleine. Darüber hinaus können so auch kulturelle und sprachliche Barrieren überwunden werden, was die breite virale Verbreitung erleichtert.

Identifiziert wurden auch folgende kontraproduktive Strategien im spontanen Antwortverhalten auf Twitter (Benesch et al. 2016b):

- Feindlicher oder aggressiver Ton, Beschimpfungen: Solche häufig vorkommenden Strategien sind oft mit Obszönitäten, Verunglimpfungen oder abfälligen Bemerkungen verbunden: Dabei besteht die Gefahr einer weiteren Eskalation, wodurch potenzielle andere Counter Speaker*innen davon abgehalten werden, ebenfalls zu intervenieren.

- **Faktencheck:** Obwohl eine solche Strategie oft empfohlen wird, scheint sie eher ineffektiv zu sein: Absender*innen von negativen Inhalten tendieren dazu, neue Fakten so zu interpretieren, wie sie in ihr eigenes Denkschema passen oder finden einen anderen Nachweis, um die eigene Meinung zu stützen.
- **Harrassment und Silencing:** Dies sind Taktiken, welche zum Beispiel durch eine Flut von oft aggressiven Antworten oder durch Androhung von Offline-Gewalt Absender*innen zum Stillschweigen gebracht werden sollen. Ein solches Vorgehen kommt besonders bei offensichtlich hasserfüllten Inhalten vor, die viral verbreitet werden.

In einer jüngeren Studie nahmen Mathew et al. (2019) die Counter Speech Kategorisierung von Benesch (2016 a,b) auf und analysierten ebenfalls auf Basis von Twitter-Daten in drei verschiedenen Communities (LGBTQI+, jüdischen Community, People of Color Community) welche spezifischen Counter Speech Strategien auf Unterstützung durch andere, (erwachsene) User*innen stoßen. Dabei zeigten sich große Unterschiede: in der LGBTQI+ Community erhielten humorvollen Strategien die meiste Unterstützung, in der jüdischen Community wurde die Warnung von Konsequenzen am stärksten unterstützt und in der People of Color Community fanden feindliche Gegenreaktionen den größten Zuspruch. Damit wird deutlich, dass produktive und kontraproduktive Strategien mit Blick auf die Mobilisierung anderer Bystander*innen je nach Kontext sehr unterschiedlich sein können. Ähnliche Studien mit der Zielgruppe Jugendliche liegen bislang noch nicht vor.

Die vorliegende Studie greift daher mit dem expliziten Fokus auf die Mobilisierung Jugendlicher zu aktiver Counter Speech einen notwendigen Forschungsbedarf auf und vermittelt einen umfassenden Einblick in das typische Interaktionsverhalten Jugendlicher, wenn diese auf negative Inhalte im Internet reagieren. Gleichzeitig wird nach Optimierungspotenzialen gesucht, um sowohl die Effizienz als auch die Attraktivität von Counter Speech durch Jugendliche zu erhöhen. Im nachfolgenden Kapitel werden die konkreten Zielsetzungen der Studie nochmals zusammengefasst.

Zielsetzung und Forschungsfragen

Die vorliegende Studie hat sich zum Ziel gesetzt, die Bedingungen und Ausprägungen jugendlichen Counter Speech Verhaltens zu analysieren und zu verstehen, sowie gemeinsam mit Jugendlichen Möglichkeiten für Counter Speech zu entwickeln und zu optimieren. Damit sollen Grundlagen geschaffen werden, um 14- bis 18-jährige Mädchen und Jungen in ihrem aktiven Auftreten als Online Counter Speaker*innen zu unterstützen und zu fördern. Wir thematisieren Counter Speech als eine Form zivilcouragierten Verhaltens, sowie als ein klares und nachhaltiges humanistisches Zeichen gegen Online Hass bzw. jede andere Form hasserfüllter Rhetorik und digitaler Gewalt.

Dazu bearbeiten wir folgenden zentralen **Forschungsfragen**:

- Wie gestaltet sich typischerweise das Interaktionsverhalten von Jugendlichen, wenn diese auf negative Inhalte im Internet reagieren?
- Welche unterschiedlichen Counter Speech Strategien sind auffindbar und inwieweit lassen sich diese als produktive oder kontraproduktive Strategien klassifizieren?
- Welche Formen der aktiven Gegenrede erscheinen aus Sicht von Jugendlichen praktikabel und wirkungsvoll bzw. welche Optimierungspotenziale lassen aus dem Counter Speech Verhalten von Jugendlichen ableiten?

Um Antworten auf diese Forschungsfragen zu finden, wurde erstmals eine umfassende empirische Datenbasis zum Counter Speech Verhalten Jugendlicher unter Einsatz innovativen Methoden (Online Rollenspiele, Inhaltsanalyse von YouTube-Daten, Online Feld-Experiment, Online-Aktionsforschung mit Jugendlichen) geschaffen. Der vorliegende Bericht gibt einen tiefgehenden Einblick in die methodische Umsetzung und die gewonnenen wissenschaftlichen Erkenntnisse.

Praktisches Projektziel war es, aufbauend auf den wissenschaftlichen Erkenntnissen jugendadäquate Unterstützungsangebote zur Mobilisierung Jugendlicher zu Counter Speech im Umgang mit negativen Inhalten im Internet auszuarbeiten: In intensiver Zusammenarbeit mit den beteiligten Praxispartner*innen ÖIAT/Saferinternet.at, Mauthausen Komitee Österreich, Büro 1.6 Kriminalprävention und Opferhilfe des Bundeskriminalamts/BM.I und Bundeskanzleramt/Referat VI/5/a Kompetenzzentrum Jugend wurde die gemeinsame Expertise mit Fokus auf die Mobilisierung Jugendlicher zu Counter Speech weiter ausgebaut und die gewonnenen Erkenntnisse genutzt, um sowohl bestehende Schulungs- und Trainingsmaßnahmen im Bereich Online Zivilcourage zu verbessern, als auch um neue Unterstützungsideen zur Förderung von Counter Speech durch Jugendliche zu entwickeln (siehe dazu Kapitel „Unterstützungsmaßnahmen der Praxispartner*innen“ in diesem Bericht)

Überblick über das methodische Gesamtdesign

Ziel der empirischen Untersuchung war es, zu erfassen, 1) wie sich das Interaktionsverhalten von Jugendlichen typischerweise gestaltet, wenn diese auf negative Inhalte im Internet reagieren, (2) welche Kommunikations-Elemente sich aus Sicht von Mädchen und Jungen als praktikabel und wirkungsvoll erweisen und (3) wie diese bei der konkreten Gestaltung von Counter Speech optimiert werden können.

Zielgruppe: 14- bis 18-Jährige: in diesem Alter finden wesentliche persönlichkeitsbildende Erfahrungen und Entwicklungsaufgaben statt, sodass Counter Speech fördernde Maßnahmen gut anknüpfen können: z.B. Entwicklung eines eigenen Werte- und Normensystems sowie eines ethischen und politischen Bewusstseins, Identitätsentwicklung, Verstärkung von Peer-Beziehungen (Lange et al. 2018), oder aus juristischer Sicht das Erreichen der Strafmündigkeit und damit einhergehend eine Sensibilisierung für strafrechtlich relevante Normverletzungen. Die Datenerhebungen fanden in Wien bzw. digital statt.

Das **Studiendesign** umfasste fünf aufeinander folgende Schritte:

1. **Online Rollenspiele:** Ziel war es, sowohl das spontane Antwortverhalten von Jugendlichen in der konkreten Online-Interaktion zu erfassen als auch Wirkungspotenziale von jugendlichen Kommunikationsstrategien aus unterschiedlichen Perspektiven (aus Sicht von Opfern, Täter*innen, Counter Speaker*innen und anderen Online-Bystander*innen) zugänglich zu machen, indem typische Settings mit Hilfe von Social Media Fake-Profilen simuliert wurden. Dabei wurden nicht nur öffentlich sichtbare, sondern auch typischerweise über die Privat-Chat Funktion stattfindende Interaktionen berücksichtigt.
2. **Inhaltsanalyse** von jugendlichen User*innen-Kommentaren auf ausgewählten YouTube Kanälen: Mit dieser Methode wurde das spontane, „alltägliche“ und öffentlich sichtbare Antwortverhalten Jugendlicher im Umgang mit negativen Inhalten in realen Settings und auf einer breiteren Ebene untersucht. Zusätzlich wurden wichtige Erkenntnisse in Bezug auf die Identifizierung und Klassifizierung von Hate- und Counter Speech Interaktionen gewonnen, die auch für weiterführende Studien (z.B. automatisierte Counter Speech Detection) eingesetzt werden können.
3. **Online Aktionsforschung zur Erarbeitung von Counter Speech Vorschlägen:** In enger Zusammenarbeit zwischen Jugendlichen, Forscher*innen und Praxispartner*innen wurden im Rahmen einer Online Aktionsforschung unterschiedliche Gestaltungsmöglichkeiten jugendgerechter Counter Speech erarbeitet und diskutiert und im konkreten Einsatz erprobt und dabei auch die Nutzbarkeit bereits bestehender Counter Speech Angebote (z.B. „Konter-Vorschläge“ im Zuge der No

Hate Speech Kampagne) berücksichtigt. Dieser Teil wurde durch eine Zusatzfinanzierung der Stadt Wien (MA 7)⁵ unterstützt.

4. **Online Feld-Experiment:** Um zu prüfen, was mit jugendgerechter Counter Speech tatsächlich bewirkt werden kann, wurde Counter Speech durch Jugendliche im Rahmen eines Online Feld-Experiments in Zusammenarbeit mit einem bekannten YouTuber im konkreten Einsatz geprüft. Dabei wurde gezielt sowohl Hate Speech als auch Counter Speech in abwechselnden Phasen durch das Forschungsteam bzw. durch involvierte Jugendliche inszeniert. Die realen Reaktionen der Online Community wurden analysiert.
5. **Entwicklung von Unterstützungsangeboten zur Förderung von Counter Speech:** Aus der Summe an Forschungsergebnissen wurden schließlich in Zusammenarbeit mit den Praxispartner*innen zentrale Erkenntnisse zur Förderung von Counter Speech sowie zu deren jugendgerechter Gestaltung abgeleitet und konkrete Unterstützungsangebote entwickelt, die im Rahmen von Informations-, Schulungs- und Trainingsmaßnahmen sowie durch unterschiedliche Informationskanäle verbreitet werden.

In den folgenden Ausführungen werden die einzelnen methodischen Herangehensweisen ausführlich dargestellt und die jeweiligen Ergebnisse präsentiert.

⁵ MAKING Cyber Heroes (2021–2022), gefördert von der Stadt Wien (MA 7, Kulturabteilung der Stadt Wien, im Rahmen des Calls *Digitaler Humanismus*)

1. Online Rollenspiele zur Erfassung des typischen Interaktionsverhaltens Jugendlicher

Ziel der Online Rollenspiele war es, spontanes Counter Speech Verhalten für verschiedene Kontexte in Sozialen Medien in einem kontrollierten Setting zu simulieren. Dabei sollten nicht nur öffentlich sichtbare, sondern auch über die Privat-Chat Funktion stattfindende Interaktionen berücksichtigt werden.

Erfasst werden sollte:

- (1) wie sich das (spontane) Interaktionsverhalten von Jugendlichen typischerweise gestaltet, wenn diese mit Online-Übergriffe konfrontiert sind und
- (2) welche Kommunikations-Elemente und Counter Speech Strategien aus Sicht von Mädchen und Jungen in ihrer Rolle als Online-Bystander*innen praktikabel und wirkungsvoll sind.

1.1 Methodisches Design

Die Durchführung von Online Rollenspielen ist methodisch innovativ und wurde auch im internationalen Umfeld bisher nur selten eingesetzt. Somit standen keine Standardverfahren oder -abläufe zur Verfügung, und die Konzeption der gesamten methodischen Umsetzung wurde sehr detailliert im Rahmen dieses Forschungsprojekts ausgearbeitet. Diese Konzeption umfasste folgende zentrale Schritte, die im Folgenden präsentiert werden: Vorbereitung von Smartphones als Erhebungsinstrumente, Entwicklung impulsgebender Szenarien und Rollenvorgaben, sowie Prüfung der forschungsethischen Angemessenheit.

1.1.1 Smartphones als Erhebungsinstrumente

Die Umsetzung der Online-Rollenspiele erfolgte mittels Endgeräten, die vom Forschungsteam bereitgestellt wurden und ausschließlich im Rahmen der Online Rollenspiele eingesetzt wurden. Damit wurde sichergestellt, dass keine persönlichen Informationen vom eigenen Handy der Jugendlichen ins Spiel gebracht wurden, dass während der Rollenspiele keine anderen, realen Online Kontakte involviert wurden und dass Zugänge zu den Accounts nicht unerwünschterweise nach den Rollenspielen weiterverwendet wurden. Zu diesem Zweck wurden insgesamt 9 Smartphones angekauft. 1 Smartphone wurde vom Forschungsteam genutzt (passive Beobachter*innen-Rolle), 8 Smartphones konnten von den Jugendlichen in ihren unterschiedlichen Rollen genutzt werden. Die Smartphones wurden mit entsprechenden Nummern gekennzeichnet.

Auf allen Smartphones wurden WhatsApp, Instagram und Snapchat eingerichtet, also jene Social Media Plattformen, die von Jugendlichen besonders gern genutzt werden. Für jedes Smartphone wurde

außerdem ein Rollenspiel-Profil (=Fake-Account) auf WhatsApp, Snapchat und Instagram vorbereitet (passwortgesichert), das von den Jugendlichen im Spiel genutzt werden konnte. Damit wurde es für die jugendlichen Rollenspieltelnehmer*innen möglich, sich von der eigenen Person (emotional und auf der Ebene der persönlichen Betroffenheit) zu distanzieren und über ihre zugeordnete Rolle z.B. als Opfer oder als Täter*in zu agieren. Die vorbereitende Profilgestaltung wurde von Jugendlichen umgesetzt (in Zusammenarbeit mit einer Handelsakademie im 10. Wiener Gemeindebezirk), die selbst nicht an den Rollenspielen teilnahmen. Dazu wurden vorab von ÖIAT/Saferinternet die erforderlichen Rohprofile und Zugangsdaten (z.B. geschlechtsneutrale Nicknames, Email-Adresse) vorbereitet. Aufgabe der Jugendlichen war es, ein typisches, möglichst authentisches Profil einer jungen Person anzulegen, wobei folgende Regeln beachtet werden sollten: (1) keine Verwendung von persönlichen Fotos oder anderen persönlichen Informationen der Jugendlichen selbst, lediglich lizenzfreie Fotos aus dem Internet. (2) Keine Kontakte zu realen Personen. Kontakte zwischen den Fake-Profilen waren aber erlaubt; ebenso durfte bekannten Influencer*innen und Celebrities gefolgt werden mit der Vorgabe, dass diese ein blaues Verifizierungshäkchen haben müssen⁶. Die Jugendlichen wurden außerdem explizit darauf hingewiesen, dass die erstellten Profile in der Folge von anderen Jugendlichen verwendet werden und daher weder eigene persönliche Daten noch die persönlichen Daten Bekannter verwendet werden dürfen. Die Gestaltung der Profile wurde in Zweier-Teams umgesetzt, eine anwesende Saferinternet-Expertin stand für Fragen oder weiteren Unterstützungsbedarf zur Verfügung. Die erstellten Profile wurden anschließend besprochen und fallweise auch von anderen aus der Gruppe hinsichtlich ihrer Authentizität kritisch geprüft (z.B. ob das Foto zu „professionell“ für Jugendliche wirkt oder als glaubwürdig gelten kann). Die Profile wurden nach Fertigstellung von der ÖIAT/Saferinternet-Expertin geprüft und in Einzelfällen bereinigt (z.B. wurden Politiker*innen entfolgt).

Auf diese Weise wurden auf Instagram sowohl Mädchen-, als auch Burschen-Profile eingerichtet, zusätzlich wurden auch geschlechterneutrale Profile erstellt. Aufgrund der schwierigen Handhabung bezüglich der wechselnden Gruppenzusammensetzung (an den Workshops nahmen in der Regel unterschiedlich viele Mädchen oder Burschen teil), wurde nach den ersten Rollenspielen auf ausschließlich geschlechterneutrale Profile umgestellt. Die WhatsApp Profile wurden von vornherein geschlechterneutral ausgerichtet, vor allem auch, weil pro Handy nur ein WhatsApp Account installiert werden konnte. Auch auf Snapchat gab es pro Handy jeweils nur einen Account, der mit Hilfe von Bitmojis gut geschlechtsneutral gestaltet werden konnte. Um den Wechsel zwischen den Profilen zu vereinfachen, wurden vom Forschungsteam geschlechtsneutrale Nicknames bereits bei der Einrichtung der Rohprofile fix vorgegeben. Die geschlechtsneutrale Gestaltung der Profile ermöglichte damit eine einfache Zuordnung der Smartphones bei der Umsetzung der Workshops.

⁶ Instagram bestätigt damit die Echtheit eines Kanals. In der Regel erhalten Profile mit großer Abonnent*innenzahl wie z.B. Prominente, bekannte Influencer oder große Marken solche Häkchen.

Für die nun neun vorbereiteten Smartphones samt Social Media Profilen und Nicknames wurden eine eigene Gruppe angelegt („Dönerr-Gruppe“⁷). Die Rollenspiele fanden ausschließlich innerhalb dieser Gruppe statt. Nach jedem Online-Rollenspiel-Workshop wurde darauf geachtet, sowohl die Daten als auch die Fake-Profilen sorgfältig zu sichern (diese wurden von den Jugendlichen am Anfang jedes Spieles adaptiert) bzw. die im Spiel entstandenen Interaktionen wieder als Vorbereitung für das nächste Spiel zu bereinigen.

Besonders herausfordernd waren Überlegungen zur Sicherung des Datenmaterials: Zentrales Anliegen war es, den gesamten Interaktionsverlauf zu dokumentieren und auch den Einsatz visueller und akustischer Kommunikationselemente (z.B. GIFs, Memes, aber auch Sprachnachrichten) zu dokumentieren. Als Lösung wurde daher auf jedem Smartphone ein Screencast-Tool installiert, welches das gesamte Spiel in Form eines Videos mitfilmte. Diese Vorgehensweise hatte den entscheidenden Vorteil, dass eine besonders reichhaltige und qualitativ hochwertige Datenbasis geschaffen werden konnte, welche exakt die Perspektive der Jugendlichen selbst wiedergibt. Berücksichtigt werden konnten auch nicht abgeschlossene Interaktionen (z.B., wenn Kommentare zwar begonnen, dann aber doch nicht gepostet wurden), interaktive, visuelle und akustische Kommunikationsformen (z.B. GIFs, Sprachnachrichten, Privat-Chats), nicht öffentliche Interaktionen in der Privat-Chat-Funktion, oder die genaue Darstellung, wann ein Wechsel zwischen unterschiedlichen Plattformen vorgenommen bzw. Interaktionen auf unterschiedlichen Plattformen umgesetzt wurden. Zusätzlich wurden die Daten auch als Screenshots gesichert bzw. verfügbare Download-Funktionen genutzt.

1.1.2 Konzeption der Umsetzung: Impulsgebende Szenarien und Rollenvorgaben

Für die Durchführung wurden von den Forscher*innen in Zusammenarbeit mit den Praxispartner*innen impulsgebende Szenarien vorbereitet, bei denen Online-Zivilcourage gefragt ist bzw. Counter Speech erwartet werden kann. Die konkrete szenische und dramaturgische Umsetzung oblag jeweils den jugendlichen Rollenspiel-Teilnehmer*innen selbst, um auch individuelle und jugendtypische Ausdrucksweisen zu gewährleisten. Bildhaft dargestellte Übergriffe wurden vom Forschungsteam vorab vorbereitet, um einerseits extrem unangemessene Bilder zu vermeiden (z.B. pornographisches Bildmaterial), andererseits aber auch, um zu „harmlose“ Settings zu umgehen (z.B. Darstellung einer Person mit minimalen, eher „kindlichen“ Veränderungen wie z.B. Mädchen mit Schnurrbart). Sowohl die Auswahl der impulsgebenden Szenarien als auch das vorbereitete Bildmaterial stammt aus einer bereits durchgeführten Studie des Forschungsteams (Studie Zivilcourage 2.0: Zivilcourage von Jugendlichen im Umgang mit wahrgenommener Gewalt im Internet, KIRAS Sicherheitsforschung; Laufzeit: Februar 2017 bis

⁷ Der Name sollte möglichst unverfänglich gewählt werden und ergab sich aus der Tatsache, dass einige der Jugendlichen, welche die Fake-Profilen erstellten, Bilder von Essen posteten.

Mai 2019) und wurde bereits auf Angemessenheit und Realitätsnähe für die untersuchte Altersgruppe der Jugendlichen getestet. Alle teilnehmenden Jugendlichen wurden über die Zielsetzung der Studie und die Motivation für die Durchführung der Rollenspiele informiert. Danach erhielten sie die impulsgebenden Szenen, welche ihre Rolle im Spiel genauer darstellte.

Folgende impulsgebenden Szenen wurden eingesetzt:

Szene 1: Jemand wird schwer beschimpft

Relevante Ausschnitte aus der Anweisung für die Rolle von Täter*innen:

„Denk dir bitte eine wirklich arge, bössartige, widerwärtige Beleidigung aus und beschimpfe eine bestimmte Person öffentlich, sodass es auch alle anderen in der Gruppe sehen können. Mach dich zum Beispiel darüber lustig, was die Person postet, beleidige das Aussehen oder die Herkunft der Person, sag, dass du sie ekelhaft findest usw. – versuche, die Person richtig fertig zu machen.“

Szene 2: Von jemandem wird ein Nacktvideo/-bild verbreitet

Relevante Ausschnitte aus der Anweisung für die Rolle von Täter*innen:

„Du erhältst gleich ein Foto, das die Füße einer bestimmten Person zeigt, es könnte auch Teil eines Nacktbildes oder Nacktvideos sein. Poste dieses Bild öffentlich in der Gruppe und behaupte, dass das ein Ausschnitt eines Nacktvideos ist, das du von dieser Person hast. Schreibe noch etwas wirklich sehr Abwertendes, Gemeines und Bloßstellendes über die Person dazu – versuch, die Person damit richtig fertig zu machen.“

Folgendes Foto wurde vom Forschungsteam zu Spielbeginn an jenes Handy geschickt, das im Spiel die Täter*innenrolle übernahm:



Abbildung 1

Szene 3: Über jemanden werden falsche Inhalte verbreitet

Relevante Ausschnitte aus der Anweisung für die Rolle von Täter*innen:

„Du erhältst gleich ein Foto, das zwei Personen vermutlich unter Alkohol- und Drogeneinfluss zeigt. Poste dieses Bild öffentlich in der Gruppe und behaupte, dass auf dem Foto eine bestimmte Person, zu sehen ist. Stell die Person als ‚Absturzkind‘ bloß, indem du zum Beispiel beschreibst, wie peinlich sie sich im betrunkenen (Drogen-)Zustand benommen hat – versuch, die Person damit richtig fertig zu machen.“

Folgendes Foto wurde vom Forschungsteam zu Spielbeginn an jenes Handy geschickt, das im Spiel die Täter*innenrolle übernahm (Anmerkung: es wurde absichtlich vermieden Personen konkret darzustellen):



Abbildung 2

Für jedes der drei Szenarien gab es prinzipiell folgende Rollen: Opfer, Täter*in, Mit-Täter*in und fünf Bystander*innen-Rollen. Das neunte Handy wurde ausschließlich vom Forschungsteam verwendet, das ebenfalls im Rollenspiel in einer passiven Beobachter*innenrolle involviert war. Damit war es möglich, nicht nur den Spielverlauf zu verfolgen, sondern auch bei Bedarf sofort reagieren zu können. Jede einzelne Rolle wurde auf einem farbigen Kärtchen, das den Jugendlichen übergeben wurde, schriftlich dargestellt, wobei die verschiedenen Rollen nicht explizit z.B. als „Opfer-Rolle“ oder „Täter*innen-Rolle“ benannt wurden. Vielmehr erhielten die Jugendlichen eine kurze Rollenbeschreibung, was in der Rolle gespielt werden soll bzw. in welchem Verhältnis man zu den anderen Personen in der Gruppe steht (z.B. es handelt sich um die Gruppe eines Vereins, man kennt nicht alle Personen).

In den ersten durchgeführten Rollenspielen wurden die Vorgaben für die fünf Bystander*innen-Rollen offengehalten und folgendermaßen dargestellt:

„Du bist seit einiger Zeit einem Verein beigetreten und nun auch in dessen WhatsApp, Instagram und Snapchat-Gruppe. Du kennst einige, die auch in dieser Gruppe sind, aber nicht alle. Verhalte dich in der Gruppe zunächst so, wie du das auch im echten Leben tun würdest, z.B. kommentieren, posten, jemandem folgen, mit anderen in Kontakt treten usw. Warte ab, was in der Gruppe passiert! Handle so, wie du das auch in Wirklichkeit tun würdest!“

Im Zuge der Umsetzung stellte sich jedoch heraus, dass es den Jugendlichen ohne weitere Hinweise zum Teil schwer fiel zu reagieren. Es wurde daher entschieden, den Bystander*innen-Rollen ebenfalls eine bestimmte Ausrichtung zu geben. Dabei wurden bisherige Befunde zur Bystander*innen-Forschung berücksichtigt, auch um zu prüfen, wie diese Rollenerwartungen von den Jugendlichen erfüllt werden konnten. Folgende Ausrichtungen wurden vorgegeben:

1. dem Opfer helfen
2. versuchen, Täter*in zu stoppen
3. die Situation insgesamt deeskalieren
4. andere Bystander*innen mobilisieren
5. offene Rolle ohne Vorgabe (wie bisher)

Mit dieser inhaltlichen Ausrichtung wurden die oben beschriebenen Rollenanweisungen entsprechend erweitert, hier beispielhaft gezeigt für die Bystander*innen-Rolle „dem Opfer helfen“:

„Warte ab, was in der Gruppe passiert! Jemand wird eine Person aus der Gruppe sehr gemein angehen. Versuch, dieser Person zu helfen oder ihr beizustehen. Handle so, wie du das auch in Wirklichkeit tun würdest!“

Innerhalb eines Online Rollenspiel Workshops (Gesamtdauer eines Workshops ca. 90-120 min) wurden in der Regel alle drei unterschiedlichen Szenarien gespielt. Eine Wiederholung eines Szenarios mit anderer Rollenbesetzung war prinzipiell nicht vorgesehen (die Rollenerfahrungen und -erwartungen wurden vorzugsweise in der anschließenden Diskussion besprochen).

Unmittelbar vor der konkreten Zuteilung der Rollen wurden die Teilnehmer*innen gebeten, sich mit dem zugeteilten Smartphone bzw. dem vorbereiteten Profil auseinanderzusetzen und diese je nach Wunsch (geringfügig) zu adaptieren. Die Jugendlichen behielten die zugeteilten Smartphones samt Profil während des gesamten Online Rollenspiel Workshops, bei dem mehrere Spiele gespielt wurden. Bei den konkreten Spielen konnte diese dabei aber unterschiedliche Rollen einnehmen (z.B. Opfer in Spiel 1, Bystander*innen Spiel 2 usw.).

Das Prozedere der Rollenzuteilung erfolgte durch zufälliges Ziehen einer Karte, auf der die Rollenbeschreibung zu lesen war. Dazu wurden farbige Karten genutzt (z.B. grün für Opfer-Rolle, gelb für offene Bystander*innen-Rolle usw.). Dies erlaubte einen Überblick über die Zuordnung zu ermöglichen eine leichtere Dokumentation der zugeordneten Rollen. Unmittelbar nach dem Ziehen der Karte wurden die Spielteilnehmer*innen gefragt, ob sie diese Rolle übernehmen möchten oder nicht. Falls nicht, konnten Rollen getauscht werden. Bei Bedarf hätten nicht vergebene oder ungewollte Rollen auch vom Forschungsteam übernommen werden können. Dies war allerdings nicht der Fall.

Um zu dokumentieren, wer welche Rolle im Spiel übernahm, wurde ein Dokumentationsbogen vorbereitet, wo die jeweilige Rolle (je nach Farbe) eingetragen wurde. Zusätzlich wurden hier auch relevante soziodemographische Merkmale (z.B. Geschlecht, Alter, Bildung, Migrationshintergrund) festgehalten. Bei jedem neuen Spiel wurde erneut eine Rolle gezogen mit der Aufforderung, eine andere Farbe als bisher zu ziehen – damit sollte der Personenwechsel gewährleistet sein. Die neuen Rollen wurden dann ebenfalls wieder im Dokumentationsbogen eingetragen.

Die Rollenspiele fanden an Schulen statt. Bei der Durchführung des Online Rollenspiels befanden sich die Teilnehmer*innen physisch in einem vom Forschungsteam begleiteten und kontrollierten Setting (z.B. im selben oder in nebeneinanderliegenden Räumen – je nach örtlicher Gegebenheit). Damit war es möglich, nach jedem Spiel das Spielgeschehen gemeinsam zu reflektieren und das Rollenerleben darzustellen und zu bewerten.

In einer abschließenden Diskussion innerhalb der Gruppe wurden die verschiedenen rollenspezifischen Erfahrungen der Workshop-Teilnehmer*innen reflektiert und in Bezug auf mögliche Anwendungssituationen im Internet bewertet (siehe dazu ausführlich Kapitel „Erkenntnisse aus den Diskussionsrunden“). Ziel war es, jene Elemente zu identifizieren, die aus Sicht der Spielenden in der Interaktion als förderlich oder hinderlich bzw. als wirkungsvoll oder nicht wirkungsvoll erlebt wurden. Dabei wurden auch vorhandene Handlungs- und Verhaltensmuster verdeutlicht und hinterfragt. Gleichzeitig konnten so auch die persönlichen Handlungsrepertoires erweitert werden.

1.1.3 Prüfung auf ethische Angemessenheit

Das methodische Prozedere der Online Rollenspiele wurde vorab der Ethikkommission der Universität Wien vorgelegt und mit einem positiven Votum beurteilt. Damit wurde gewährleistet, dass das Vorhaben hinsichtlich der Angemessenheit für diese Altersgruppe und bezüglich möglicher ethischer Problemfelder abgesichert ist.

Im Zuge dieses Prüfverfahrens wurde das gesamte methodische Prozedere in Bezug auf ethische und datenschutzrechtliche Anforderungen detailliert verfeinert. So war es eine wesentliche Bedingung, sowohl für die Erhaltung der persönlichen Integrität der Rollenspieler*innen zu sorgen als auch eine Entlastung auf der Ebene der persönlichen Betroffenheit anzubieten. Aus Sicht des Forschungsteams hätte sich etwa die Einnahme der Täter*innen-Rolle als besonders schwierig gestalten und ethische Probleme beinhalten können: Jugendliche Online-Rollenspieler*innen in der Täter*innenrolle hätten durch die Spieldynamiken zu Online Interaktionen motiviert werden können, die ihnen im Alltag zutiefst widersprechen, wodurch ihre persönliche Integrität Schaden nehmen hätte könnte. Das wäre insbesondere dann belastend gewesen, wenn diese mit Online-Rollenspiel-Teilnehmer*innen befreundet waren, was bei den geplanten Workshops sehr wahrscheinlich war. Aus diesem Grund wurde die Möglichkeit geschaffen, die Täter*innenrolle (oder alternativ eine andere schwierige Rolle) abzukoppeln, indem diese von einer jungen Person aus dem Umfeld von Saferinternet.at übernommen wurde (vorgesehen und dafür vorbereitet wurde eine 18-jährige Mitarbeiterin von Saferinternet.at; diese befand sich in einem örtlich getrennten Setting und war den Online Rollenspiel-Teilnehmer*innen nicht bekannt). Allerdings wurde diese Möglichkeit von den teilnehmenden Jugendlichen nicht in Anspruch genommen. Als weitere Entlastungsmaßnahme wurde die Täter*innen Rolle als Teamhandlung (z.B. es gibt sowohl Täter*in als auch Mittäter*in) umgesetzt.

Weiters wurden im Zuge der Einreichung bei der Ethikkommission auch entsprechende Informationsmaterialien zur Untersuchung sowie Einverständniserklärungen für Jugendliche und beteiligte schulische Einrichtungen ausgearbeitet, die dann bei der Umsetzung der Online Rollenspiele eingesetzt wurden. Zusätzlich wurde den beteiligten Jugendlichen Informationsmaterial der Praxis-

Partner*innen (v.a. ÖIAT-Saferinternet.at) mit relevanten Hinweisen auf Beratungsstellen und Verhaltensvorschläge bei Cybermobbing mitgegeben. Dies war auch relevant für den Fall, dass die beteiligten Mädchen und Jungen durch die Teilnahme für das Thema sensibilisiert wurden oder ev. vergangene Opfer- oder Täter*innenerfahrungen wachgerufen wurden.

Das methodische Vorgehen grenzt sich von ähnlichen durchgeführten Studien (z.B. Weber & Dickter 2015) dahingehend ab, dass die Teilnehmer*innen bereits zu Beginn explizit darauf aufmerksam gemacht wurden, dass es um eine Simulation von Online-Übergriffen zu Forschungszwecken ging und nicht – wie häufig in experimentellen Settings zu realen Verhaltensbeobachtungen – erst im Nachhinein über den eigentlichen Zweck der Untersuchung aufgeklärt wurden. Entsprechend der jungen Zielgruppe erschien uns ein solches Vorgehen ethisch angemessen. Die Jugendlichen wurden noch vor Beginn des Spiels ausdrücklich darauf hingewiesen, dass ein Abbruch des Spiels jederzeit möglich sei, für den Fall, dass der Als-ob-Charakter trotz der getätigten Vorkehrungen nicht aufrechterhalten werden konnte und die inszenierten Online-Übergriffe tatsächlich die Ebene der persönlichen Betroffenheit erreicht hätten.

Zur Vermeidung möglicher Spannungen wurden alle Rollenspiel-Teilnehmer*innen am Ende eines Spiels vom Forschungsteam zu einer versöhnlichen Geste aufgefordert, insbesondere zwischen Täter*in und Opfer. Es sollte noch einmal klar kommuniziert werden, dass alles nur ein Spiel war und die Spieler*innen keinerlei böse Absichten hegten.

1.1.4 Umsetzung der Rollenspiel-Workshops, Sampleübersicht und Analyse­methode

Insgesamt wurden 21 explorativ ausgerichtete Online-Rollenspiel Workshops mit rund 170 14- bis 18-jährigen Jugendlichen mit unterschiedlichem Geschlecht und Bildungshintergrund durchgeführt (Erhebungszeitraum: Oktober bis Dezember 2019). Die folgende Tabelle zeigt eine Auflistung nach Schulformen, Erhebungsort, sowie mit den Merkmalen Alter, Geschlecht und Migrationshintergrund der jeweils 8 teilnehmenden Schüler*innen pro Online-Rollenspiel Workshop:

Tabelle 1

Sampleübersicht: Online Rollenspiele					
Gruppe Nr.	Schulformen*	Bezirk	Alter in Jahren	Anzahl Geschlecht Mädchen:Burschen	Migration** ja:nein
G01	HAK	1100 Wien	16 - 17	4 : 4	8 : 0
G02	HAK	1100 Wien	15 - 17	4 : 4	8 : 0
G03	Gymnasium	1090 Wien	17 - 18	0 : 8	4 : 4
G04	PTS	1220 Wien	14 - 15	3 : 5	5 : 3
G05	Gymnasium	1090 Wien	16 - 16	8 : 0	1 : 7
G06	PTS	1220 Wien	14 - 15	6 : 2	5 : 3
G07	Berufsschule	1060 Wien	17 - 19	0 : 8	3 : 5
G08	FMS / PTS	1100 Wien	15 - 16	4 : 4	2 : 6
G09	HAS/HAK	1220 Wien	16 - 18	4 : 4	3 : 5
G10	Berufsschule	1160 Wien	16 - 21	4 : 4	1 : 8
G11	Gymnasium	1190 Wien	14 - 18	5 : 3	1 : 7
G12	Berufsschule	1160 Wien	16 - 21	4 : 4	7 : 1
G13	FMS / PTS	1100 Wien	14 - 15	4 : 4	6 : 2
G14	FMS	1030 Wien	14 - 15	5 : 3	5 : 3
G15	HAS/HAK	1220 Wien	15 - 18	8 [1]	4 : 4
G16	HAS/HAK	1220 Wien	16 - 18	7 : 1	7 : 1
G17	Berufsschule	1210 Wien	17 - 18	4 : 4	8 : 0
G18	Berufsschule	1060 Wien	15 - 16	0 : 8	8 : 0
G19	FMS	1030 Wien	14 - 16 Jahre	1 : 7	6 : 2
G20	Gymnasium	1050 Wien	14 - 17 Jahre	8 : 0	8 : 0
G21	Gymnasium	1050 Wien	15 - 18 Jahre	2 : 6	8 : 0
*HAK = Handelsakademie, PTS = Polytechnische Schule, FMS = Fachmittelschule, HAS = Handelsschule					
** Migrationshintergrund (1. und 2. Generation), dh. ja=allochthone : nein=autochthone Jugendliche					

In jedem Workshop waren 8 Teilnehmer*innen anwesend, entsprechend der Anzahl der zur Verfügung gestellten 8 Smartphones. Es wurden 5 geschlechterhomogene Gruppe (3 Mädchen- /2 Burschengruppen) und 16 geschlechterheterogenen Gruppen umgesetzt. Um unterschiedliche Bildungsebenen zu berücksichtigen, wurden die Jugendlichen aus unterschiedlichen Schultypen rekrutiert (Höhere Schulen mit Maturaabschluss, berufsbildende Schulen, Berufsschulen, Polytechnische Schulen /

Fachmittelschulen). Durchgeführt wurden die Online Rollenspiele in den Räumen (z.B. Klassenzimmer, Gruppenraum, Konferenzzimmer, Freizeitraum, Werkraum) der jeweiligen Schule in Wien und dauerten durchschnittlich 2,5 bis 3 Stunden.

Im Folgenden wird der Ablauf der Online Rollenspiele im Überblick zusammengefasst:

- Einstieg, Aufwärmphase: allgemeine Erklärung (Anonymität, Datenschutz), kurze Vorstellungsrunde, Einstimmung auf Online Rollenspiel, Unterzeichnen der Einverständniserklärung
- Aneignung der Rollen-Profile: Zuteilung der Smartphones mit Profilen und Nicknames inkl. Eintragung am Dokumentationsbogen, Aktivierung und Erklärung des Screencast-Tools; Auseinandersetzung und Adaptierung der Profile durch die Jugendlichen
- Zuteilung der Rollen für das erste Spiel (Ziehung) inkl. Erfassung am Dokumentationsbogen
- Spielstart, Spielende durch Posten von ENDE
- Diskussion: Rollenerleben, Rollenerwartungen
- Ausgleich und Versöhnung mit Täter*in / Opfer
- Neue Spielrunde mit neuerlicher Ziehung der Rollen und Wiederholung des Prozederes
- Nach dem dritten Spiel: Abschlussdiskussion/-reflexion im Workshop, Entlassung aus den Rollen (Trennung Person und Rolle).
- Übergabe einer Kinokarte als Dankeschön für die Teilnahme

Für die **Dokumentation der Online Rollenspiele** wurden grundsätzlich zwei Datenebenen berücksichtigt: Erstens wurden reale Gespräche (insbesondere die Diskussion im Anschluss an die umgesetzten Online Rollenspiele) sowohl mittels Videos als auch auf Tonband aufgezeichnet, wobei die Videoaufzeichnung vorrangig eine Erleichterung für die Transkription darstellte, da in solchen Gruppensettings einzelne Stimmen ohne bildliche Unterstützung oft kaum zuzuordnen sind. Die Aufnahmen wurden anschließend anonymisiert transkribiert und für die Analyse vorbereitet. Zweitens wurden die Daten sowohl mittels Screencasts als auch mittels Screenshots gesichert, um die zentralen Online Interaktionen der Jugendlichen während der Rollenspiele zugänglich zu machen. Bei den Online Rollenspielen verwendeten die jugendlichen Teilnehmer*innen Pseudonyme für ihre Interaktionen bzw. wurden diese Pseudonyme gemäß den Regieanweisungen bestimmten Rollen zugeordnet. Die im Zuge der Online Rollenspiele erstellten Social Media Accounts wurden unmittelbar nach Sicherung der Daten wieder gelöscht.

Die **Auswertung** der erhobenen qualitativen Daten erfolgte inhaltsanalytisch nach Mayring (2003). Dies erlaubt als reduktives und strukturierendes Verfahren eine gute Deskription des Forschungsfeldes, ermöglicht einen systematischen, regelgeleiteten Überblick und ist zudem auch für größere Datenmengen geeignet. Die Datenbasis war sehr umfassend und reichhaltig: insgesamt ca. 60 Spielprotokolle (21 Gruppen à ca. 3 Spiele); zusätzlich 21 Protokolle (1 Protokoll pro Gruppe) zu den Diskussionen während der Spiele und zum Abschluss der Online Rollenspiel Workshops. Aufgrund der Fülle an Material wurde

eine eigene Analysegruppe eingerichtet, deren Aufgabe es war, in einem ersten Schritt das Material mittels eines entwickelten Analyseprotokolls zu reduzieren (festgehalten wurde hier z.B. Zusammenfassungen zur Charakteristik des Spiels und zu den einzelnen Rollen, eingesetzte Counter Speech Strategien und deren Folgen inkl. Originalzitaten). In einem zweiten Schritt wurden die Erkenntnisse in Kategorien verdichtet und zusammengefasst.

1.2 Ergebnisse

Im Folgenden stellen wir erste **zentrale Counter Speech Strategien** der beteiligten Jugendlichen vor, die auf Basis der Analyse der Online-Rollenspiele sichtbar wurden. Jugendliche verwenden oft mehrere Strategien gleichzeitig bzw. sind diese manchmal auch an mehrere Adressat*innen gerichtet (z.B. an Täter*in und Opfer gleichzeitig). Die einzelnen Strategien variieren auch beträchtlich in Tonalität und Intensität (z.B. ernst, humorvoll, sachlich bis übergriffig), wodurch ein unterschiedlicher Bedeutungsgehalt signalisiert wird. Weiters werden die einzelnen Strategien in Abhängigkeit der verschiedenen Adressat*innen (z.B. Täter*innen, Opfer, andere Bystander*innen, an alle beteiligten Personen) unterschiedlich formuliert. Anzumerken ist auch, dass die analysierten Strategien vorrangig auf Basis von Interaktionen auf WhatsApp basieren und somit auch zum Teil plattformspezifisch zu werten sind. So wird WhatsApp überwiegend für die Kommunikation unter Bekannten / Freund*innen genutzt und verläuft daher anderes als z.B. auf Instagram oder auch TikTok, wo häufig auch Kontakte mit wenig oder unbekannt Personen stattfinden. Zudem unterscheiden sich die einzelnen Plattformen auch in den technischen Möglichkeiten, wie kommuniziert werden kann (z.B. Posten von Memes, GIFs oder ausschließlich von Textkommentaren).

In den folgenden Ausführungen wird die Bandbreite an unterschiedlichen Strategien zunächst deskriptiv vorgestellt und dabei auf die jeweils unterschiedlichen Zielrichtungen fokussiert (Täter*innen-bezogene, Opfer-bezogene, Bystander*innen-bezogene und an alle gerichtete Strategien). Danach erfolgt eine differenzierte Analyse der Unterschiede im Bystander*innenverhalten nach Geschlecht, Bildung und Migrationshintergrund. Abschließend gehen wir auf Erfolg und Misserfolg von Counter Speech Strategien sowie auf die Wahrnehmung der Rollenspiele aus Sicht der Teilnehmer*innen ein. Die im Folgenden präsentierten Strategien werden mit wörtlichen Zitaten aus dem Datenmaterial illustriert, um ihre Anwendung zu verdeutlichen. Etwaige Rechtschreib- und Tippfehler oder grammatikalische Ungenauigkeiten wurden bewusst beibehalten und nicht korrigiert, um einen authentischen Einblick in die jugendlichen Lebenswelten zu geben.

1.2.1 Verwendete Counter Speech Strategien

Täter*innen-bezogene Strategien

Die überwiegende Mehrheit an Counter Speech Strategien fokussiert auf Täter*innen. Insgesamt konnten aus den Daten der Online Rollenspiele elf unterschiedliche Hauptstrategien identifiziert werden, wie Jugendliche auf Hasskommentare gegenüber Täter*innen reagieren. Innerhalb dieser Hauptstrategien ließen sich zahlreiche Sub-Strategien eruieren, sodass mehr als 50 unterschiedliche Strategien gefunden werden konnten. Dies ist eine deutlich stärkere Ausdifferenzierung als in bisherigen Studien – die sich allerdings in der Regel nicht auf Jugendliche beziehen (vgl. z.B. Benesch et al. 2016a, 2016b).

(1) Warnung

Das Aufzeigen von möglichen Konsequenzen ist für Jugendliche eine gängige Counter Speech Strategie und umfasst Warnungen vor rechtlichen Folgen („Wenn ihr keine Anzeige wollt dann hört mit der ganzen scheiße auf“, G10,Sz3,S67⁸; „Das ist strafbar. Das ist Mobbing“, G10,Sz2,S24), die mögliche Involvierung Erwachsener („hör auf damit sonst werde ich mit deinen Eltern reden“, G20,Sz2,S22), oder Ankündigungen eines möglichen Ausschlusses aus der Gruppe („Ich block dich“, G20,Sz3,S45). Es wird aber auch mit gewalthaltigen Mitteln gedroht, oft in Verbindung mit aggressiven Ausdrucksweisen („Schau ich geb dir noch eine Chance, lass sie in Ruhe oder ich brech dir den Gesicht“, G18,Sz2,S31) oder unbestimmte Drohungen kommuniziert, um sich für das Opfer zu rächen („Alter ich habe auch von euch genügend peinliche Fotos. Könnte sie auch alle posten“, G21,Sz2,S33). Darüber hinaus finden sich Warnungen mit dem Hinweis, dass sich Täter*innen durch ihr Verhalten selbst schädigen könnten („Du machst dich selber damit peinlich“, G16,Sz2,S22).

(2) Abwertung

Besonders vielfältig sind Strategien, die auf die gezielte Beleidigung von Täter*innen oder auf aggressive Gegenangriffe abzielen: Dazu gehören gezielte persönliche Abwertungen („Peinlich bist du“, G16,Sz2,S31), Hinweise auf Mängel bzw. fehlende Kompetenzen („Rechtschreibung hast in der Schule nicht gelernt oder?“, G16,Sz3,S21), Entzug der Ernsthaftigkeit und des Respekts („Boah dem hast du es aber richtig gegeben *hust *hust“, G11,Sz2,S8), ein allgemein feindlicher, aggressiver Tonfall („was ist mit dir du idiot“, G13,Sz2,S4), oder Sarkasmus („Von wo dieses Selbstvertrauen? 😏“, G16,Sz1,S13). Die Strategie der Abwertung wird manchmal auch auf die moralische Ebene verlagert („Schämst du dich nicht?“, G12,Sz3,S15). Mitunter wird auch der Selbstausschluss aus der Gruppe nahegelegt („Keiner zwingt dich hier zu sein“, G11,Sz1,S3). Eine weitere abwertende Strategie ist es, auf die Unangemessenheit des

⁸ Die Abkürzungen geben die Herkunft der Zitate an und erklären sich wie folgt: G = Nummer der Gruppe (1-21; Anzahl der durchgeführten Online-Rollenspiel Workshops); Sz= Nummer der Szene (1-3; innerhalb jeder Gruppe wurden max. 3 Online-Rollenspielszenen durchgeführt), S = Seitennummer im Transkript.

Verhaltens in dieser Altersgruppe zu verweisen („Ihr streitet wie die Kindergartenkinder“, G13,Sz3,S24) oder die Sinnlosigkeit solcher Online-Übergriffe zu betonen („Leute es ist so traurig so etwas zu Posten wem bringt das bitte etwas?“, G11,Sz2,S10). Aber auch der Ausdruck von Desinteresse und Aufmerksamkeitsentzug wird als gezielte Strategie eingesetzt, um Täter*innen abzuwerten („ES JUCKT KEINEN“, G1,Sz1,S73).

(3) Aufforderung aufzuhören

Die Aufforderung, Täter*innen zum Aufhören zu bewegen, kann sehr unterschiedlich ausgedrückt werden: Dazu zählen direkte Aufforderungen unterschiedlicher Aggressivitätsgrade („Lass ihn einfach in Ruhe man“, G14,Sz3,S29; „Kannst du aufhören mit der Scheiße“, G20,Sz3,S29), abfällige Versuche, Täter*innen in die Schranken zu weisen („Gusch“, G9,Sz1,S13; „Nerv nicht“, G9,Sz1,S14), das Einfordern von Entschuldigungen und notwendigen Handlungen („Lösch und entschuldige dich“, G16,Sz2,S21), aber auch Hinweise auf die eigene Befindlichkeit als dritte, indirekt involvierte Person („Ich hab so keine Lust mehr auf eure Kindergartendiskussion“, G12,Sz2,S16). Im Rahmen dieser Strategie werden auch GIFs oder Memes eingesetzt, um den Wunsch nach Beendigung der Online-Übergriff auf der Bild-Ebene zu kommunizieren. Diese indirekte Form der Intervention hat den Vorteil, dass damit auch z.B. eine humorvolle Ebene mittransportiert werden kann, die Zugänge zu anderen Peers möglicherweise leichter eröffnet.



Stop It! - Arnold | Meme Generator

Abbildung 3: GIF „Stop it“ (G5, Sz3, S8)

(4) Faktencheck

Diese Form der sachlichen Gegenrede wird häufig in Form einer Frage an Täter*innen adressiert und dabei entweder die Grundlage des Online-Übergriffs in Frage gestellt („Woher hast du das Bild überhaupt?“, G11,Sz2,S11), Erklärungen eingefordert („Wieso beleidigst du“, G13,Sz1,S13; „Dann sag mir warum sie eine schlampe ist“, G14,Sz2,S26), oder aber auch vermeintliche Motive für übergriffiges Verhalten des Täters bzw. der Täterin unterstellt („Kann es sein, dass sie nicht mit dir zusammen sein will und du deswegen so ne bitch bist Alex?“, G3,Sz1,S34). Dieser Zugang setzt Täter*innen unter Rechtfertigungszwang und kann entweder zu Eingeständnissen oder auch zur vehementen Verteidigung der eigenen Vorgehensweise führen und unter Umständen auch Übergriffe verstärken.

(5) Moralischer Appell

Eine gängige Strategie der beteiligten Jugendlichen ist es, humane und moralische Werte einzumahnen, indem etwa aufgefordert wird, die Perspektive des Opfers einzunehmen („Stell dir vor jemand postet sowas über dich“, G16,Sz3,S13), auf Normvorgaben verwiesen wird („Sowas macht man nicht“, G20,Sz2,S10; „Kein schimpfen hier“, G19,Sz3,S11) oder dezidiert darauf hingewiesen wird, dass die Verletzung anderer Menschen unmoralisch ist („HÖRT auf man ihr verletzt andere Menschen“, G10,Sz1,S51). Die teilnehmenden Jugendlichen verwenden auch die Strategie, den/die Täter*in auf die eigene Zugehörigkeit zur betroffenen Opfergruppe hinzuweisen („Du bist auch ein Mädchen, vergiss das nicht“, G21,Sz2,S40), oder aber auch mit der Handlungsbewertung aus Sicht einer dem/der Täter*in nahestehenden Person zu argumentieren („was würden deine Eltern sagen, wenn sie das lesen würden?“), G12,Sz3,S16). Der Einsatz von GIFs oder Memes ist hier eine attraktive weitere Strategie, um Appelle an die Moral auf der Bildebene zu transportieren:



Abbildung 4: GIF „Your words hurt like fists!“ (G5, Sz3, S33)

(6) Verteidigung des Opfers – an Täter*innen adressiert

Eng anknüpfend an die Strategie des moralischen Apells ist jene der öffentlichen Verteidigung des Opfers⁹, die aber direkt an Täter*innen adressiert ist. Jugendliche setzen hier unterschiedliche Strategien ein, wie die Betonung der Selbstbestimmtheit („Lasst sie doch, sie soll machen, was sie will“, G19,Sz2,S6), der Hinweis auf individuelle Rechte („Jeder Mensch hat seine daseinsberechtigung“, G1,Sz1,S48) und des Anspruchs auf Individualität des Opfers („Jeder ist hübsch auf eigener Art und Weise“, G10,Sz3,S68). Auch öffentliche Solidaritätsbekundungen mit dem Opfer werden auf unterschiedliche Weise zum Ausdruck gebracht („Leb ani“ als Reaktion auf „stirb ani“, G10,Sz3,S55; mehrmaliges Posten des Namens des Opfers: „Samii“ G7,Sz2,S21; „Wenn Micha raus muss dann geh ich auch“, G11,Sz1,S6). Es wird aber auch versucht, die Berechtigung von Online-Übergriffen in Frage zu stellen, indem die Unschuld des Opfers hervorgehoben wird („geh sie nicht an sie hat doch nichts getan“, G7,Sz2,S13), oder mögliche Ursachen relativiert („Was soll da jetzt so schlimm daran sein?“, G16,Sz2,S10; „Es kann passieren“, G12,Sz3,S7) und Vorverurteilungen abgewehrt werden („Sie ist keine schlampe“, G14,Sz2,S16).

⁹ Opfer werden auch im Privat-Chat kontaktiert. Diese Strategien werden hier aber nicht dargestellt, da sie sich nicht direkt an den/die Täter*in richten.

(7) Deeskalation

Häufig versuchen Jugendliche, Täter*innen zur Deeskalation aufzurufen mit einfachen Aufforderungen, sich zu beruhigen („Leute ohne Spaß kommt bissi runter“, G16,Sz1,S25). Eine wichtige Rolle spielt auch hier der Einsatz von GIFs oder Memes mit dem Ziel, die Stimmung insgesamt aufzulockern. Gleichzeitig wirken diese auch ablenkend. Diese indirekten und oft humorvollen Interventionen auf Bildebene vereinfachen für Jugendliche das Handeln als Counter Speaker*innen und bieten sich oft als attraktive Handlungsoption an.



Abbildung 5: GIF Kind, das „Nein“ deutet (G12, Sz2, S5)



Abbildung 6: Humorvolles Tom & Jerry GIF (G8, Sz3, S30)

(8) Ablenkung

Insbesondere wenn die Situation zu eskalieren droht, versuchen die beteiligten Jugendlichen auch, gezielt Ablenkungsstrategien einzusetzen, indem z.B. ein im früheren Interaktionsverlauf bereits angesprochenes harmloses Thema wieder aufgegriffen und hervorgehoben wird („Also. Ich will immer noch ins Kino...Kinoooo...Kiiiiinooooo...LEUTE...KINOO“, G21,Sz2,S45), oder ein neues Thema eingebracht wird, z.B. Hinweis auf neu erschienene Filme oder andere harmlose Inhalte. Aber auch das Hervorrufen einer Irritation wird als Ablenkungsstrategie eingesetzt, z.B. indem völlig unerwartete (z.B.: Songtext posten, G8,Sz2,S20), oder auch unverständliche Inhalte („Dudepp dudepp palalalappalalakandjdhf dedepp jfirjsksodudhdjd“; „Danke für die Aufmerksamkeit ich hoffe ihr mögt mein Lyrics“, G10,Sz1,S26) gepostet werden. In eine ähnliche Richtung geht auch die Strategie des absichtlichen Zu-Spamens, mit dem Ziel, Interaktionen zu unterbrechen („STOP STOP ... STOP STOP STOP STOP...“, G10,Sz3,S2).

(9) Empathie

Die Strategie, empathisch gegenüber Täter*innen aufzutreten, wird von den beteiligten Jugendlichen selten angewendet, stellt jedoch eine wichtige Möglichkeit dar, mit dem*der Täter*in in Austausch treten zu können. Beispielsweise wird Verständnis gezeigt und Hilfe angeboten („Ich weiß, dass bei dir zu Hause nicht alles gut läuft aber wir können zusammen die Sache regeln“, G21,Sz1,S24), versöhnliche („Ani komm runter, magst du mit zum Essen?“, G21,Sz1,S19) oder verlockende Angebote und Belohnungen gemacht

(„Jordan hör auf, wenn du dich entschuldigst, kauf ich dir Tschick“, G18,Sz3,S14) mit dem Ziel, Täter*innen damit zum Aufhören zu bringen.

(10) Offline Lösung

Eine weitere Strategie ist die Suche nach Lösungsmöglichkeiten außerhalb des digitalen Kontakts, die zur direkten face-to-face Konfrontation auffordern – entweder mit dem Ziel einer Austragung über physische Gewalthandlungen („Komm 1gg1“, G14,Sz2,S31) oder mit dem Ziel der Deeskalation, um in realen und persönlichen Settings weitere Eskalationen zu vermeiden (z.B. „Lasst uns bei nem bier drüber reden“, G15,Sz3,S6; „Treffen wir uns alle persönlich und schauen, wie das endet“, G16,Sz1,S10).

(11) Kontakt zu Täter*in über Privat-Chat

Wenn Jugendliche sich dazu entscheiden, den*die Täter*in im Privat-Chat anzusprechen, macht sie das verletzlicher und erhöht die Möglichkeit, selbst in einen Konflikt zu geraten. Die entsprechenden Nachrichten werden überwiegend neutral bis deeskalierend formuliert. Sie bitten den*die Täter*in aufzuhören („Bitte hör auf. Bitte können wir normal sein.“ G21,Sz2,S60; „Das ist nicht okay“ G3,Sz2,S41), fordern ihn*sie zur Änderung des Umgangstons auf („Digga rede normal“), fordern Erklärungen ein („Was stört dich an ihm so?“ G11,Sz2,S7), oder weisen den*die Täter*in auf die Auswirkungen seiner*ihrer Äußerungen hin („Du weißt schon dass man Leute verletzen kann übers Internet“ G5,Sz3,S49).

Opfer-bezogene Strategien

Bei jenen Counter Speech Strategien, die sich an Opfer von Übergriffen wenden, konnten zwei Hauptkategorien festgemacht werden, die jeweils wieder zahlreiche Sub-Kategorien aufwiesen. Zentral ging es darum, sich in der digitalen Öffentlichkeit schützend vor das Opfer zu stellen oder den privaten Kontakt zum Opfer zu suchen.

(1) Opfer öffentlich in Schutz nehmen

Unterstützung für das Opfer wird vorrangig indirekt formuliert und ist großteils ident mit jenen Strategien, die sich an Täter*innen richten, da diese die eigentlichen Adressat*innen sind. Diese Kategorie umfasst zunächst einfache Aufforderungen, das Opfer in Ruhe zu lassen („hej leute hört auf sie so anzugehen“ G7,Sz2,S9; „Lass Reda doch Mal in ruhe“ G5, Sz3, S14) und die Abwehr von Vorverurteilungen („Sie ist keine schlampe“ G14,Sz2,S16), die auch durch öffentliche Solidaritätsbekundungen unterstützt werden können („Ejjjj balee wer mit Jordan Problem hat muss auch mit mir rechnen ...“ G7,Sz3,S14; „Er ist ein guter Freund und außerdem geht es euch nix an“ G11,Sz3,S8).

Weiters finden sich hier Strategien, die versuchen, Täter*innen die Legitimationsgrundlage für ihr Handeln zu entziehen und damit gleichzeitig das Opfer zu unterstützen. Dazu gehört etwa die Betonung der Selbstbestimmtheit („Er ist schon alt genug er weißt was falsch und was richtig ist“ G13,Sz3,S17; „Sein Leben Bro du kannst dich nicht einmischen“ G13,Sz3,S15), der Recht des Opfers („Er hat genauso ein Recht

hier zu sein wie du“ G11,Sz1,S5), sowie des Anspruchs auf Individualität („Jeder ist hübsch auf eigener Art und Weise“ G10,Sz3,S68; „Jeder hat eigenen Geschmack, lass sie“ G12,Sz1,S14). Weiters sind hier Strategien einzuordnen, die die Unschuld und die ungerechte Behandlung des Opfers hervorheben („Er dachte man kann dir vertrauen“ G 13,Sz3,S19) bzw. das vermeintliche Motiv für die Übergriffe relativieren („Was soll da jetzt so schlimm daran sein?“ G16,Sz2,S10; „Ich finde das Foto nicht Problematik“ G20,Sz3, S11) und sogar positiv darstellen („Ist doch langweilig, normal zu sein“ G12,Sz3,S5).

Vereinzelt wird das Opfer auch öffentlich unterstützt, indem dessen Stärke betont wird und Möglichkeiten für Rache signalisiert werden („Sie [=Opfer] wird dich eh fetzen“ G3,Sz1,S25) bzw. Interventionen des Opfers selbst öffentlich unterstützt werden (*Reaktion auf Intervention des Opfers* „Gut so“ G15,Sz2,S21).

(2) Opfer im Privat-Chat kontaktieren

Die Kontaktaufnahme mit dem Opfer stellt die wichtigste Form der Unterstützung des Opfers dar. Für diesen persönlichen Zugang zum Opfer ist der persönliche Rahmen der Privat-Chat -Funktion erforderlich, um z.B. durch persönliches Nachfragen („Micha was los“ G14,Sz2,S27) Kontextbedingungen zu klären („Hä bist das wirklich du auf dem Foto?“ G11,Sz3,S22; „Was hat er gegen dich?“ G3,Sz1,S43). Dies stellt auch eine wichtige Informationsgrundlage für das eigene Verhalten als Counter Speaker*in dar. Mit der persönlichen Kontaktaufnahme wird dem Opfer aber auch Unterstützung („Wenn du Unterstützung brauchst dabei kann ich gerne mitkommen also wenn du zur Polizei gehst ... oder wenn du Mal reden willst“ G5,Sz2,S54; „Ich bin da für dich“ G9, Sz2, S18) und Solidarität („ich glaube dir dass du nichts falsches gemacht hast“ G5, Sz2, S32) vermittelt.

Auch der Zuspruch von Trost wird praktisch ausschließlich im Privat-Chat übermittelt („lass sie reden sie sind neidisch“  [GIF trauriges Kätzchen] G9,Sz2,S24; „Mach dir keine Sorgen“ G9,Sz2,S18) und manchmal der persönliche Kontakt durch Anbieten eines Offline-Kontaktes verstärkt („Sam geh ma gemeinsam saufen“ G18,Sz3,S3; „Date?“ G3,Sz1,S43). Manchmal ist dies aber auch damit verknüpft, Übergriffe wohlwollend zu verharmlosen („die machen nur spaß, kein stress ... aber es geht auch zu weit langsam“ G3,Sz2,S43) bzw. wird empfohlen, diese nicht ernst zu nehmen („Aber nimm das alles nicht ernst“ G3,Sz1, S43) und dabei auch mit vermeintlich verwerflichen Motiven oder negativen Eigenschaften von Täter*innen untermauert („sie sind neidisch“ G9, Sz2,S24; „alex brudi ... lass dem jordan er soll schreiben was er will er denkt vllt dass er cool ist aber egl“ G13,Sz1,S31).

Die Kontaktaufnahme im Privat-Chat wird auch dafür genutzt, Handlungstipps zu übermitteln („Du kannst sie anzeigen“ G9,Sz2,S18; „Dann blockiere ihn“ G11,Sz2, S18f.). In erster Linie wird allerdings dazu aufgefordert, die Übergriffe einfach zu ignorieren („Ani ignorier ihn einfach er ist ein Arschloch 😞“ G9,Sz2,S8; „Scheiß auf die. Stell es lautlos“ G18,Sz3,S24).

Bystander*innen-bezogene Strategien

Insgesamt konnten vergleichsweise wenige Counter Speech Strategien identifiziert werden, die sich auf andere Bystander*innen, d.h. mitlesende Peers, beziehen. Auch hier konnten wieder zwei Hauptkategorien mit mehreren Subkategorien festgestellt werden: Unterstützung der Gegenrede sowie Absprachen im Privat-Chat .

(1) Gegenrede anderer Bystander*innen unterstützen

Strategien, die auf eine Unterstützung anderer Bystander*innen abzielen, sind gekennzeichnet durch relativ vage Formulierungen, oft in Form von einfachen Ausdrücken der Unterstützung ohne weitere inhaltliche Positionierung („Ja genau“ G15,Sz2,S6; „isso“ G15,Sz2,S12; Like / Daumen hoch posten G5,Sz2,S21; „JJJAAAAEEE“ G14,Sz3,S25). Es kommt aber auch vor, dass an Aussagen anderer angeknüpft wird oder diese verstärkt werden (... aknüpfend an vorangegangene Aufforderung ein Bild zu löschen „Ehrlich nimm das raus“ G16,Sz3,S12; „Das sehe ich genauso“ G9,Sz1,S5).

(2) Mit Bystander*innen in Privat-Chat absprechen

Auch die Idee einer Mobilisierung anderer Bystander*innen zur Unterstützung wird kaum aufgegriffen bzw. wird dies praktisch ausschließlich im Privat-Chat versucht. Neben der Abklärung des Konfliktkontextes („Was hat Dani eigentlich?“ G12,Sz1,S20) wird hier vor allem versucht, sich gegen Täter*innen zu verbünden (*Posting an mehrere Bystander*innen*: „Bist du dafür das wir Ani kicken?“ G10,Sz3,S92) bzw. das weitere Vorgehen abzusprechen („Wir müssen eine Lösung finden zusammen“ G14,Sz2,S30), wobei hier vor allem allgemeines Ignorieren als beste Strategie empfohlen wird („Antworten wir einfach alle nicht mehr dann hören sie auf“ G10,Sz3,S90; „Ignorieren ist beste Wahl“ G12, Sz1, S20) bzw. auch zum Teil dazu aufgerufen wird („Hört nicht auf ihn“ G21,Sz2,S37). Es finden sich aber auch Aufrufe zu Nutzung der Meldefunktion („meldet einfach das foto“ G20,Sz3,S28; „Leute geht Jordan melden“ G16, Sz3, S13). Nur vereinzelt finden sich explizite Aufrufe zur eigenen Unterstützung als Bystander*innen („Leute kommt seit nicht so schüchtern verteidigt ihn Mal ;! Ich kann das nicht allei e“ G11,Sz2,S30; „Leute unterstützt mich mal“ G11,S2,S3).

An alle gerichtete Strategien

Counter Speech Strategien, die sich an alle richten, fokussieren vor allem auf Beruhigung und Deeskalation bzw. auf die möglichst sofortige Beendigung des Konfliktes. Häufig wird damit vermieden, als Bystander*innen selbst eine klare Position zu beziehen. Neben den überwiegend auf Täter*innen bezogenen Strategien stehen solche allgemein deeskalierenden Ansätze in ihrem Einsatz und ihrer Bedeutung an zweiter Stelle. Auch in diesem Strategienbündel wurden zwei Hauptkategorien mit mehreren Subkategorien analysiert, nämlich Bemühungen um ein rasches Ende sowie um Deeskalation.

zur Deeskalation dar („Langsam Leute ist ja nichts passiert“ G21,Sz3,S14; „Was soll da jetzt so schlimm daran sein?“ G16,Sz2,S10; „Omg es ist ein Bild was seid ihr alle nur für special snowflakes“ G10,Sz2,S11“). Ein anderer Zugang der Deeskalation ist jener der Stimmungsauflockerung, der auch an die Strategie des Ablenkens erinnert. Dies wird in der Regel mittels GIFs oder Memes umgesetzt, wie die folgenden Beispiele zeigen:



Abbildung 9: GIF eines tanzenden Hundes (G8, Sz2, S19)

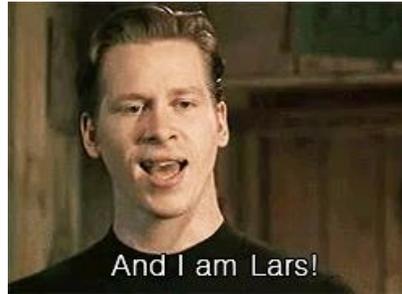


Abbildung 8: GIF "And I am Lars!" (G5, Sz3, S5)



Abbildung 7: liebevolles "Schlaf gut" GIF (G5, Sz1, S21)

Insgesamt verweisen die analysierten Strategien auf eine große Bandbreite: Jugendliche verfügen über ein großes Verhaltensrepertoire, um auf negative Inhalte im Internet zu reagieren. Dabei dominieren ganz klar Strategien, die auf den*die Täter*in bezogen sind. Die folgenden Ausführungen beschäftigen sich näher damit, wie diese Strategien eingesetzt werden und welche Rolle dabei dem Geschlecht und dem Bildungshintergrund von Jugendlichen zukommt. Zusätzlich wird auch der Migrationshintergrund der Teilnehmer*innen berücksichtigt. Außerdem wird eruiert, welche Counter Speech Strategien aus der Sicht von Jugendlichen letztlich zielführend sind.

1.2.2 Unterschiede im Bystander*innenverhalten nach Geschlecht, Bildung und Migrationshintergrund

Die im vorangegangenen Kapitel beschriebenen Counter Speech Strategien wurden von den jugendlichen Rollenspielteilnehmer*innen sehr unterschiedlich eingesetzt. Dies hängt zum einen mit den unterschiedlichen Rollenvorgaben zusammen, die von den Jugendlichen mit Blick auf Geschlecht, Bildung und teilweise auch Migrationshintergrund sehr unterschiedlich angenommen und umgesetzt wurden. Zum anderen war aber auch der jeweilige Spielkontext von großer Bedeutung, das heißt, auf welche Art und Weise zum Beispiel die Übergriffe gestaltet wurden (z.B. ausfällig mit vielen Schimpfwörtern oder eher gemäßigt) und inwieweit dies die Gestaltung von Gegenreaktionen beeinflusst hat.

In der folgenden Ergebnisdarstellung werden daher geschlechts- und bildungsspezifische Unterschiede (Darstellung pro Schultyp) im Bystander*innenverhalten vor dem Hintergrund der jeweiligen Spielkontexte dargestellt. Weiters wird auch auf Besonderheiten von Jugendlichen mit und jenen ohne

Migrationshintergrund hingewiesen. Bei der Analyse der unterschiedlichen Bystander*innen-Rollen mit expliziten Vorgaben (dem Opfer helfen; Täter*in versuchen zu stoppen; die Situation insgesamt deeskalieren; andere Bystander*innen mobilisieren) wurde stets die offene Bystander*innen-Rolle ohne Vorgaben als Referenzfolie herangezogen. Auf diese Weise können geschlechts- und bildungsspezifische Aussagen darüber getroffen werden, wie jugendliche Bystander*innen typischerweise reagieren. Der Bildungshintergrund der teilnehmenden Jugendlichen wird über den Schultyp, den sie zum Zeitpunkt der Datenerhebung besuchten, abgebildet. In die Studie inkludiert wurden Allgemeinbildende Höhere Schulen (AHS Oberstufe), Berufsbildende Höhere und Mittlere Schulen (BHS, BMS: Handelsakademie, Handelsschule), Berufsschulen sowie Fachmittelschulen (PTS: Polytechnische Schule). Die Ergebnisse werden im Folgenden differenziert nach den einzelnen inkludierten Schultypen dargestellt. Für die Differenzierung des Bildungshintergrundes wurde einerseits der angestrebte formale Bildungsabschluss als Annäherung herangezogen (z.B. AHS Oberstufe mit Maturabschluss und Fokussierung auf Allgemeinbildung), andererseits aber auch auffällige Differenzen und Ähnlichkeiten in der Gestaltung des Spielkontexte: So unterschieden sich z.B. die inszenierten Rollenspiele der AHS Oberstufen-Schüler*innen, die aus Schulen in bürgerlichen Bezirken in Wien rekrutiert wurden, deutlich von den Spielkontexten der Schüler*innen aus den Berufsbildende Höheren Schulen (ebenfalls Maturaabschluss aber mit Fokus auf den Erwerb berufsspezifischer Kompetenzen), die aus traditionellen Arbeiter*innen-Bezirken in Wien rekrutiert wurden. Die Spielkontexte der BHS-Schüler*innen ähnelten viel stärker jenen der Berufsbildenden Mittleren Schulen, was auch deshalb schlüssig ist, da die meisten der kooperierenden Schulen sowohl BHS- als auch BMS Ausbildungszweige anboten und daher auch die durchgeführten Rollenspiel-Workshops zum Teil in den gleichen Schulen stattfanden. In der Ergebnisdarstellung wurden daher Schultypen BHS und BMS zusammengezogen.

AHS Oberstufe

Kurzbeschreibung der inszenierten Spielkontexte

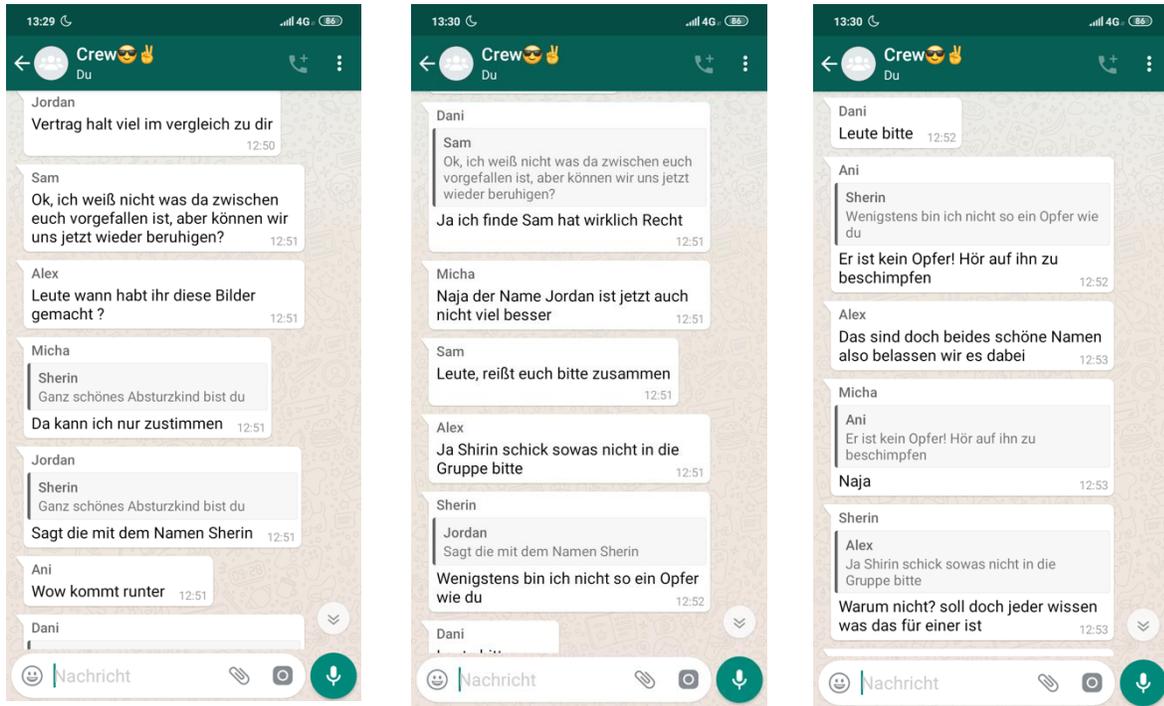


Abbildung 10: Beispielhafter Ausschnitt aus einem Rollenspiel in der AHS (G11, Sz3)

Ein zentrales Charakteristikum der teilnehmenden Schüler*innen aus AHS Oberstufen ist, dass die Inszenierungen der Online Rollenspiele im Vergleich zu den anderen Schultypen in relativ gemäßigttem Umgangston umgesetzt wurden. Trotzdem sind auch hier Schimpfwörter und Kraftausdrücke zu finden, allerdings in einem deutlich geringeren Ausmaß als bei den Inszenierungen der anderen Schultypen und hauptsächlich beschränkt auf die Täter*innenrolle. In der Regel versuchten die Teilnehmer*innen, in einem gemessenen Ton zu intervenieren, es gab aber auch Formen von Gegenattacken, die sich einer angriffigen bis aggressiven Ausdrucksweise bedienten. Im Vergleich zu den übrigen Gruppen wurden auffällig wenige GIFs in den Spielen eingesetzt. Wenn diese verwendet wurden, dann fand dies sehr zielgerichtet statt, entweder zur Ablenkung, zur Stimmungsauflockerung oder um Botschaften an Täter*innen zu vermitteln.



Abbildung 11: Beispiele eingesetzte GIFs (G3,G21)

Burschen signalisieren: „Du bist unreif, nicht ernst zu nehmen und du nervst“

Insgesamt ist das Bystander*innenverhalten der männlichen Rollenspiel-Teilnehmer mit höherem Bildungshintergrund stark täter*innenorientiert ausgerichtet. Das Verhalten von Täter*innen wird generell in Frage gestellt, indem es als unnötig bewertet wird („lasst es doch sein. Was bringt es euch“ G3,Sz1,S23), Unverständnis gegenüber diesem Verhalten signalisiert wird („Was ist dein Problem eigentlich?“ G3,Sz1,S44), die Grundlage der Übergriffe hinterfragt wird („Woher hast du das Bild überhaupt?“ G11,Sz2,S11) und vermeintliche Ursachen relativiert werden („Chillts jeder geht mal meier“ G3,Sz3,S13). Damit wird einerseits die Legitimationsgrundlage für Übergriffe entzogen, andererseits wird das Verhalten von Täter*innen insgesamt als unreif („Ani wie alt bist du? 4?“ G21,Sz1,S8) und Täter*innen insbesondere von Burschen mit Migrationshintergrund als nicht ernst zu nehmend („Und du bist wer?“ G21,Sz1,S,6) dargestellt. Um sich selbst als entsprechend „reif“ zu präsentieren, distanzieren sich die Bystander*innen vom Niveau her gegenüber Täter*innen („Ich hab so keine Lust mehr auf eure Kindergartendiskussion“ G11,Sz2,S23) und bringen eher ihre Genervtheit zum Ausdruck („Euda. Ihr macht mich krank“ G21,Sz2,S50). Burschen mit Migrationshintergrund stellen Täter*innen zum Teil auch als moralisch verwerflich dar („Der Inbegriff von ehrenlos“ G21,Sz2,S28).

Entsprechend gestalten sich die Strategien bzw. Lösungsversuche eher herablassend und reichen von sanft disziplinierend mit der einfachen Aufforderung, sich zu beruhigen („relaxt alle mal ein bisschen. Soll ja nicht noch ausarten“ G11,S1,S10), über sich lustig machend (*postet Bild eines hübschen Mädchens mit Buch und kommentiert dazu*: „Da habt ihr was zum Runterkommen“ G21,Sz1,S7) bis hin zu ernsteren bzw. disziplinierenden Aufforderungen aufzuhören („Leute es reicht“ G21,Sz2,S37) und auch aggressiveren Gegenattacken („Alter, Fresse jetzt“ G21,Sz2,S8).

Weniger um dem Opfer zu helfen oder um eine konstruktive Lösung zu finden, sondern eher um die eigene Genervtheit zu beenden, werden auch Themenwechsel („wie wär's mit einem Themenwechsel bevor hier eskalante Kante passiert“ G3,Sz1,S33), Zu-Spamen („Kino“, „Kino“, „Kino“... G21,Sz2,S45f) und andere Ablenkungstaktiken eingebracht, wie etwa die Unterbreitung attraktiver Angebote bzw. Bestechungsversuche („Okay. Hör mir mal zu. Ein einmaliges Angebot. Kino Auf mein Nacken“ G21,Sz2,S49f.).

Insgesamt zeichnet sich damit im Gruppenchat ein Bild eines*einer mäßig engagierten, von der Unreife der Hater*innen genervten Bystanders/Bystanderin ab, der*die einfach seine Ruhe haben will und auch kein besonderes Engagement zur Unterstützung des Opfers aufweist. Dieses Bild deckt sich auch gut mit den analysierten offenen Bystander*innen-Rollen ohne konkrete Vorgaben: Die männlichen Spieler reagierten eher wenig engagiert, zurückhaltend und vermieden es in der Regel, eine klare Position zu beziehen. Zu finden waren eher einfache Aufrufe zur Beruhigung sowie Ablenkungsversuche. Im Zeitverlauf wurde zunehmend auch eine gewisse Genervtheit zum Ausdruck gebracht, nur bei sehr

aggressiven Angriffen von Täter*innen reagierten die teilnehmenden männlichen Jugendlichen in ihrer Bystander*innenrolle ebenfalls angriffig.

Analysiert man allerdings die eher vereinzelt vorkommenden Interventionen im Privat-Chat, so sind deutliche Abweichungen von diesem Bild zu erkennen: Die Interventionen fallen hier deutlich stärker aus und sind als direkte, moralisch aufgeladene Zurechtweisungen zu charakterisieren („Bro lösche das jetzt man wtf stell dir vor das bist du“ G3,Sz3,S39,PC; „chill bisschen... Das wird zu persönlich“ G21,Sz2,S61,PC), auch das Opfer wird unterstützt („Jordan nimm es nicht ernst. Du weißt eh, wie sie immer Scheiße labbert“ G21,Sz2,S64,PC). Weiters finden sich auch Absprachen mit anderen Bystander*innen im Privat-Chat („Antworten wir einfach alle nicht mehr dann hören sie auf ... Dann verstehen sie hoffentlich das es falsch ist“ G11,Sz2,S32,PC). Es scheint also nicht durchgehend zuzutreffen, dass männliche Bystander mit höherem Bildungshintergrund eher desinteressiert und genervt gegenüber Online-Übergriffen sind und wenig Engagement für das Opfer zeigen. Vielmehr dürfte ein solches Auftreten eher der erwünschten öffentlichen Selbstdarstellung in der Bystander*innenrolle gegenüber Peers in digitalen Medien entsprechen.

Mädchen signalisieren: „Das gehört sich nicht, das darf man nicht!“

Das Bystanderinnenverhalten der weiblichen Rollenspiel-Teilnehmerinnen aus AHS Oberstufen ist ebenfalls überwiegend täter*innenorientiert ausgerichtet. Sie nehmen allerdings stärker auch auf das Opfer Bezug. Im Vordergrund stehen die Einforderung nach Wahrung der Gesprächskultur und zivilisierter Umgangsformen („Können wir uns jetzt wieder wie zivilisierte Menschen verhalten“ G11,Sz3,S6), sowie normative Zurechtweisungen („das sagt man aber nicht“ G5,Sz1,S26; „Recht am eigenen Bild?! Du darfst sowas nicht weiterschicken“ G5,Sz2,S42), wobei Übergriffe zum Teil auch klar als Cyber-Mobbing deklariert werden („Hab keinen Bock auf Cyber Mobbing“ G5,Sz3,S36) und damit als Normbruch dargestellt werden, teilweise unterstützt mit inhaltlich passenden GIFs (*GIF mit dem Inhalt „Stop it“* G5,Sz3,S8).

Das Aufzeigen dieser Regelverstöße liefert den passenden Kontext, um von Täter*innen Rechtfertigungen und Erklärungen für ihr Handeln einzufordern („Hey, ganz kurz, was hat Micha dir getan dass du ihn so wenig leiden kannst?“ G11,Sz1,S19; „Warum teilst du das?“ G20,Sz3,S5), wobei dies zum Teil auch angriffig bzw. sich lustig machend formuliert wird („Ich mein Hallo? Gehts noch?“ G11,Sz1,S4; „Boah dem hast du es aber richtig gegeben *hust *hust“ G11,Sz2,S8; „Im ernst hör auf“ „Es zeigt deine Dummheit“ G20,Sz3,S40). Teilweise werden Täter*innen aufgrund dieser Regelverstöße auch aufgefordert, die Gruppe zu verlassen („Ihr könnt auch gerne einfach die Gruppe verlassen wenn ihr so ein großes Problem mit der Sherin habt“ G5,S2,S29) bzw. wird dadurch auch deren aktiver Ausschluss aus der Gruppe legitimiert und dabei auch Unterstützung durch andere Bystander*innen im Privat-Chat gesucht („Wie können wir diese zwei rauswerfen“ G5,Sz3,S46).

Neben diesen normativen Zurechtweisungen gibt es auf einer zweiten Ebene aber auch Appelle an die Moral bzw. an das Einfühlungsvermögen von Täter*innen, wobei hier Unterschiede zwischen Mädchen mit und ohne Migrationshintergrund auffällig sind: Mädchen ohne Migrationshintergrund vermitteln moralische Appelle nicht öffentlich, sondern nutzen dafür die Privat-Chat -Funktion („Hey, ich finde es echt nicht okay wie du Reda die ganze Zeit persönlich in der Gruppe attackierst mit irgendwelchen blöden ausreden, sorry aber davon wirst du echt nicht cooler wirken...“ G5,Sz3,S41,PC); „Du weißt schon dass man Leute verletzen kann übers Internet“ G5,Sz3,S49,PC). Sie übermitteln öffentliche moralische Appelle eher indirekt, z.B. als allgemeine Aussagen über Toleranz („ich finde es sollte egal sein woher man kommt, solange man sich untereinander versteht!“ G5,Sz1,S35), mittels passender GIFs (GIF mit dem Inhalt „Your words hurt like fists!“ G5,Sz3,S33), oder einer entsprechenden Änderung des Gruppennamens (*Änderung Gruppename in „Peace“* G5,Sz3,S27). Mädchen mit Migrationshintergrund zeigen sich hingegen deutlich stärker emotional aufgebracht und moralisierend gegenüber Täter*innen, auch im Gruppenkontext. Aus dieser Empörung heraus wirken die Interventionen angriffig, auch durch die Verwendung einer sehr expliziten Sprache, die aber nicht vordergründig als Beleidigung gedacht ist, sondern zur Verdeutlichung der moralischen Verwerflichkeit eingesetzt wird („Ja wenn ich jz a Foto von deiner Vagina mach, dann schau wie leben läuft“ G20,S2,S10). Aus dieser Empörung heraus werden Täter*innen dann auch direkter aufgefordert, aufzuhören und sich zu entschuldigen („Entschuldige dich jz bei ihm alter ... Uns lösche das Bild ... Stop this shit“ G20,Sz3,S24).

Konkrete aggressive Gegenattacken gegenüber Täter*innen kommen – auch vor dem Hintergrund der eher gemäßigten Inszenierungen der Spiele – praktisch nicht vor. Allerdings wird zum Teil Genervtheit ausgedrückt bzw. finden sich vereinzelt Versuche, die Übergriffe durch Ablenkung zu umgehen („Kennt ihr diesen meme, RISEEEE AND SHINE“ G5,Sz1,S25; „Und jetzt reden wir über was anderes“ G5,Sz3,S18).

Neben den täter*innenorientierten Strategien sind auch Bemühungen zur Unterstützung des Opfers zu finden: Öffentlich findet dies im Gruppenchat durch Solidaritätsbekundungen („Wenn Micha raus muss dann geh ich auch“ G11,Sz1,S6) bzw. durch ein öffentliches Partei-Ergreifen statt („Er ist kein Opfer! Hör auf ihn zu beschimpfen“ G11,Sz3,S5), oder es wird vermeintlich falsches Verhalten des Opfers öffentlich relativiert („kann jedem passieren“ G20,Sz3,S6). Der Privat-Chat wird hingegen genutzt, um Unterstützung, Mut und Trost zuzusprechen bzw. Ratschläge zu vermitteln („Wenn du Unterstützung brauchst dabei kann ich gerne mitkommen also wenn du zur Polizei gehst“, „Oder wenn du Mal reden willst“, „Ich bin immer für dich da“ G5,S2,S54,PC; „Hi hoffe die Kommentare der anderen verletzen dich nicht zu sehr“ G5,Sz3,S50,PC).

Insgesamt wirken weibliche Bystander*innen im Vergleich zu männlichen engagierter und zeigen auch ein breiteres und vielfältigeres Interventionsrepertoire. Gegenüber Täter*innen steht das Verurteilen des Nicht-Einhaltens von normativen Regeln und zivilisierten Umgangsformen im Fokus. Dabei treten die teilnehmenden Mädchen in der Öffentlichkeit bzw. im Gruppenchat zumindest indirekt moralisierend auf

und bekunden auch ihre Solidarität mit dem Opfer. In privateren Interaktionen kommen moralisierende und das Opfer unterstützende Strategien aber stärker zur Geltung.

Auch hier deckt sich das Bild mit den offenen Bystander*innenrollen: Einige Mädchen übernahmen diese Rolle mit viel Engagement und einem breiten Spektrum an Interventionen (Zurechtweisungen von Täter*innen, Unterstützung für das Opfer, Nutzung von Privat- und Gruppenchat) mit dem Ziel Täter*innen aufgrund von Regelverletzungen klare Grenzen und Zurechtweisungen aufzuzeigen oder moralisierend die Opferperspektive als Spiegel vorzuhalten. Es gab aber auch Spielteilnehmer*innen, die sich ähnlich wie die männlichen Bystander verhielten und kaum reagierten, sich eher desinteressiert und genervt zeigten bzw. die Situation verharmlosten und Täter*innen als „kindisch“ bezeichneten und dann auch nur wenig energisch zur allgemeinen Beruhigung aufriefen („Ganz ehrlich, komm bitte Mal wieder runter“ G11,S2,S13).

Handelsschule (HAS)/ Handelsakademie (HAK):

Kurzbeschreibung der inszenierten Spielkontexte

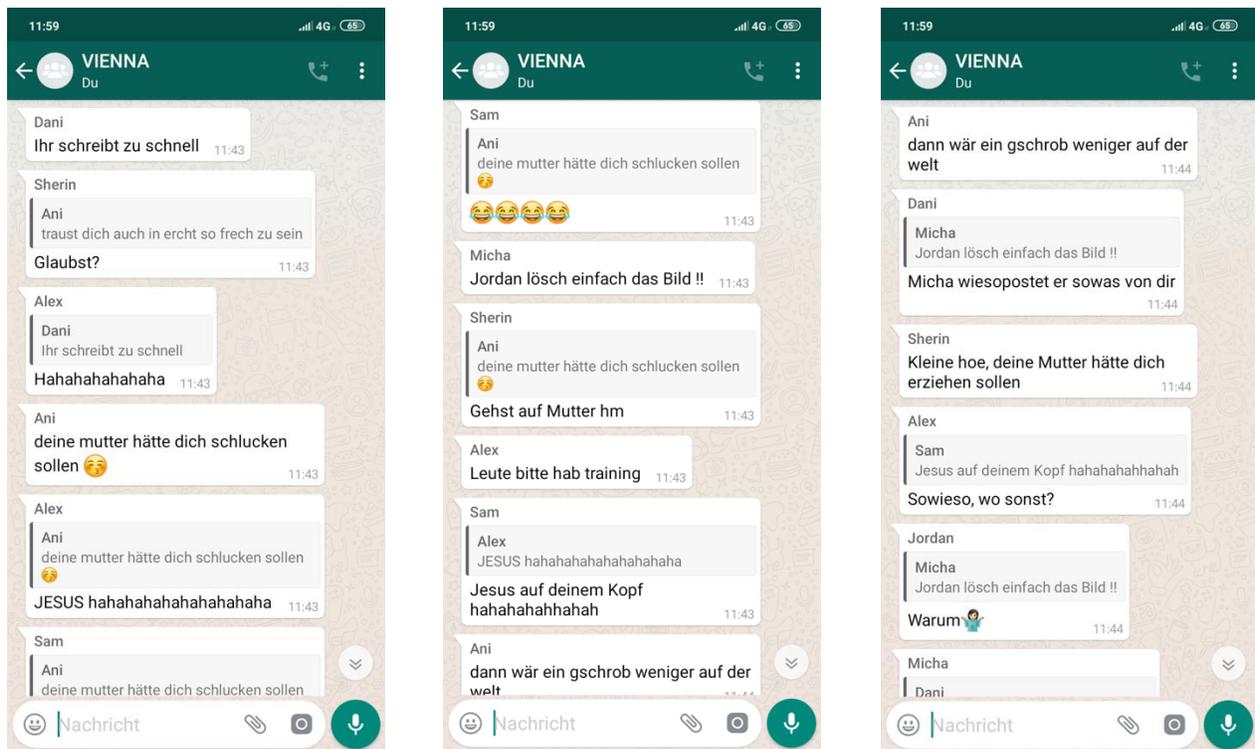


Abbildung 12: Beispielhafter Ausschnitt aus einem Rollenspiel in einer HAS (G16, Sz3)

Die Übergriffe in den Rollenspielen werden im Kontext der Handelsschulen und Handelsakademien sehr ausfällig inszeniert, es herrscht ein abfälliger Sprachgebrauch mit derben und auch vulgären Kraftausdrücken, die insgesamt sehr sexistisch geprägt sind. Trotzdem ist die Sprache weniger vulgär als

in den Gruppen mit Schüler*innen aus der Polytechnischen Schule und aus Berufsschulen. Es werden auch relativ häufig GIFs gepostet, die sowohl der Unterhaltung als auch der Ablenkung dienen und teilweise auch mit konkreten Botschaften verknüpft sind.



Abbildung 13: Beispiele eingesetzter GIFs (G1, G2, G9)

Burschen signalisieren: „Halt die F... und hör mir zu! Dann können wir darüber reden.“

Wieder stehen Täter*innen-bezogene Strategien im Vordergrund, wobei hier aggressive Gegenattacken eine wichtige Rolle spielen („Micha hör auf amk“ G15,Sz2,S10; „Fresse es reicht“ G15,Sz3,S12; „Gusch“ G9,Sz1,S13). Obwohl auf diese Weise auch nach schnellen Lösungen gesucht wird („Man fresse jz mach 1 gegen 1 mit ihm bevor du weiter redest“ G15,Sz2,S24) und Täter*innen dazu auch im Privat-Chat kontaktiert werden („du fuckst so ab“ „halt jz deine fresse“ G1,Sz2,S71,PC), entwickeln sich diese angriffigen Strategien aber oft erst im Laufe des Spiels, um eine Reaktion zu provozieren, nachdem auf die ersten Kommentare nicht oder kaum reagiert wurde (*nachdem andere Strategien keine Wirkung gezeigt haben*: „Habt ihr echt so wenig im Kopf dass nur Schläge die Probleme lösen können?“ G15,Sz2,S27). Insofern scheint dies ein Weg zu sein, um sich selbst Gehör zu verschaffen.

Neben diesen aggressiven Zugangsweisen finden sich aber auch diplomatische, lösungsorientierte bzw. vermittelnde Vorgehensweisen, um z.B. die Ursache des Konflikts zu klären, aber auch um Grenzen aufzuzeigen und Lösungswege anzubieten („Was habe ich verpasst das ihr ihn so behandelt obwohl er ein Mensch ist genau wie ihr“ G9,Sz1,S12; „Hey das geht zu weit“ G9,Sz1,S5; Beruhigt euch und besprecht es doch“ G9,Sz1,S18). Immer wieder werden allgemeine und deseskalierende Aufrufe zur Beruhigung eingebracht („Leuteee“ G1,Sz2,S32; „Heast beruhigt euch alle“ G16,Sz1,S21; „Kommt alle runter und lasst uns bei nem bier drüber reden“ G15,Sz3,S6) oder ein zivilisierter Umgangston eingemahnt („Leute ich will hier keine Beleidigungen sehen“ G1,Sz1,S46; „Leute verhaltet euch mal wie halbwegs zivilisierte Menschen sonst führt das zu nix“ G15,Sz2,S15). Aber auch moralisierende Strategien, die gleichzeitig auch der Herabwürdigung dienen, werden verwendet („Andere runter machen weil man selbst ein Wrack ist ist einfach nur peinlich.“ G15,Sz2,S19) bzw. wird das Handeln der Täter*innen hinterfragt („Was bringt euch das die ganze zeit zu beleidigen“ G1,Sz2,S59; „ist sowas normal bei euch?“ G1,Sz1,S45) und als unreif („Hörts auf ihr seid wie kleine Kinder“ G15,Sz1,S17) dargestellt. Burschen mit Migrationshintergrund

betonen auch die Unehrenhaftigkeit des Verhaltens („Ehrenlose“ G1,Sz2,S27). Immer wieder zeigen sich manche Bystander*innen auch einfach genervt („Bespricht das privat“ G1,Sz2,S60).

Gelingt es nicht, sich für konstruktive Lösungsvorschläge Gehör zu verschaffen, dann werden Strategien der Ablenkung oder Bestechung eingesetzt, um die Wortgefechte zu durchbrechen (G1,Sz1,S64; *postet Bild mit hübscher Frau und fragt* „Leute wie findet ihr das Bild“ G1,Sz2,S30; „WENN IHR AUFHÖRT GEHEN WIR ALLE ZU IBRAHIMS VATER UND DANN BEZAHL ICH FÜR JEDEN DEN DÖNER“ G1,Sz2,S47). Alternativ kommen auch Aufforderung zum Ignorieren vor („Micha ignorier alle und ihr seit Mal leise“ G9,Sz3,S17).

In Bezug auf Strategien, die sich gezielt an das Opfer richten, zeigen sich auffallende Unterschiede zwischen Burschen mit und ohne Migrationshintergrund: Während jene ohne Migrationshintergrund das Opfer kaum unterstützen, am ehesten in Form einer Verharmlosung der vermeintlichen Konfliktursache oder eines versöhnlichen Lösungsvorschlags für beide Parteien („Leute jeder darf doch mal die Sau raus lassen....“ G15,Sz3,S12; „...Also löscht doch den Blödsinn und lacht darüber“ G15,Sz3,S12), zeigen Burschen mit Migrationshintergrund deutlich mehr Unterstützung: Das Opfer wird getröstet und beraten („Ani ignorier ihn einfach er ist ein arschloch“ *Trauriges Emoji* G9,S2,S8; „Du kannst sie anzeigen“ G9,S2,S18) und gleichzeitig gegenüber Täter*innen scharf verteidigt, oft durchaus aggressiv („Hört auf Sam zu mobben“ G15,Sz2,S12; „Sei ruhig! Erstens ist Dani ein Mädchen und zweitens redest du nicht so über sie.“ G1,Sz1,S66; *schlägt Park vor wo 1gg1*: „In park brate“ G15,Sz2,S25).

Insgesamt entsteht der Eindruck, dass die männlichen Bystander zwar durchaus Potenzial für konstruktive Lösungsvorschläge haben, sich aber zunächst entsprechendes Gehör verschaffen müssen und dies auf aggressive Weise einfordern. In Bezug auf das Opfer scheinen nur Burschen mit Migrationshintergrund auch die Rolle als „Beschützer“ zu übernehmen.

Dieses Bild fügt sich gut in die Rollenanalyse jener Bystander*innen ein, die keine expliziten Rollenvorgaben erhielten: auch hier agierten die Bystander*innen Täter*innen gegenüber eher mit kurzen aggressiven und konfrontativen Zurechtweisungen, wobei zum Teil auch die direkte Kontaktaufnahme mit Täter*innen im Privat-Chat genutzt wurde, oder versucht wurde zu deeskalieren, zu beruhigen und die Sinnlosigkeit des Handelns zu betonen und wenn nötig mit Ablenkung die Kommunikation zu durchbrechen. Das Opfer erhielt bei den offenen Bystander*innen-Rollen allerdings keine gezielte Unterstützung, auch nicht von Jugendlichen mit Migrationshintergrund. Stattdessen wurde der Übergriff verharmlost. Möglicherweise wurde diese „Beschützer-Rolle“ auch etwas überinszeniert bzw. kommt erst dann zum Tragen, wenn tatsächlich Opfer betroffen sind, die einem nahe stehen.

Mädchen signalisieren: „Hör auf! Warum machst du das?“

Auch hier sind die Strategien wieder Täter*innen orientiert ausgerichtet. Um aufgrund der ausfälligen Inszenierungen die Gesamtsituation zu deeskalieren, wird allgemein zur Beruhigung aufgerufen („Hey Leute beruhigt euch mal“ G15,Sz1,S17), zum Teil wird das auch etwas energischer und mitunter übergriffig formuliert („OK LEUTE HÖRT ECHT AUF UND CHILLTS MAL“ G16,Sz2,S32; „gusch. Es reicht. Hört auf“ G1,Sz1,S63). Weiters wird die Wahrung einer vernünftigen Gesprächskultur eingefordert („Anstatt zu streiten könnten wir darüber normal reden“ G15,Sz1,S4) oder vermittelt, dass das übergriffige Verhalten der Täter*innen nicht ernst genommen („HEAST IHR KLEINKINDER BERUHIGT EUCH“ G16,Sz3,S18) und als lächerlich wahrgenommen wird („habt ihr eure Tage oder was“ G2,Sz2,S15), zum Teil zeigen sich die Mädchen aber auch genervt von den vielen Übergriffen („wenn ihr streiten wollt geht privat“ G2,Sz2,S12). Mädchen mit Migrationshintergrund äußern sich zum Teil etwas angriffslustiger gegenüber Täter*innen als jene ohne Migrationshintergrund.

Dominierend sind hier Strategien, die das Handeln von Täter*innen in Frage stellen und insgesamt Unverständnis gegenüber den getätigten Übergriffen zum Ausdruck zu bringen. Dies wird zum Teil wiederholend formuliert („Sie hat doch nix gemacht?“ G9,Sz1,S11; „Was hat sie den bitte gemacht das sie ein Problem ist?“ G9,Sz1,S20; „Why“ G9,Sz2,S4; „Wieso seid ihr so gemein?“ G16,Sz1,S18). Dabei wird immer wieder auch an die Moral von Täter*innen appelliert („das ist org unnötig“ „sowas macht man nicht“ G15,Sz2,S6; „Stell dir vor jemand postet sowas über dich“ G16,Sz3,S13).

Vor diesem Hintergrund wird das Handeln von Täter*innen als „unangebrachtes Einmischen“ interpretiert („Was interessiert euch das was er macht es ist seine Sache“ G9,Sz3,S14) und daher Täter*innen auch immer wieder dazu aufgefordert, das Opfer in Ruhe zu lassen bzw. aufzuhören und entsprechende Handlungen, wie z.B. das Löschen von Bildern zu setzen, zum Teil erfolgt dies auch sehr explizit insistierend und energisch („Lass sie doch einfach in Ruhe“ G9,Sz1,S5; „halt stop“ G15,Sz2,S8; „Lösch das Bild“ G9,Sz2,S5; „Lösch und entschuldige dich“ G16,Sz2,S21; „Du sollst es löschen. Was bist du für eine Person?“ G16,Sz2,S26).

Opferorientierte Interventionen sind in einem vergleichsweise geringen Ausmaß zu finden und sind eher darauf ausgerichtet, das Unverständnis bezüglich der getätigten Übergriffe zu unterstreichen („Sam hat nichts gemacht“ G15,Sz2,S11). Opfern wird daher der Ratschlag gegeben, Täter*innen einfach zu ignorieren („Ignorire sie einfach“ G9,Sz1,S22; „Micha ignorier die einfach“ G9,Sz3,S10). Eine Mobilisierung von anderen Bystander*innen zur Unterstützung kommt in der Regel nicht vor, mit Ausnahme von einzelnen Aufforderungen, Täter*innen zu melden („Leute geht Jordan melden“ G16,Sz3,S13).

Insgesamt sind die Counter Speech Strategien demnach hauptsächlich darauf ausgerichtet, Täter*innen zu vermitteln, dass die getätigten Übergriffe auf allgemeines Unverständnis stoßen und daher nicht

akzeptabel sind. Vor diesem Hintergrund wird nachdrücklich, maßregelnd und zum Teil lautstark versucht, Täter*innen zum Aufhören zu bewegen.

Dieses Bild spiegelt sich auch bei der Analyse der offenen Bystander*innenrollen wider, wobei hier auch ein provokant/konfrontatives bzw. teilweise auch ein herablassendes Auftreten gegenüber Täter*innen zu finden ist (z.B. „Dann poste doch mal ein Bild von dir wennst dich traust“ G16,Sz3,S18; „Rechtschreibung hast in der Schule nicht gelernt oder?“ G16,Sz3,S21), was allerdings auch aus als ein emotionales Handeln aus diesem Unverständnis heraus zu erklären ist. Vereinzelt zeigen sich Mädchen in ihrer offenen Bystanderinnenrolle eher desinteressiert oder genervt. Sie setzen keine nennenswerten Interventionen und posten relativ unkonkrete Beruhigungsaufrufe (z.B. „Chill bissl“ G16,Sz1,S30), was als eine Überforderung im Umgang mit der ausfälligen Inszenierung gewertet werden könnte.

Berufsschule

Kurzbeschreibung der inszenierten Spielkontexte

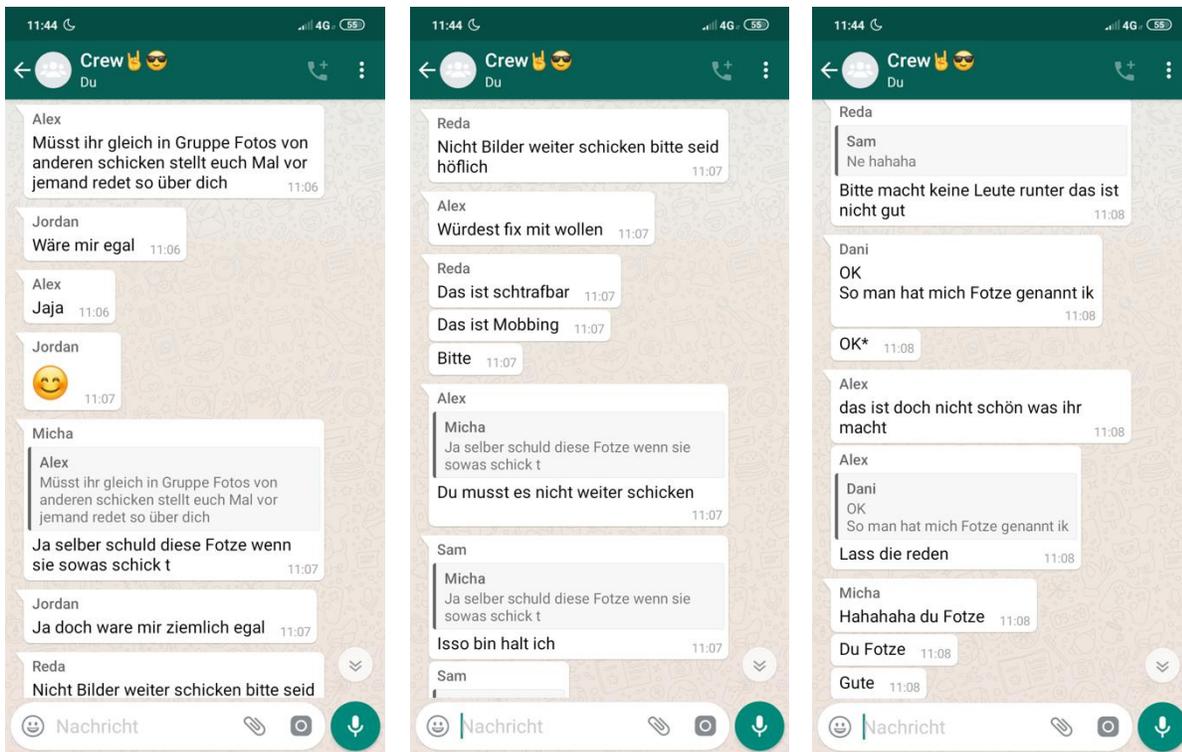


Abbildung 14: Beispielhafter Ausschnitt aus einem Rollenspiel in der Berufsschule (G10, Sz2)

Bei der Inszenierung der Rollenspiele in Berufsschulen sind ausfällige und vulgäre Ausdrucksweise sehr dominant, wobei deutliche Unterschiede zwischen ausschließlich männlich zusammengesetzten und ausschließlich weiblichen Gruppen erkennbar sind: In männlich besetzten Gruppen ist die Sprache noch ausfälliger bzw. abfälliger, sehr häufig gespickt mit derb sexistischen Beschimpfungen, Aufforderungen zu

direkten Konfrontationen und auch gewalthaltigen Drohungen (z.B. „ich schlage dir die Zähne aus“ G7,Sz2,S16). Bei Gruppen mit ausschließlich weiblichen Teilnehmerinnen fallen die Übergriffe deutlich weniger vulgär aus, trotzdem kommen sexistisch aufgeladene Beschimpfungen als „Schlampe“ oder „Hure“ vor. In den Burschengruppen präsentieren sich auch die Opfer selbst als sehr aggressiv, sodass die Rollen zwischen Täter*innen und Opfern oft nicht klar trennbar sind. Während der Rollenspiele werden häufig belustigende, oft auch sexualisierte GIFs eingesetzt. Diese dienen vorwiegend der Unterhaltung, vermitteln in einzelnen Fällen aber auch Botschaften wie z.B. Aufforderungen, sich zu entschuldigen. Das wird vor allem von Mädchen genutzt.



Abbildung 15: Beispiele eingesetzter GIFs (G7, G10, G12)

Burschen signalisieren: „Ich geb dir noch eine letzte Chance du H...“

Die vulgäre und ausfällige Art der Inszenierung macht es Jugendlichen schwer, einigermaßen gelassen und vernunftbezogen zu intervenieren. Es wird versucht, allgemein zu deeskalieren, was aber eher resignierend wirkt („Schluss jetzt!!!“ G17,Sz2,S35; „Hört auf mit schimpfen“ G18,Sz2,S31; „Ihr schimpft zu viel. Haram“ G18,Sz1,S13) und auch z.B. durch die Nutzung von Großbuchstaben zur Verstärkung von Beruhigungsaufrufen kaum Chancen auf Gehör findet („BLEIBT ALLE MAL RUHIG“ G7,Sz3,S27; „RUHE MAL JEDER SCHREIBT JETZT MAL WAS ER NICHT MAG AN DEN MENSCHEN HIER“ G10,Sz2,S24). Die Strategien verlagern sich daher eher auf weniger ernstgemeinte Aufrufe zur Deeskalation („Chillt“ mit lustigem Baby-GIF dabei G7,Sz1,S10; „Heast“ (mit GIF, wo ein Mann mit schlechten Zähnen derb lacht G7,Sz1,S14); „Nein nein nein“ „Das darf man nicht“ G10,Sz3,S1; „Frieden“ mit lustigem GIF G7,Sz1,S20). Andere zeigen sich eher desinteressiert und reagieren nur vereinzelt mit Kommentaren, ohne dabei eine klare Position zu beziehen („Alles wird gut“ G7,Sz2,S7), oder eher genervt („Omg es ist ein Bild was seid ihr alle nur für special snowflakes“ G10,Sz2,S10).

Da sich Beruhigungsaufrufe nicht durchsetzen können, wird vor allem auf Ablenkung als Strategie gesetzt, um die Dynamik der Übergriffe zu durchbrechen („Gemma Kaugummi Automat fetzen“ G10,Sz1,S24; „Will

Neben den vorrangig aggressiven Einschüchterungen und Drohungen durch Burschen mit Migrationshintergrund sind auch Strategien zu finden, die das Verhalten von Täter*innen als moralisch verwerflich darstellen, allerdings wirken diese vergleichsweise hilflos („Du bist eine schlechte Person“ G12,Sz2,S3; „☹️ ☹️ bitte nicht immer blöd über andere reden“ G12,Sz3,S7; „Mann ur absturz kind bist du alex“ G17,Sz2,S30).

Auffallend ist ein insgesamt – sowohl bei Burschen mit als auch ohne Migrationshintergrund – relativ hohes Engagement dem Opfer gegenüber (sowohl im Privat- als auch im Gruppen-Chat). Allerdings präsentieren sich wie oben erwähnt die Opfer selbst als sehr angriffig und konfrontativ, sodass ein solches Engagement eher als Stärkung oder auch Anfeuerung des Opfers zu werten ist. Unterstützt wird in Form von Aufforderungen, das Opfer in Ruhe zu lassen („Lasst Ani in Ruhe lasst sie in Ruhe“ *fünf weinende Emojis* G10,Sz3,S52,PC), durch Verteidigung („geh sie nicht an sie hat doch nichts getan“ G7,Sz2,S13), durch Abschwächung des vermeintlichen Fehlverhaltens des Opfers („Ja und dann fickt sie halt mit ihm was ist so schlimm daran?“ G10,Sz3,S70), der Betonung des Rechts zur selbstbestimmten Lebensgestaltung („Ihr Leben nicht unseres, sie macht, was sie will.“ G12,Sz2,S10), oder auch, indem das Opfer beraten wird („Ignorier sie einfach Ani“ G10,Sz3,S54, PC; „Lass dir das nicht gefallen“ G7,Sz3,S16; „Scheiß auf die. Stell es lautlos“ G18,Sz3,S24) und getröstet wird („Sam geh ma gemeinsam saufen“ G18,Sz3,S3). Auch öffentliche Solidaritätsbekundungen mit dem Opfer kommen vor („Samii“ G7,Sz2,S21 *mehrere Bystander*innen schreiben Nachrichten mit dem Namen des Opfers mit positiven Emojis wie Herzen*; „Leb ani“ *als Reaktion auf* „stirb ani“ G10,Sz3,S55; „Ejjjj balee wer mit Jordan Problem hat muss auch mit mir rechnen ...“ G7,Sz3,S14).

Insgesamt verdeutlicht die Analyse der männlichen Spielteilnehmer*innen aus Berufsschulen, dass die ausfällige Art und Weise der Übergriffe kaum Spielraum für konstruktive Lösungsversuche lässt. Als Ausweg bleiben lediglich (1) aggressive und eindringliche Gegenattacken und Einschüchterungsversuche, um Täter*innen in die Schranken zu weisen; (2) Ablenkungsmanöver, um jegliche Kommunikation zur durchbrechen, sowie (3) eine klare Gegenposition zugunsten des Opfers zu beziehen, um das Machtungleichgewicht etwas auszugleichen. Wer keinen dieser Wege beschreitet, zieht sich von Interventionen eher zurück.

Dieses Bild spiegelt sich auch in der Analyse der offenen Bystander*innen-Rollen wider: Burschen ohne Migrationshintergrund zeigen sich hier deutlich stärker desinteressiert, bringen sich eher wenig ein und zeigen sich genervt (z.B. „Könnt ihr das privat klären?“ G4,Sz1,S11). Burschen mit Migrationshintergrund in der offenen Bystander*innenrolle setzen eher auf Ablenkung durch den Einsatz von humorvollen GIFs, wodurch aber auch vermittelt wird, die Situation insgesamt eher wenig ernst zu nehmen, wenn auch ab und zu halbherzige Beruhigungsaufrufe hinzukommen. Ein vehementes Auftreten gegenüber Täter*innen oder auch Opfern kommt hier aber nicht vor und scheint wohl auch aus der Inszenierungsvorgabe heraus

entstanden zu sein bzw. dient dies, wie bereits erwähnt, eher der Anfeuerung eines ohnehin starken Opfers.

Mädchen signalisieren: „Was glaubst du, wer du bist? Entschuldige dich!“

Unter den Mädchen aus Berufsschulen gab es nur vereinzelt Teilnehmerinnen ohne Migrationshintergrund, daher stehen die folgenden Ergebnisse vor allem für Mädchen mit Migrationshintergrund. Aufgrund der ausfälligen Inszenierungen der Übergriffe sind die Strategien einerseits auf die normative Zurechtweisung im Umgang miteinander ausgerichtet („Leute keine Schimpfwörter“ G10,Sz2,S4; „leute bitte es ist nicht nett“ G17,Sz1,S12; „Halooo verhält's euch normal was ist los mit euch wieso so frech“ G17,Sz1,S18), andererseits wird versucht, die Übergriffe durch einfache Aufforderungen zum Aufhören zu beenden, die teils aggressiv, teils lustlos oder genervt wirken („HÖRT DOCH AUF“ G10,Sz2,S21; „Ja eh hör auf es nervt schon“ G12,Sz2,S16; „Kannst du einfach die Fresse“ G10,Sz1,S43; „es reichttttt“ G17,Sz1,S19). Vereinzelt wird auch Desinteresse signalisiert („Macht das bitte privat aus“ G10,Sz1,S56) oder allgemein zur Beruhigung aufgerufen („Beruhigst euch“ G17,Sz1,S7).

Im Vorgehen gegen Täter*innen werden die Aufforderungen aufzuhören zum Teil auch durch die Thematisierung von Konsequenzen für Täter*innen verstärkt, indem z.B. ihr Verhalten als strafbar benannt wird („Sowas kann man auch anzeigen“ G10,Sz2,S20; „Das ist schtrafbar“ „Das ist Mobbing“ G10,Sz2,S9). Vereinzelt wird auch mit dem Ausschluss von Täter*innen gedroht („Ich bin dafür das Alex aus der Gruppe raus soll G10,Sz1,S25). Gleichzeitig wird von einigen auch versucht, die Autorität der Täter*innen abzuschwächen, indem das eigene Auftreten provokant gestaltet wird und man sich lustig macht („Glaubst auch nur du oder?“ G12,Sz1,S3; „Wenn du das sagst, dann mache ich es sofort“ G12,Sz1,S10; „Weißt du was ich lustig finde? Dass du ihn verarscht, weil er Erfahrungen im Leben macht! Und du? Was machst du? Jeden Tag denselben Ablauf? G12,Sz3,S21). Teilweise wird auch an die Moral von Täter*innen appelliert („HÖRT auf man ihr verletzt andere Menschen benehmt euch doch“ G10,Sz1,S51; „Müsst ihr gleich in Gruppe Fotos von anderen schicken stellt euch Mal vor jemand redet so über dich“ G10,Sz2,S7).

Entsprechend werden als unmittelbare Lösungsvorschläge eine Entschuldigung und entsprechende Handlungen von Täter*innen (z.B. Löschen des Fotos) eingefordert („Er muss sich jetzt entschuldigen.“ G12,Sz3,S17; „Lösch das sofort!!!“ G17,Sz3 S15). Solche täter*innenorientierten Strategien finden sich auch im Privat-Chat mit Täter*innen und werden durchaus sowohl konfrontativ eingesetzt („Was ist dein Problem mit Jordan ... Warum bist do so gemein“ G17,Sz1,S44,PC) als auch normativ zurechtweisend („Wenn du Problem hast rede normal“ G17,Sz3 S20). Auch hier werden Täter*innen klar aufgefordert, aufzuhören („hört bitte auf jz“ G17,Sz2 S26; „Ja sei einfach nett“ G17,Sz3 S12), verbunden mit konkreten Handlungsaufforderungen und zum Teil aggressiven Zurechtweisungen („Lösch das sofort!!!“ G17,Sz3 S15; vallah gusch ja“ G17,Sz3 S19).

Opferorientierte Strategien kommen in einem vergleichsweise geringem Umfang vor, die Strategien sind aber vielschichtig: Das Opfer wird in ein besseres Licht gerückt („Ist doch langweilig, normal zu sein“ G12,Sz3,S5), die Unschuld beteuert („Was hat sie dir getan?“ G10,Sz2,S5), das Recht auf Eigenverantwortlichkeit betont („Das ist seine Sache, das interessiert überhaupt nicht.“ G12,Sz3,S6), Toleranz eingemahnt („Jeder ist hübsch auf eigener Art und Weise also hört auf mit sowas“ G10,Sz3,S68) und die Folgen für das Opfer zumindest indirekt thematisiert („Schämst du dich nicht, dass du einem Freund das Herz brichst“ G12,Sz3,S15f.). Auffallend ist, dass unter den wenigen Mädchen ohne Migrationshintergrund diese opferorientierten Strategien sogar dominanter sind als täter*innenorientierte Strategien.

Insgesamt entsteht der Eindruck, dass die ausfälligen Inszenierungen wenig Chancen für inhaltliche Gegenargumente zulassen, sodass lediglich der Versuch gestartet werden kann, Täter*innen mögliche Konsequenzen aufzuzeigen bzw. zu versuchen, diesen durch moralische und normativ aufgeladene Zurechtweisungen die Grundlage für ihr Handeln zu entziehen.

Der Vergleich mit der Analyse der offenen Bystander*innenrollen bestätigt dieses Bild: Mädchen aus Berufsschulen (v.a. mit Migrationshintergrund) bauen ihre Interventionsstrategien auf Gegenangriffen gegenüber Täter*innen und gleichzeitiger Verteidigung des Opfers auf. Sie versuchen, sich gegenüber Täter*innen als stark zu präsentieren und agieren auch angriffig, kontern häufig und versuchen dabei die Autorität von Täter*innen zu schwächen. Angriffe und Unterstellungen gegenüber dem Opfer werden abgewehrt.

Polytechnische Schule

Kurzbeschreibung der inszenierten Spielkontexte

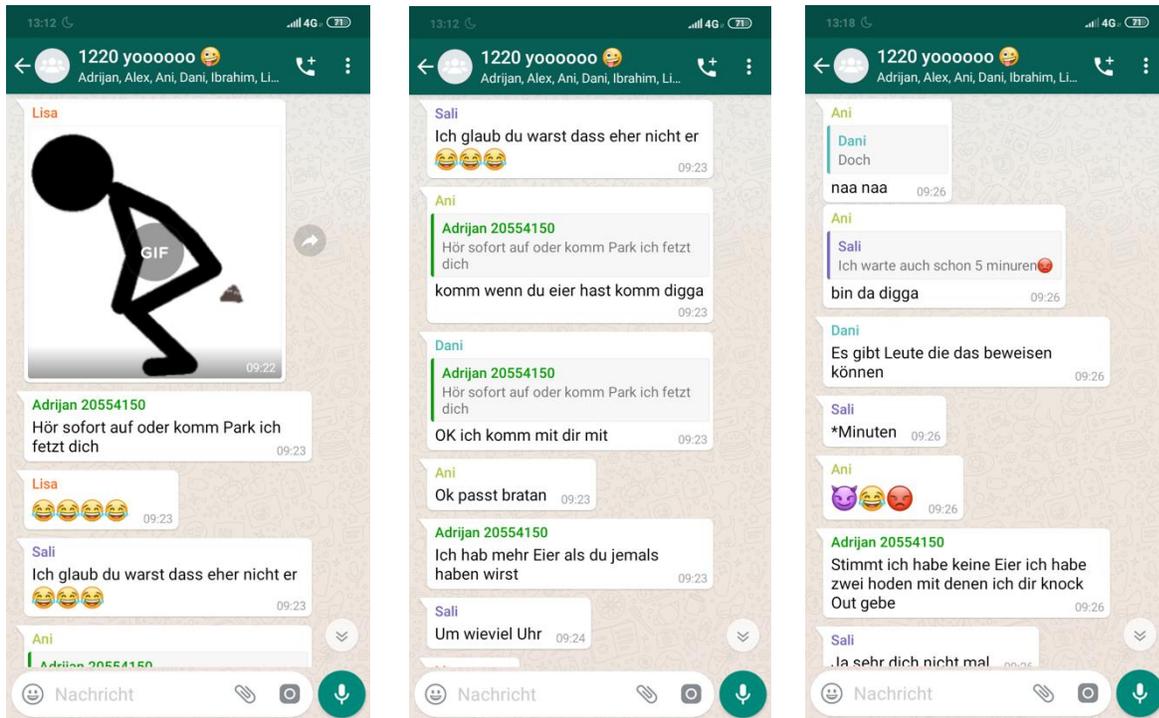


Abbildung 16: Beispielhafter Ausschnitt aus einem Rollenspiel in der Berufsschule (G4, Sz3)

In den Gruppen aus Polytechnischen Schulen herrscht eine Mischung aus sehr aggressiver und derb humorvoller Grundstimmung. Die Inszenierungen sind überwiegend ausfällig mit vielen, oft sexualisierten Schimpfwörtern. Obszöne Beschimpfungen oder Aufforderungen zur direkten Konfrontation (z.B. „komm Park ich fetzt dich“ G4,Sz3,S9) kommen zum Teil auch ohne konkrete Übergriffe vor. In manchen Fällen gelingt es den Jugendlichen nicht oder nur schwer, überhaupt eine Inszenierung zu entwickeln bzw. scheint auch einigen Mitspieler*innen häufig nicht klar zu sein, welche Rolle sie haben, sodass die Spiele fast ausschließlich aus einer Aneinanderreihung von Beschimpfungen bestehen bzw. mögliche Interaktionen regelrecht zugespamt werden. Entsprechend finden sich manchmal entweder keine wirklichen Gegenreaktionen, oder es kommen hilflos erscheinende Aufrufe mit der Bitte, nicht zu schimpfen. Teilweise kommt es auch zu einer exzessiven Verwendung von Emojis und GIFs, die überwiegend zu Unterhaltungszwecken eingesetzt werden.



Abbildung 17: Beispiele eingesetzter GIFs (G4)

Burschen signalisieren: „Hör endlich ihr auf du sch... H... oder ich dreh dich ab“

Aufgrund der ausfälligen Inszenierungen, die kaum Spielraum für inhaltliche Gegenreaktionen bieten, sind allgemein deeskalierende Strategien bzw. Aufforderungen zu Beruhigung hier sehr dominierend („Hey nicht so ausrasten jetzt okey“ *Smiley mit Schweißtropfen und lachen* G4,Sz1,S6; „Leute beruhigt euch“ G13,Sz1,S16; „HALT STOP CHILLT ALLE MAL!!!!!!“ G14,Sz3,S14; „JORDAN SEI ENDLICH LEISE MANN“ G13,Sz1,S20). Dabei wird auch die Unreife von Täter*innen thematisiert („Leute checkt ihr es eigentlich das ihr gerade Ur kindisch und peinlich seit ihr Streit was besser ist das mach nich Mal 10 jährige“ G13,Sz1,S22). Es wird aber auch an die Einhaltung von Normvorgaben appelliert („Bitte verwende keine Schimpfwörter“ G8,Sz1,S38), gleichzeitig wird das aber von einigen auch als aggressive Gegenattacke formuliert („Nicht beleidigen ihr durch Gefickten schieß hurensöhne“ G19,Sz2,S8). Wenige andere wiederum drücken dies durch eine versöhnliche und Frieden stiftende Art und Weise aus („wir sind alle freunde es ist nicht gut wenn wir uns gegenseitig beschimpfen“ G8,Sz1,S27; „Frieden“ G8,Sz2,S22).

Eine besonders wichtige Rolle spielt auch hier Ablenkung und Irritation als Strategie, um die negative Kommunikation zu durchbrechen (*postet Songtext* G8,Sz2,S20; *postet Link zu einer Filmvorschau* G13,Sz3,S18; *ändert Gruppennamen in: hahahahah*“ G8,Sz3,S32), wobei auch Zu-Spamen als Mittel eingesetzt wird (*merhmaliges Posten von: „BERUHIGT EUCH ODER ICH STEH EUCH AB* G8,Sz3,S13ff.“ *Reaktion darauf* mehrmaliges Posten von „FRIEDEN“ G8,Sz3,S54ff.; ständiges Posten eines Stickers G8,Sz3,S55ff.; „Stop Stop Stop Stop Stop Stop Stop“ G19,Sz3,S38) Dabei werden auch z.B. GIFs/Memes genutzt, oft ohne Bezug zur Situation, zum Teil um auszudrücken, dass die Situation nicht ernst genommen wird (*postet obszöne Geste und lachende Emojis* G13,Sz3).

Werden dennoch konkrete Interventionen gesetzt, fallen diese eher aggressiv, beschimpfend und konfrontativ aus („Was is heast. Hört auf damit“ G19,Sz2,S29; „was ist mit dir du idiot“ G13,Sz2,S4; „herst du opfer lösch das Bild“ G14,Sz2,S3; sam du HDF sonst kassierst du von mir richtige watschen und fäuste“ G13,Sz2,S7; „HDF WILLST 1GG1“ G14,Sz3,S19),

Auffallend viele Jugendliche und darunter auffallend viele ohne Migrationshintergrund wirken vor dem Hintergrund der ausfälligen Übergriffe überfordert: Sie wirken im Online-Rollenspiel eher unbeteiligt bzw. beteiligen sich nicht inhaltlich, sondern posten scheinbar willkürlich zur allgemeinen Erheiterung in die Gruppe (z.B. *postet Fotos von Frauen im Bikini* G13,Sz2,S6ff.). Bei einigen anderen fällt auf, dass diese relativ lange inaktiv bleiben und öfter beginnen, etwas zu schreiben, um es dann wieder zu löschen, wie am Screencast-Tool Aufzeichnen ersichtlich. Es entsteht der Eindruck einer gewissen Hilflosigkeit im Umgang mit Interventionen. Dies drückt sich beispielsweise auch dadurch aus, dass dazu aufgerufen wird, etwas zu unternehmen, ohne jedoch konkrete Interventionsvorschläge anzubieten („Wir müssen irgendwas dagegen tun“ G14,Sz2,S27; „Wir müssen eine Lösung finden zusammen“ G14,Sz2,S30). Auch ein Austausch

im Privat-Chat findet dazu statt („Sagen wir der Amina dass sie aufhören soll mit der Scheiße“ G4,Sz2,S43). Vereinzelt wird auch Desinteresse signalisiert („Könnt ihr das privat klären?“ G4,Sz1,S11)

Da es offenbar Hauptaufgabe der Jugendlichen ist, sich überhaupt Gehör zu verschaffen, kommen moralisierende Inhalte („Wieso mischt du seine Eltern ein stell dir vor jmd macht das bei dir“ G13,Sz3,S10), oder versöhnliche und deeskalierende Vorschläge („Ihr solltet euch wirklich Mal gemeinsam unterhalten damit der Streit nicht weiter eskaliert“ G14,Sz3,S16, „Beruhigt euch bitte Mal trifft euch und redet unter vier Augen“ G14,Sz3,S22) in diesen Gruppen eher selten vor.

Ebenso finden sich nur wenige Strategien, die das Opfer unterstützen bzw. in Schutz nehmen („Lasst mein Bro“ G4,Sz3,S5; „Wer was gegen ihn sagt 1vs1“ G4,Sz3,6). In der Regel werden diese dann im Privat-Chat vermittelt („alex brudi“ „lass dem jordan er soll schreiben was er will er denkt vllt dass er cool ist aber egl“ „diskutiere einfach mit ihm nicht“ „ok mein freund“ G13,Sz1,S31, PC; „Jordan“ „Lass die Fetzen in 1110“ G14,Sz3,S55, PC). Nur vereinzelt gibt es Versuche, dafür auch Unterstützung durch andere Bystander*innen zu finden („helf mir bitte dass wir dem dani beschützen“ G13,S2,S32,PC). Der geringe Anteil an opferbezogenen Strategien lässt sich aber auch dadurch erklären, dass in den Inszenierungen die Opferrolle zum Teil nicht mehr klar von den Täter*innenrollen trennbar ist.

Auch hier bestätigt die Analyse der offenen Bystander*innenrollen die gewonnenen Erkenntnisse: Einige versuchen, mit derb humorvollen Inhalten abzulenken (z.B. *postet 2 vulgäre Fotos von Frauen im Bikini* G13,S2,S6ff), andere können wiederum die Rollen nicht zuordnen und attackieren das Opfer bzw. nehmen die Situation insgesamt wenig ernst. Interventionen gegenüber Täter*innen werden vor allem als Gegenangriffe formuliert oder es kommen einfache Aufforderungen, aufzuhören.

Mädchen signalisieren: „Im Ernst, Hört auf! Ich kann schimpfen wie du.“

Aufgrund der ausfälligen Inszenierung der Spiele sind auch hier inhaltliche Gegenargumente kaum möglich. Es wird daher versucht, die Situation zu deeskalieren, indem zur Beruhigung aufgerufen wird („LEUTE“ G6,Sz2,S31; Chillt eure Lage mal“ G8,Sz1,S25; „Beruhigt euch mal!!“ G14,Sz2,S29) und ein angemessener Umgangston einfordert wird („ihr dürft nicht sagen kahbas“ G6,Sz1,S31; „Heast klärt das normal und nicht so“ G14,Sz2,S36; „Kein schimpfen hier“ G19,Sz3,S11; „Digga rede normal“ G8,Sz3,S95, PC), wobei manche auch genervt und angriffig wirken („Heast ihr hs kinder chillts mal“ G8,Sz2,S27; „Heast ihr nervt“ G8,Sz2,S28; *Gruppenname wird umbenannt in „Nervt nicht Alex und reda“, anschließend weitere Änderungen durch andere* G8,Sz3,S37). Vereinzelt wird auch versucht, die Situation durch Ablenkung zu deeskalieren, indem etwa ein Themenwechsel eingebracht wird („Qdh <3 <3“ Anm: Titel eines deutschrap Songs G8,Sz1,S47).

Täter*innen werden dazu aufgerufen, aufzuhören und dabei gleichzeitig darauf hingewiesen, dass dies ernst zu nehmen sei („man könnt ihr damit aufhören es ist nicht lustig“ G4,Sz2,S19; „Alex hör auf ohne spaß checkst du es nicht du bist der einzige der gerade Rum stresst alle sind gegen dich“ G8,Sz3,S35; „Lass ihn einfach in Ruhe man“ G14,Sz3,S29; „Wer bist du dass du so redest“ G8,Sz3,S95). Dabei werden zum Teil auch Konsequenzen angedroht, indem z.B. auf mögliche rechtliche Folgen verwiesen wird („Ich kenne dich auch“ „Also lieber sei leise“ G14,Sz2,S3; „Und außerdem das ist strafbar“ G14,Sz2,S19; „ich blockiere euch“ G6,Sz2,S10). Das Handeln von Täter*innen wird als unsinnig dargestellt („Das ist so unnötig was ihr da macht mit dem ganzen beleidigen usw“ G14,Sz2,S26, „Das bringt euch im leben nicht weiter wenn ihr euch gegenseitig fertig macht“ G14,Sz2,S29) bzw. in Frage gestellt („Was ist dein Problem Sam!?“ G14,Sz2,S25; „Dann sag mir warum sie eine schlampe ist“ G14,Sz2,S26). Es werden auch Appelle an die Moral eingebracht („Du könntest genau in diese Situation sein“ G13,Sz3,S10; „Er dachte man kann dir vertrauen“ G13,Sz3,S19; „Reda geht's noch!? Das kannst du nicht machen!“ G14,Sz2,S16) und zum Teil auch konstruktive Lösungsvorschläge präsentiert („Streit ist keine LÖSUNG“ „!!!!“ „Ihr müsst reden! G14,Sz2,S39).

Trotz dieser vielfältigen Bemühungen entsteht der Eindruck, dass sich solche „gemäßigten“ Interventionen in Anbetracht der generell abschätzigen Atmosphäre offenbar nicht durchsetzen können. Daher versuchen insbesondere Mädchen mit Migrationshintergrund, mit aggressiven und beleidigenden Gegenattacken Täter*innen die Stirn zu bieten. Die Ausdrucksweisen passen sich dabei stark den getätigten Übergriffen an und fallen oft derb, gewalthaltig und sexualisiert aus („Hab deinen Vater gestern xnxx gesehen. Er hat 7cm Schwanz wie konntest du gezeugt werden“ G8,Sz2,S17; „Vallah sei lieber leise oder Fäuste fliegen“ G8,Sz3,S12; „Drecks Kind glaubst du wärst was wer bist du dass du Mutter einmischt du verficcker Schwuchtel“ G8,Sz3,S73, PC; „Komm 1gg1“ G14,Sz2,S31). Bei weniger aggressiven Gegenattacken versuchen die Teilnehmerinnen, sich über die Täter*innen lustig zu machen („Anscheinend bis du Hobbylos“ *lachendes Smiley* G13,Sz3,S6). Auffällig ist, dass es neben diesen sehr aktiven Mädchen, die mittels Gegenattacken intervenieren, auch einige andere weibliche Spielteilnehmerinnen gibt, die sich nicht positionieren möchten („Jeder hat eine eigene Meinung“ G13,Sz1,S15) bzw. inaktiv, desinteressiert und eher ohnmächtig wirken.

Trotz der Dominanz von täter*innenorientierte Strategien wird auch versucht, das Opfer zu unterstützen. Dafür wird vor allem die Privat-Chat -Funktion genutzt („Ich kann dir helfen“ G8,Sz1,S36,PC; „Ich bin für dich da“ G8,Sz1,S37,PC; „Sherin Bro lass dich nicht runter machen jeder kann Mal einen Fehler machen ... Sowas ist nicht peinlich kann passieren“ G13,Sz3,S35,PC). Auch im Gruppenchat wird das Opfer generell verteidigt bzw. in Schutz genommen („Sie ist keine schlampe“ G14,Sz2,S16; „Lass Jordan in ruhe!?!“ G14,Sz3,S15), auch mit indirekten Unterstützungsformen, die sich gleichzeitig direkt an Täter*innen wenden, um diesen die Legitimationsgrundlage für ihre Übergriffe zu entziehen („Wieso gehst du jetzt auf Ani los“ G13,S2,S10; „Sein Leben Bro du kannst dich nicht einmischen“ G13,Sz3,S15; „Das ist ihr Privatleben“ G14,Sz2,S5; „Lasst sie doch, sie soll machen, was sie will“ G19,Sz2,S6).

Insgesamt entsteht der Eindruck, dass zwar die weiblichen Spielerinnen prinzipiell Versuche zur Lösung des Konflikts starten, sich letztlich aber kaum Gehör verschaffen können und dann insbesondere Mädchen mit Migrationshintergrund versuchen, mit Gegenattacken entsprechende Aufmerksamkeit zu erzielen, indem sie sich sprachlich an die Ausdrucksweise der Täter*innen anpassen und damit Stärke beweisen wollen. Dies gelingt aber nur wenigen Mädchen. Viele andere sind überfordert und reagieren sehr verhalten oder gar nicht. Da die meisten damit beschäftigt sind, sich durchzusetzen, gehen auch die vorhandenen Unterstützungsversuche für das Opfer eher unter.

Dieses Bild spiegelt sich auch in den offenen Bystander*innenrollen wider: Mädchen mit Migrationshintergrund präsentieren sich Täter*innen gegenüber als sehr angriffig und aggressiv und werden im Laufe des Rollenspiels immer ausfälliger. Sie sind dabei auch provokant, verwickeln sich in Diskussionen und versuchen, klare Grenzen aufzuzeigen. Dabei wird auch auf die Sinnlosigkeit des Handelns von Täter*innen hingewiesen und zur Beruhigung und Beendigung der Übergriffe aufgerufen. Es gibt auch weniger engagierte Mädchen, die Beruhigungsaufrufe tätigen, trotzdem aber klar aggressiv-zurechtweisend (z.B. „gusch“ „Kein schimpfen hier“ G19,Sz3,S11) auftreten. Mädchen ohne Migrationshintergrund bringen sich in der offenen Bystander*innenrolle aber nur wenig ins Spiel und wirken generell eher desinteressiert, gekennzeichnet durch halbherzige Aufrufe zur Beruhigung.

1.2.3 Hinweis auf Erfolg und Misserfolg von Counter Speech Strategien

Bei der Analyse der unterschiedlichen Counter Speech Strategien wurde auch versucht herauszufinden, inwieweit sich spezifische Strategien insgesamt als erfolgreicher erweisen als andere. Diese Bewertung war allerdings sehr schwierig, da schon allein aufgrund der Rollenvorgaben praktisch kein Abrücken von negativ besetzten Rollen (Täter*innen- und Mittäter*innen-Rollen) vorgesehen war. Das Verharren in der jeweiligen Rolle trifft aber sehr wahrscheinlich auch für realen Settings zu, da sich beispielsweise auch in den Analysen der YouTube Daten (siehe dazu Ausführungen zur Inhaltsanalyse in diesem Bericht) sowie in bisherigen Studien zu dieser Thematik (siehe z.B. Benesch et al. 2016a, 2016b) kaum Hinweise auf ein Zeichen der Einsicht oder gar eine Entschuldigung festmachen lassen.

Insgesamt kann – wie auch zu erwarten war – festgehalten werden, dass keine einzige Counter Speech Strategie in Bezug auf Täter*innen als erfolgversprechend charakterisiert werden kann. Wenn auf Counter Speech direkte Reaktionen von Täter*innen folgten, dann stets in Form von Gegenangriffen, Widerreden oder weiteren Provokationen. Wird Counter Speech in Form von Gegenattacken und Provokationen formuliert, so kann das zwar (kurzzeitig) vom Opfer ablenken, allerdings riskieren Bystander*innen dadurch zum nächsten Opfer zu werden. In vielen Fällen blieb nach Counter Speech eine Folgereaktion durch Täter*innen aus, was sehr vorsichtig zumindest als Teilerfolg interpretiert werden könnte. Allerdings kann dies keinen spezifischen Counter Speech Strategien zugeordnet werden.

Bewertet man den Erfolg von Counter Speech im Sinne einer erfolgreichen Mobilisierung von anderen Bystander*innen, indem diese die Interventionen anderer unterstützen (z.B. durch kurze, unterstützende Statements oder auch einfache Likes), sich anschließen (z.B., indem sie ebenfalls inhaltlich dazu kommentieren) oder sich zusammenschließen (z.B. durch gemeinsame Absprachen), so lassen sich bestimmte Erfolgs-Tendenzen in Zusammenhang mit spezifischen Counter Speech Strategien festmachen – wenn auch die Form der tatsächlichen Unterstützung durch andere Bystander*innen nicht immer gängigen Erwartungen in Hinblick auf zivilcouragiertes Handeln entspricht – die Jugendlichen verwenden z.B. aggressive Gegenattacken oder unterstützen Ablenkungsmanöver ohne Versuch einer Konfliktlösung.

Im Folgenden werden zentrale Counter Speech Strategien vorgestellt, bei denen es auf Grundlage der analysierten Daten aus den Online Rollenspielen noch am ehesten möglich zu sein scheint, andere Bystander*innen zu mobilisieren:

„Ruhe haben wollen“ durch Einforderung zivilisierter und „reifer“ Umgangsformen

Dazu zählen Counter Speech Strategien, die den unzivilisierten Umgangston und das unangemessene, unreife Verhalten („Kindergarten“) thematisieren, allgemeine Aufforderung zur Beruhigung, auch mit dem Hinweis auf die Sinnlosigkeit der Übergriffe. Weiters sind dies auch Kommentare von Counter Speaker*innen, die in diesem Zusammenhang betonen, dass sie eigentlich nur ihre Ruhe haben wollen,

sich selbst von den Übergriffen genervt fühlen und daher auch die Beteiligten auffordern, diese Dinge privat zu klären. Vor diesem Hintergrund wird häufig zusätzlich der vermeintlichen Anlassfall für den Übergriff als harmlos bezeichnet (z.B. das ist doch nur ein Bild) bzw. Übergriffe als unrechtmäßiges Einmischen interpretiert (z.B. ist ihr Leben, sie kann machen was sie will) – wodurch gleichzeitig implizit ein Einmischen generell, d.h. auch für Bystander*innen in Frage gestellt wird.

Wenn Counter Speaker*innen durch die genannten Strategien durch andere Bystander*innen Unterstützung finden, steht dabei zwar auch, aber nicht zwingend im Vordergrund, dem Opfer helfen zu wollen, sondern es geht auch immer und manchmal in erster Linie darum, die eigene, unangenehme Situation als Bystander*innen zu verbessern. Bystander*innen haben oft ganz einfach keine Lust zu intervenieren oder eine Intervention ist unangenehm, besonders dann, wenn Jugendliche eher unvorbereitet mit Übergriffen konfrontiert sind, ein erfolgreiches Eingreifen als aussichtslos erscheint oder sie mit der Situation überfordert sind. Daher stoßen solche Aufrufe mit dem Ziel, eine angenehme Gesprächsatmosphäre (wieder)herzustellen, auch auf Anklang bei anderen Bystandern*innen.

Klare Benennung als moralisch verwerflicher, rechtlich sanktionierbarer Normverstoß

Grundlage dieser Strategie sind vor allem klare Hinweise auf die moralische Verwerflichkeit der Übergriffe, indem Täter*innen als schlechte (z.B. unehrenhafte, peinliche) Persönlichkeiten bezeichnet werden, denen aus der Opferperspektive ein klarer Spiegel vorgehalten wird (z.B. „stell dir vor, das bist du“), oder klar gemacht wird, dass ihr Verhalten gesellschaftlich verankerten Moralvorstellungen klar widerspricht (z.B. Einfordern von Toleranz gegenüber Herkunft). Gleichzeitig wird das Handeln auch als klarer Normverstoß bezeichnet (z.B. „ist wirklich unterste Schublade“) und deutlich gemacht, dass Grenzen erreicht oder überschritten wurden. In der Folge wird auch deutlich auf mögliche rechtliche Konsequenzen verwiesen (z.B. „das ist strafbar“). Das explizite Bezeichnen und Benennen als Grenzüberschreitung, wodurch auch mögliche Konsequenzen z.B. auch in rechtlicher Hinsicht legitimiert werden, schafft für andere Bystander*innen offenbar eine gewisse Orientierung, die es leichter macht, ein klar als falsch deklariertes Verhalten auch zu erkennen und dagegen aufzutreten.

Einbringen von Lösungsvorschlägen und klaren Handlungsaufforderungen zur Beendigung des Konflikts

Bystander*innen unterstützen andere Counter Speaker*innen eher dann, wenn diese zum Aufhören aufrufen und dabei gleichzeitig auch Täter*innen zu einer klaren Handlung auffordern, wie z.B. sich zu entschuldigen, ein Bild zu löschen, aus der Gruppe auszutreten oder auch, sich einer Face-to-Face Konfrontation (mitunter auch körperliche Austragungen „1gg1“) zu stellen. Unterstützung finden auch konfrontative Fragen, die Täter*innen in einen Rechtfertigungszwang bringen (z.B. „Warum machst du das?“) sowie Vorschläge zur konstruktiven Konfliktlösung, indem z.B. ein Gespräch unter vier Augen außerhalb des Social Media Kontexts vorgeschlagen wird. Mit der Aussicht auf möglichst schnelle Lösungen, die noch dazu in einem Offline-Kontext ausgetragen werden sollten, sehen andere Bystander*innen neben der schnellen Klärung für das Opfer auch wieder für sich selbst eine Chance, sich

aus der unangenehmen Situation zu bringen. Zudem werden hier keine unmittelbar eigenen Beiträge zur Lösung erwartet, vielmehr wird dies Rolle anderen Parteien zugeschrieben.

Mobilisierung durch kollektive Handlungsaufforderungen und öffentliche Zeichen der Solidarität

Counter Speaker*innen fordern andere Bystander*innen nur selten dezidiert zur Mithilfe, zur Unterstützung oder zur Solidarisierung auf. Trotzdem folgen in der Regel zumindest einige Personen diesem Aufruf, besonders dann, wenn diese als kollektive Handlungsaufforderungen (z.B. ignorieren wir alle, melden wir alle – in der Regel im Privat-Chat übermittelt) oder öffentliche Zeichen der Solidarität gesetzt werden (z.B. unterstützende Nachrichten mit dem Namen des Opfers mit positiven Emojis, Bystander*innen folgen mit weiteren Nachrichten mit Namen des Opfers und positiven Emojis). Klare Handlungsanweisungen und „Vorbilder“, Markierungen und Zeichen für die Gestaltung von Interventionen (z.B. immer mit Namen des Opfers; ähnlich auch Markierungen mit Hashtags) vereinfachen es Bystander*innen, Interventionen auch tatsächlich umzusetzen. Gleichzeitig entsteht damit eine Solidarisierung mit anderen Bystandern*innen, die das gemeinsame Vorgehen stärkt.

Insgesamt zeigen die Ausführungen, dass es vor allem dann besser gelingt, Bystander*innen als Unterstützung zu mobilisieren, wenn damit auch in Aussicht gestellt wird, die Befindlichkeit der Bystander*innen selbst zu verbessern (z.B. schnelle Lösung/Beseitigung zur Wiederherstellung einer angenehmen Gesprächsatmosphäre), oder wenn ihnen eine klare Orientierung vermittelt wird, dass ein Eingreifen tatsächlich angebracht ist (z.B. Benennung als Normverstoß/ moralische Verwerflichkeit). Im Idealfall enthalten solche Interventionen Anweisungen, was zu tun ist.

1.2.4 Wahrnehmung der Online-Rollenspiele aus der Sicht der Teilnehmer*innen

Im Zuge der Online-Rollenspiele wurden nicht nur die Interaktionen auf den Smartphones der Jugendlichen dokumentiert, sondern auch Videos von den Diskussionen mit den Jugendlichen aufgenommen. Es wurden sowohl Diskussions- und Reflexionsrunden zwischen den einzelnen Spielen durchgeführt als auch abschließende Diskussionen nach Beendigung der Rollenspiele.

Diese zielten vor allem darauf ab, dass die Spieler*innen

- ihre Rollen, insbesondere ihre gespielten Bystander*innenrollen, gemeinsam reflektierten,
- die Beurteilung ihrer Handlungen hinsichtlich ihrer Wirksamkeit und Auswirkungen miteinander austauschten und
- die Online Rollenspiele mit ihrer Lebensrealität und eigenen (Konflikt-)Erfahrungen verglichen und diskutierten.

Die Diskussionsrunden wurden anhand eines Leitfadens moderiert. Die teilnehmenden Jugendlichen konnten ihren spontanen Eindrücken zu den Rollenspielen mitteilen und diskutierten folgende Themenbereiche¹⁰:

- Fragen zu den eingenommenen Rollen (Rollenerleben, Rollenerwartungen etc.)
- Fragen zur Entscheidung zur Handlung (Was waren hier eure Überlegungen? Wie habt ihr entschieden, was ihr tun sollt? Warum habt ihr dann so gehandelt?)
- Fragen zur Beurteilung der Handlungen (bezogen auf eigene Handlungen als auch jene der anderen Spieler*innen, Wirksamkeit, Auswirkungen, Nutzen)
- Fragen zur Abschätzung der Realitätsnähe (ähnliche Situation beobachtet/erlebt, Unterschied zwischen den Handlungen in Realität und Spiel)
- Fragen zur Einschätzung bestmöglicher Interventionen (Welche Handlungen wären zielführend und sinnvoll? Wie lässt sich eine Eskalation stoppen?)
- Fragen zur Einschätzung aller drei Rollenspiele im Vergleich (Spaßfaktor, Rollen, Realitätsnähe, Interventionen etc.)

Die Diskussionen nach bzw. zwischen den Online Rollenspielen zeigten sehr deutlich auf, dass die teilnehmenden Jugendlichen aufgrund ihrer heterogenen Lebenskontexte (Bildung, Alter, Geschlecht etc.) sehr unterschiedliche und auch entgegengesetzte Erfahrungen und Sichtweisen zu den Online Rollenspielen einbrachten. Durch die kurze Zeit zwischen den einzelnen Spielen konzentrierten sich die

¹⁰ Die Themenbereiche wurden aufgrund unterschiedlicher Rahmenbedingungen (z.B. Zeitausmaß, Zusammensetzung der Gruppe, Diskussionsverlauf und -dynamik etc.) in den Gruppendiskussionen sehr unterschiedlich ausgeführt.

Diskussionen vor allem auf das Rollenerleben der Schüler*innen, die Realitätsnähe und die Einschätzung der bestmöglichen Interventionsoptionen in den jeweiligen vorgegebenen Szenen.

Im folgenden **Ergebnisteil** wurden die Diskussionen auf ihre inhaltlichen Aussagen ausgewertet. Die Diskussionen zeigen deutlich auf, dass in den konkreten Rollenspielen die vorgegebenen Social Media Plattformen (WhatsApp, SnapChat und Instagram) von den Schüler*innen sehr unterschiedlich eingesetzt wurden (öffentlicher und privater Modus; Einsatz von Texten, Bildern, Emojis; Nutzung nur von WhatsApp bis hin zu alle drei Foren gleichzeitig etc.). Ziel war es, in der jeweils vorgegebenen Spielrolle (Opfer, Täter*innen, Bystander*innen) möglichst so zu handeln, *als ob* es Wirklichkeit wäre.

Persönliche Befindlichkeit während der Rollenspiele

Auf die Einstiegsfrage nach jedem Rollenspiel: *Wie war es für Euch? Wie geht es Euch?*, antworteten die Schüler*innen sehr unterschiedlich. Wurden nach dem ersten Rollenspiel (Szene mit Beschimpfung) noch häufig die Begriffe *lustig, ganz lustig spannend, super, very good*, also positive Begriffe für das persönliche Erleben des Rollenspiels gewählt, so kamen im zweiten und dritten Rollenspiel vermehrt Formulierungen mit der Steigerung des Adjektivs wie *viel besser, leichter, eh gut, besser gelaufen, eindeutig besser*. Das kann auch damit begründet werden, dass das erste Rollenspiel in vielen Gruppen die Funktion des „Aufwärmens“ hatte. Die Spieler*innen mussten sich zu Beginn erst mit dem Handy, dem eigenem Profil auf drei Plattformen, den Spielregeln sowie mit den eigenen Rollen vertraut machen. Besonders bei jüngeren Jugendlichen mit niedrigem Bildungshintergrund (Polytechnische Schule) kam es öfter vor, dass die Rollenbeschreibungen allein auf Basis der ausgeteilten Karten und der dort zur Verfügung gestellten Sätze nicht klar war. Diese Jugendlichen wurden vom Forschungsteam unterstützt, um ein gutes Verständnis von der Spielanleitung und der eigenen zu spielenden Rolle zu entwickeln.

Neben den vermehrt positiven Sichtweisen zum zweiten Spiel (Szene mit Nacktfoto) wurden in den Diskussionen zusätzlich auch Begriffe wie *chaotisch unangenehm, stressig, langweilig, beschissen, überfordernd* etc. genannt. Mit Blick auf Geschlecht und Schulform können hier kaum Unterschiede in den Gesprächen erkannt werden, hingegen verbalisieren eher jüngere Teilnehmer*innen in den Rollenspielen vermehrt mehr Spaß zu haben („*urgeil*“) als ältere Jugendliche. Sie scheinen – wie auch auf den Handy-Transkripten zu erkennen ist – mehr Spaß und Unterhaltung beim Schimpfen und Posten zu haben und, wie sie selbst auch sagen, vieles nicht so ernst zu nehmen. Während des Spiels konnte bei einigen jüngeren Schüler*innen immer wieder lautes Lachen, Kichern und teilweise auch Kontaktaufnahme durch lautes Reden und Fragen „*Wer bist Du?*“, „*Welchen Nickname hast du?*“ beobachtet werden.

Auffallend ist auch, dass in den zweiten und dritten Rollenspielen von den jugendlichen Workshopteilnehmer*innen vermehrt mitgeteilt wurde, dass nun das Agieren leichter fiele, da es auch einen Grund (Nacktfoto, Fakebild) gibt: „*Es war anders, weil nun ein Grund*“, aber „*sehr okay*“, da die

Teilnehmer*innen sich nicht wahllos gegenseitig beschimpften. Vermehrt wurde aber auch darauf verwiesen, dass es „echt anstrengend“ oder auch „wahnsinnig anstrengend“ sei und man Kopfschmerzen bekomme, wenn man seiner Rolle gerecht werden wollte und doch wirkungslos bliebe. Vereinzelt wurde auch gesagt, dass es „schwierig“, „kompliziert“ und „überfordernd“ sei, sich einerseits in eine (fremde) Rolle einzufühlen und diese zu spielen und andererseits sehr realitätsnah zu handeln.

Die Jugendlichen definierten auch ihre Befindlichkeit in Abhängigkeit von der gespielten Rolle. So war es beispielsweise für einige Mädchen sehr „amüsant und unterhaltsam“ (G11), in die Rolle des Täters bzw. der Täterin zu schlüpfen, andererseits teilten Schüler*innen mit, dass für sie die Bystander*innen-Rolle sehr anstrengend und schwierig war, vor allem, wenn sie helfen wollten. Vereinzelt stiegen auch Jugendliche während des Rollenspiels aus und surfen im Internet. In einigen Gruppen nahmen auch Mädchen und Burschen teil, die sich im Rollenspiel nicht wirklich zurecht fanden und sich sichtlich unwohl fühlten, aber ihre persönliche Befindlichkeit nicht verbalisierten wollten oder konnten. Sie erlebten es als angenehm bzw. vorteilhaft, in der Gruppe untertauchen zu können und nicht angesprochen zu werden.

Insgesamt kann festgehalten werden, dass es den Rollenspieler*innen mehrheitlich Spaß gemacht hat, sich in die verschiedenen Rollen einzufühlen. Sie wurden angeregt, über mögliche Interventionen und Deeskalationsstrategien nachzudenken und sich darüber auszutauschen.

Einschätzung der Realitätsnähe der Online-Rollenspiele

Im Folgenden werden die Diskussionsbeiträge der Jugendlichen über die Realitätsnähe der gespielten Online Rollenkonflikte getrennt nach den Spielszenen dargestellt.

Spielszene 1: Beschimpfung

Auffallend in den Diskussionen ist die sehr große Bandbreite bzgl. der Einschätzung der Realitätsnähe des virtuellen Rollenspiels mit Streit und Beschimpfungen auf Social Media. Die Erfahrungen der teilnehmenden Jugendlichen erweisen sich als sehr divers. Sie reichen von der Feststellung, dass ähnliche Situationen bislang weder persönlich erlebt noch beobachtet wurden („Das hab’ ich noch nie selber erlebt“ (G1), über vorsichtige und vage Bestätigungen „Glaub’ schon“ oder „Schon auch in echt“ (G21) bis hin zur Behauptung der Realitätsnähe aufgrund eigener Erfahrungen und Erlebnisse im Internet: „Ja, kommt vor!“ (G11), „Ja, so läuft es im Internet“ (G4) oder „sehr sogar, auch eins zu eins die Beschimpfungen“ (G19). Ebenso gibt es Jugendliche mit Social Media Erfahrungen, die konkrete Spielszenen als „absolut“ realitätsfremd einschätzten. Ein differenzierter Blick auf die Einwände zeigt, dass vor allem a) die Themen für den Anlass des Streits im Rollenspiel, b) die sprachliche Ausdrucksweise bzw. die Intensität und Wortwahl bei den Beschimpfungen sowie c) der Bekanntheitsgrad der beteiligten Personen im Rollenspiel ausschlaggebend für die unterschiedliche Einschätzung der Realitätsnähe des Rollenspiels waren.

ad a) Themenwahl:

Nach Einschätzung der Teilnehmer*innen würden in Sozialen Medien Themen wie Basketball, Umwelt, Macht etc. nicht zu gegenseitigen Beschimpfungen führen, dies sei „unnötig“. Überhaupt beschimpfe man sich nicht wegen einer Meinungsverschiedenheit, sondern höchstens, wenn man eine bestimmte Person nicht leiden könne. Denn, so die Jugendlichen, es gäbe bessere Möglichkeiten, wie unangenehme Dinge oder Geheimnisse zu posten oder Situationen auszunutzen und sicherlich „nicht nur ein hin und her schimpfen“ (G16).

ad b) Sprachliche Ausdrucksweise:

Bezüglich der Sprache meinten einige Diskutant*innen, dass die Beschimpfungen sehr harmlos, „volksschulmäßig“ und „in echt ganz anders“ oder auch viel schlimmer wären („in Wirklichkeit wäre ich gemeiner“) und begründeten dies beispielsweise so: „ich hab’ mich nicht getraut zu beleidigen“ (G2). Auch wenn es nur ein Spiel war, erzählten die Jugendlichen, dass sie sich teilweise beim „richtigen“ Schimpfen gehemmt gefühlt hätten „wir beschimpfen - nur nicht so richtig“ (G6). In Wirklichkeit würden User*innen anders agieren und „wirklich ausrasten“, was sie aber hier im Spiel nicht gemacht hätten, weil sie sich doch untereinander kennen würden. Hingegen meinte eine 16-jährige Gymnasiastin, dass die Beschimpfungen im Internet weniger aggressiv und verletzend seien und „nicht oft so übertrieben mit den Beschimpfungen“ (G05) wie im Rollenspiel.

ad c) Bekanntheitsgrad:

Einige Diskutant*innen waren auch davon überzeugt, dass man auf bestimmten Plattformen wie beispielsweise auf WhatsApp nicht mit unbekanntem Personen streiten respektive sich beschimpfen würde (G4). Eventuell würde man auf Instagram und YouTube schlimme Beleidigungen finden. Einige Burschen meinten wiederum, wenn man Personen nicht kenne, dann schimpfe man auch nicht in dieser Weise zurück, denn „so viel herumreden wie jetzt, tun wir nicht“ (G18). Fremde Personen würde man nicht ernstnehmen und sofort ignorieren.

Spielszene 2: Nacktbild

Bei diesem Rollenspiel waren sich fast alle Spieler*innen einig, dass diese Geschichte sehr realitätsnah sei, jedoch das gepostete Nacktbild nicht der Realität entspreche. „Das gibts schon (lachen), aber jetzt im Spiel nicht so schlimm, da nur Füße (lachen)“. In Echt würde das Foto nicht nur einen kleinen unklaren Ausschnitt von nackten Beinen zeigen, sondern deutlich mehr vom Körper und dann würde auch die Situation wirklich eskalieren, so ein Jugendlicher. Im Spiel gab es aufgrund des Fotos jedoch wenig Veranlassung sich einzumischen, aber „Im Leben ist alles brutaler als so was!“ (16-Jähriger, HAK, G16.) Mädchen meinten auch, dass sie diese Situationen kennen würden und dass man diese Form der Belästigung einfach nicht ernst nehmen sollte. „Es ist ganz normal, dass jemand so was schickt oder Nacktbild oder solche Sachen – das ist lustig!“ (15-Jährige, PTS, G6). Fotos würden zudem „aus Liebe“ verschickt werden und das sei üblich. Dass es dann auch in falsche Hände kommen kann, sei deshalb

einfach Realität, so ein Mädchen aus einer Berufsschule. Wichtig sei, dass man dazu eben stehe: „Das ist so!“ (G12). Weiters wird die Realitätsnähe aufgrund der Verstrickung von unterschiedlichen Personen mit vielen Postings erwähnt.

In einer einzigen Gruppe mit 15-/16-jährigen Berufsschüler*innen wurde die Szene jedoch als absolut unrealistisch eingeschätzt. Die Burschen meinten, dass es „ganz anders in Wirklichkeit“ zugehe. Man würde nie ein Foto nur als Spaß verschicken, sondern immer aus einem bestimmten Grund und wenn man ein peinliches Foto von jemanden zugeschickt bekommen würde, dann würde man dieses zunächst speichern und zur Verteidigung in Konfliktsituationen verwenden: „Keiner schreibt, ich hab das, das, das von dir sondern man speichert sich vielleicht was, und wenn die Person dich vielleicht angeht, dann wirft man das ein [...] das einzige, was dich sozusagen verteidigt und für nichts rausschmeißt, dann hast das nicht mehr in der Hand [...] und darum geht es, dass man auf der sicheren Seite ist.“ (16-Jähriger, BS, G18), Überdies streite man sich nicht in der öffentlichen Gruppe und wenn doch, dann würden sich in der Realität keine anderen Personen beteiligen, sondern nur die tatsächlich involvierten Personen posten (Täter*in versus Opfer). Nur vereinzelt gab es Teilnehmer*innen, die auf keine eigenen Erfahrungen oder Erlebnissen von Bekannten mit ähnlichen Situationen verweisen können.

Spielszene 3: Fake-Profil

Die letzte gespielte Szene wurde von den Teilnehmer*innen als am glaubwürdigsten eingeschätzt. „So könnte es im Internet ausgesehen haben“, meinte ein 17-Jähriger aus einer AHS-Oberstufe. Von vielen beteiligten Jugendlichen wird die Geschichte, bei dem mittels einer Hate-Seite falsche Inhalte verbreitet werden, als „echt“ eingeschätzt, vor allem wenn man diese mit den vorhergehenden Szenen (Beschimpfung, Nacktfoto) vergleiche. Aber auch hier gab es vereinzelt Gruppenteilnehmer*innen, die diese Situation als unrealistisch sahen und einige, die diesbezüglich keine Erfahrungen gemacht oder beobachtet hatten und deshalb auch nicht abschätzen konnten, ob diese virtuellen Konflikte auch die Alltagswirklichkeit heutiger Jugendlicher darstellen.

(Spiel-)Rollenerleben und Rollenerwartungen

Im folgenden Teil werden die Diskussionsbeiträge der Jugendlichen über ihre eingenommenen Rollen in den jeweiligen Rollenspielen zusammenfassend dargestellt.

Spielszene 1: Beschimpfung

Die Reflexion des ersten Rollenspiels mit den Teilnehmer*innen zeigt deutlich, dass viele Schüler*innen es nicht gewohnt waren, Rollenspiele zu spielen, und Zeit brauchten, sich mit der zufällig gezogenen Rolle zurechtzufinden. Die Ausführungen der Rollenskripte schienen für manche Spieler*innen zu allgemein formuliert zu sein, aber die Neugierde und Lust, mit dem Handy zu spielen, erleichterten den Einstieg. Das von den Jugendlichen gewählte (Konflikt-)Thema wie Klimapolitik, Fußball, Veganismus etc. initiierte

anfangs nur zögerliche Diskussionen, führte aber schließlich in vielen Onlinespielen zu allgemeinen Beschimpfungen, die nichts mehr mit dem Thema zu tun hatten. Meinungsunterschiede konnten nicht mehr wahrgenommen werden und allgemeine Beleidigungen wurden stattdessen gepostet. Die Rollenanweisungen wurden häufig nicht mehr ausgeführt bzw. verließen Jugendliche ihre Rollen und konnten ihre Relationen zueinander (Täter*in, Opfer, Bystander*innen) immer unklarer erkennen. Die Rolle der Täter*innen verschwamm mit der Opferrolle, denn alle schienen sich gegenseitig zu beschimpfen. Diese unerwarteten Dynamiken in vielen Gruppen führten zu unterschiedlicher Wahrnehmung und Ausführung der eigenen Spielrolle und beeinflussten die Rollenspiele wesentlich. Im Folgenden geben wir einen kurzen Überblick zu den unterschiedlichen Workshop-Gruppen auf Grundlage der Diskussionen:

- ⇒ Spielgruppen, die die Online-Auseinandersetzung mit GIFs, Bildern, Emojis führten („es wurde mit GIFs gespart“), eine Fülle von Postings produzierten („jeder hat einfach losgeschrieben), ähnlich einem Ping-Pong-Wettbewerb, und ihre eigentlichen Rollen „nicht checkten“, aber viel Spaß daran hatten.
- ⇒ Spielgruppen, die sich, unabhängig von ihren Rollenanweisungen, gegenseitig scheinbar grundlos („um was geht’s eigentlich“) beschimpften: „alle [sind] aufeinander losgegangen“ (G19) oder „eine Person hat alle beschimpft“ (G12). Viele der Rollenspieler*innen hätten am liebsten die Gruppe verlassen oder wie ein 17-jähriger formuliert: „Ich hab drauf g’schissen“ (G7).
- ⇒ Spielgruppen, die versuchten, ihre Rollen bis zum Abbruch des Spiels zu spielen und sich trotzdem nicht involvierten, weil es für sie nicht in den Kontext passte: „Ich hätte eine gemeinsame Gegeninitiative starten sollen, aber hatte in diesem kurzen Chatverlauf keine Möglichkeit gesehen“ (18-Jähriger, BRG, G11). Viele dieser Teilnehmer*innen fanden es überdies ziemlich „nervig über unnötige Sachen“ zu diskutieren und „über Dinge mit Leuten zu reden, die sich nicht kennen“ (G12).

Die Rollenerfahrungen und -erwartungen der Jugendlichen werden nun im folgenden Teil nach den spezifischen Spielrollen (Täter*in, Opfer, Bystander*innen) differenziert dargestellt:

*a) Die Rolle der (Mit-)Täter*in:*

Auffallend in den 21 durchgeführten Gruppendiskussionen ist, dass die Rolle der Täter*innen sowohl positiv als auch negativ erlebt wurde. Ein Teil der Jugendlichen teilte mit, dass es Spaß gemacht habe, auch die „böse“ Rolle zu spielen, für andere hingegen war die Täter*innenrolle unangenehm und anstrengend, „weil ich das normalerweise nicht mache“ (G20) oder es einfach nicht möglich war „in echt“ zu spielen. Aus der Sicht der Jugendlichen hat die Täter*innenrolle überwiegend gut funktioniert, denn der*die Täter*in und auch der*die Mittäter*in haben jeden Versuch der Deeskalation verhindert und immer wieder mit neuen Angriffen reagiert. So meinte ein Jugendlicher entschuldigend in seiner Täterrolle, dass er so reagieren musste, obwohl er sich im Unrecht fühlte. Eine Strategie war deshalb, die Einwände der anderen zu ignorieren und sich nicht von den eigenen Interventionen abbringen zu lassen (G11). Ein 17-jähriger Berufsschüler meinte dazu: „Es ist so, wenn man aggressiv drauf ist, dann merkt man das nicht, die guten Sachen, man schaut nicht drauf, man erwartet nur von der einen Person [Opfer] die Antwort.“

(G17, 17J, BS). Deshalb hätten auch die Interventionen keinen Effekt gehabt bzw. hätten den Konflikt nicht stoppen können. Es gab aber auch die Sichtweise, dass die Täter*innenrolle eine sehr schwierige Rolle sei, da es im vorliegenden Rollenspiel zu wenig Kontextwissen gab, um jemanden wirklich zu beleidigen. Überdies verliere man den Spaß beim Beschimpfen, so ein 17-Jähriger Jugendlicher (G3). In der Rolle der Mittäterin meinte ein Mädchen: „Ich hab’ mich nicht getraut zu beleidigen“ und gestand sehr offen, dass sie in Wirklichkeit viel „gemeiner“ wäre als jetzt im Rollenspiel (G2, 16J, HAS) – allerdings sahen das, wie oben erwähnt, viele andere in der Täter*innen-Rolle genau umgekehrt.

*b) Die Rolle der Bystander*innen:*

Die Spieler*innen haben ihre Rollen als Bystander*innen überwiegend als sehr anstrengend erlebt. Wer aktiv seine Rolle spielen wollte, musste viel lesen, um dem Verlauf folgen zu können. Es waren sichtlich zu viele Postings: man wäre immer „zu spät gewesen, wenn ich gekontert hätte“ und „man fühlt sich ignoriert“ (G21), so ein Mädchen in der Diskussion. Bystander*innen hatten oft „keine Lust“ zum Schreiben und deshalb sei auch die Intervention nicht gelungen, meinten einige Spielteilnehmer*innen (G1). Auch Hilflosigkeit und Mitgefühl mit dem Opfer machte sich unter den Bystander*innen breit, denn Geschriebenes konnte nicht mehr rückgängig gemacht werden und Interventionen wurden ignoriert und blieben wirkungslos: „Ich hatte nicht das Gefühl, dass meine ‚Beruhigungsversuche‘ wirklich gut funktioniert haben. [...] Sie haben immer weitergemacht mit ihren Hänseleien quasi.“ (15-Jährige, BRG, G11)¹¹. Für die Mobilisierung anderer Unterstützer*innen bräuchte es mehr Kontextwissen und eine längere Auseinandersetzung, so ein Diskutant (18-Jähriger, BRG, G11). Viele Bystander*innen haben nicht reagiert und ein Gymnasiast meinte dazu, dass er sich bei kurzen Streitigkeiten nie einbringen würde – da hätte man halt einen „schlechten Tag“ erlebt, aber bei Mobbing sehr wohl: „Am Anfang war es nicht schlimm, es ist halt so ‚ich mag ihn nicht‘ aber später im Verlauf, ja - weiß nicht ... Das ist halt jetzt ein Chat sozusagen über 20 min aber normalerweise ist so etwas wochenlang und wenn etwas wochenlang so geht, dann berührt es einem viel mehr und es ist intensiver und dann ich würde eher einschreiten... also wenn es länger wäre.“ (17-Jähriger, BRG, G11).

c) Die Rolle des Opfers:

Über die Rolle des Opfers wurde im vorliegenden Rollenspiel wenig diskutiert, jedoch eine interessante Bemerkung geäußert. Ein 16-Jähriger meinte: „Solange ich mich wehren kann, interessiert mich die Hilfe von anderen nicht“ (G2) und zeigt damit unmissverständlich auf, dass er Interventionen von Bystander*innen als Einmischen in eine private Angelegenheit verstehe. Diese Einstellung teilt auch ein anderer Teilnehmer: „Es ist eine Sache zwischen diesen beiden (zeigt auf die Burschen) und andere müssen sich hier nicht einmischen“. (14-Jähriger, FMS, G19).

¹¹ Da die Bystander*innenrolle für die Analyse im besonderen Fokus steht, werden in den angeführten Zitaten auch das Geschlecht, Alter und der Schultyp der Jugendlichen angegeben

Spielszene 2: Nacktbild

Im zweiten Online-Rollenspiel konnte aufgrund des Vorwissens und der Erfahrungen aus dem ersten Rollenspiel viel Unsicherheit abgelegt werden. Die Teilnehmer*innen fühlten sich fast durchgängig beim Spielen sehr wohl und versuchten, ihre zugewiesenen Rollen so zu spielen, wie es ihrem Skript entsprach. Vereinzelt gab es auch Rollen- und Spieldiffusion:

- ⇒ Spieler*innen verloren den Überblick: „Ich bin nicht mitgekommen! [...] Es war ein durcheinander, um ehrlich zu sein!“ (16-Jähriger, BS, G18),
- ⇒ Mittäter*innen fehlte „ihr Opfer“: „Wusste nicht wen ich angreifen soll“ (G 3), „Ich glaube, die, die die böse Rolle hatten, haben den Streit verkannt“ (17-Jähriger, BS, G18),
- ⇒ Bystander*innen identifizierten weder den*die Absender*in des Nacktfotos noch die dargestellte Person (G10, G12)
- ⇒ Mittäter*innen blieben im Verlauf des Rollenspiels unerkant: „Ich dachte jeder ist böse“, (G 18) u.v.m.

Manche Spieler*innen erklärten in den Diskussionsrunden, dass sie im Rollenspiel unbedingt auch posten und engagiert auftreten wollten, obwohl sie normalerweise nicht so handeln würden. Es war einigen trotz „wahnsinniger“ Anstrengung sehr wichtig, sich zu beteiligen und zu intervenieren („Im Spiel war ich immer dabei“), um neue Erfahrungen zu machen: „[Ich wollte] kontern und hab’ aber so was noch nie erlebt“ (G20).

Einige Diskutant*innen schätzten die „Mobber“ im Vergleich zum ersten Rollenspiel „härter“ ein, was vermutlich mit einer geringeren Hemmschwelle im zweiten Durchlauf zu tun hatte. Sie stellten auch eine ziemliche Ausgewogenheit zwischen jenen fest, die sich schützend vor andere Teilnehmer*innen gestellt hatten, und jenen, die andere angegriffen hatten, und erzählten auch von zeitweisem Chaos und gegenseitigen Beschimpfungen (G16). Bis auf eine Gruppendiskussion waren sich alle Spieler*innen darüber einig, dass im zweiten Rollenspiel ein Motiv für den Konflikt (unautorisiertes Nacktfoto) das Eingreifen Anderer rechtfertige. Dagegen stellten etliche 15-jährige Mädchen in einer Diskussionsrunde im Polytechnikum fest: „Nein, es ist ganz normal, dass jemand so was schickt oder Nacktbild oder solche Sachen, das ist lustig“ (G6). Durch ein Foto ließe man sich nicht provozieren, denn das sei Normalität und könnte immer wieder passieren.

*a) Die Rolle der Täter*innen:*

Die (Mit-)Täter*innenrolle war aus Sicht der Workshopteilnehmer*innen gut und leicht zu spielen, wenngleich es schwierig war, die Aufhetzungen aufrecht zu erhalten. So formulierten zwei Burschen in derselben Diskussionsgruppe: „Nach einer Zeit konnte ich nicht mehr mobben!“ und „Ich war auch jemand, der gemobbt hat, und ich habe mich gut dabei gefühlt. Ich wusste aber nicht was ich schreiben sollte“ (G2). In einer anderen Gruppe meinten Mädchen, dass es in der Rolle des Täters bzw. der Täterin eigentlich sehr „amüsant“ und „unterhaltsam“ sei, wenn Bystander*innen sich postend einmischen

würden (G11). Und in einer weiteren Gruppendiskussion stand die Person, die die Täter*innerolle übernahm, vor der Herausforderung, dass sich eine „bestimmte Person“ (= Opfer) durch das Veröffentlichen des unautorisierten Nacktfotos nicht angesprochen fühlte und deshalb den Angriff sehr erschwerte (G6). Es kam auch vor, dass „Leute gemein“ waren, obwohl sie gar nicht die Rolle inne hatten (G2). Die Tatsache, dass sich alle Interventionen in einem Rollenspiel auf die Täter*innen konzentrierten (und nicht auf das Opfer) wurde in einer Diskussionsrunde ebenso erstaunt festgestellt und hinterfragt: Es könnte doch auch anders sein, so die Jugendlichen (G12).

*b) Die Rolle der Bystander*innen:*

Die Rollen der Bystander*innen gestalteten sich im zweiten Online-Rollenspiel wesentlich aktiver und wurden auch anstrengender wahrgenommen als im ersten Spiel. Die Workshopteilnehmer*innen in allen Gruppendiskussionen stellten mehrheitlich fest, dass ihr Intervenieren kaum langfristig zur Beruhigung der Situation und Einsicht der Täter*innen beigetragen habe. Kurzfristige Beruhigung bzw. Unterbrechung der aufgeheizten Konfliktsituation konnte jedoch mit Interventionen wie beispielsweise dem Senden von lustigen GIFs, dem Posten von banalen Themen wie Essen („Ich hab’ Hunger!“) oder dem gezielten Einsatz von Privat-Chat s mit Involvierten erreicht werden. Insgesamt bestand der Tenor der Bystander*innen, die deeskalieren wollten, darin, dass man nicht „gehört“ wurde. Die Vielzahl, Schnelligkeit und Gleichzeitigkeit der Statements im öffentlichen wie auch im privaten Chat auf unterschiedlichen Plattformen konnten laut Spieler*innen die Beschimpfungen und Angriffe nur marginal stoppen. Vielmehr führten diese zahlreichen Interventionen zu mehr Chaos und vor allem zu Unzufriedenheit und Hilflosigkeit bei den Helfer*innen. „Ich wollte auch die Mobber stoppen und so, aber viel konnte ich nicht machen“, meinte ein Mädchen in der Diskussionsrunde (G2). Einige Bystander*innen teilten auch mit, dass sie durch Postings nichts bewirken konnten und deshalb auch die „Hoffnung aufgegeben“ hätten, dass weiteres Einmischen sinnvoll sei. (G20). Die unterschiedliche Einschätzung darüber, ob es sich im vorliegenden Rollenspiel um einen „richtigen“ Konflikt handle, führte auch dazu, dass keine Interventionen gesetzt wurden: „Ich hatte eine gute Rolle, aber ich konnte nichts Gutes machen, weil es war kein richtiger Streit hier, ja also. Ich glaube, die, die die böse Rolle hatten, haben den Streit verkannt.“ (17-Jähriger, BS, G18)

Im zweiten Rollenspiel schien es auch für einige Bystander*innen schwierig zu sein „anständig zu bleiben und nicht zu schimpfen“, denn manchmal wurden auch die Unterstützer*innen nach Erhalt des Nacktfotos in ihrer Wortwahl bei den Interventionen beleidigend „obwohl ich das gar nicht wollte“ (G2). Ein Helfer erklärte seine Ambivalenz im Rollenkonflikt folgendermaßen: „Also ich war einer der Guten, ich hab auch in die Gruppe geschrieben ‚Ich bin stolz auf Euch wir haben gewonnen und letztes [Fußball-]Spiel gut gemacht‘ und dann hab ich auch Feuer gemacht und meine Rolle erfüllt und beruhigt und dann wieder Feuer gemacht und beruhigt.“ (16-Jähriger, BS, G18)

In einigen Workshops erzählten Jugendliche vereinzelt von ihren Befürchtungen, gemobbt zu werden, wenn sie sich aktiv als Bystander*innen engagieren würden (G2, 10, 12 u.a.). Dies Befürchtungen bestätigten sich zum Teil auch in den Erfahrungen im Rollenspiel: Ein Mädchen berichtete beispielsweise, dass es beschimpft und angegriffen wurde, weil es sich für das Opfer eingesetzt habe. (G2). Wichtig sei, so

einige Spieler*innen, sich selbst zu schützen, und das führe in der Realität oft zu Situationen, in die man sich (besser) nicht einmische (G1).

Warum sich einige Bystander*innen trotzdem engagiert im Spiel einmischten, war u.a. auch dem Umstand zu verdanken, dass sie sich aufgrund der Postings des Opfers über die persönliche Enttäuschung und Kränkung bzgl. der Verbreitung des Nacktfotos aufgefordert fühlten, einzugreifen. Zudem hatte sich das öffentliche Bedanken des Opfers bei den Verteidiger*innen über ihre Interventionen wiederum positiv ausgewirkt und das zivilcouragierte Handeln der Bystander*innen verstärkt (G5).

c) Die Rolle des Opfers:

Die Rolle des Opfers wurde unterschiedlich erlebt und kommentiert: So meinten Burschen, wenn ein Opfer – wie im gespielten Rollenspiel – zurückpostet und sich selbst deutlich zur Wehr setzt, dann mache es auch den Angreifern wenig Spaß, weiter zu machen. „Wenn jemand stark ist und etwas zurückschreibt, dann glaub’ ich hören sie eh auf, denn das Ziel von Mobbern ist ja jemanden zu demütigen und der hat immer zurückgeredet.“ (G2) Wenn die attackierte Person Mut zeige, gegen ihre Angreifer poste und sich selbst stark verteidige, brauche es deshalb auch keine unterstützenden Bystander*innen, (G12). Hilfe könne ein Opfer scheinbar nur dann erfahren, wenn es selbst zeige, dass es Unterstützung möchte, so ein Jugendlicher: „Ich glaub’ nicht, dass man reagiert, wenn nicht das Opfer zeigt, dass es beleidigt ist.“ (G2). In einer homogenen Diskussionsrunde äußerte sich ein Jugendlicher, der die Opferrolle spielte, über die Rollenwahrnehmung der Bystander*innen und merkte an, dass es ihm angenehm auffiel, dass sich Bystander*innen für ihn einsetzten: „Es tat gut nach dem sechsten mal ‚Scheiße‘ was Positives zu hören (G3).

Am sinnvollsten scheine es jedoch zu sein, so die Meinung einiger 15-jähriger Mädchen, dass der bzw. die „Gemobbte“ sich selbst nicht in der Rolle eines Opfers sehe und sich auch nicht durch ein Nacktfoto, das „ganz normal“ sei, provozieren lasse: „Es ist ganz normal, dass jemand so was schickt oder Nacktbild oder solche Sachen, das ist lustig“ (G6). In der Realität würde man sich hier nicht einmischen, sondern nur lachen, meinten wiederum andere Mädchen (G2).

Spielszene 3: Verbreitung falscher Inhalte über eine Person

Die Durchführung des letzten und dritten Rollenspiels verlief in den Workshops überwiegend routiniert in Bezug auf Rollenübernahme und Spielverlauf. Die Rollenspieler*innen wirkten sicherer und zielgerichteter in der Umsetzung ihres Rollenskripts. Demensprechend kurz war auch die Anlaufzeit des Konflikts durch die Verbreitung eines Fake-Bildes samt Anfeindungen eines Rollenspielteilnehmers in den sozialen Medien und die Reaktionen durch zahlreiche Gegenreden. Auffallend ist, dass mit zunehmender Spielerfahrung und Reflexion über die Spielrollen in den Diskussionsrunden auch unterschiedliche Interventionsstrategien aus den Rollenanweisungen von Bystander*innen verstärkt ausprobiert wurden. Zum Beispiel: Unterstützer*innen für eine gemeinsame Strategie mobilisieren, Opfer helfen/trösten, (Mit-) Täter*innen zur Einsicht bringen und Angriffe stoppen, Entschuldigung seitens von Täter*in einfordern usw.

*a) Die Rolle der (Mit-)Täter*innen:*

Die Umsetzung der Rolle des (Mit-)Täters bzw. der (Mit-)Täterin wird im Allgemeinen als gut und gelungen eingeschätzt obwohl hier unterschiedliche Erfahrungen in den Rollenspielen gemacht wurden. Interessant bei diesem dritten Rollenspiel sind vor allem folgende Aussagen über das Rollenerleben des*der (Mit-)Täter*innen: So resümieren einige Jugendliche, dass sie ihre „böse“ Rolle eine bestimmte Zeit lang leicht und unbeirrt spielen konnten, indem sie ihre Online-Mitspieler*innen sehr gut provozieren konnten. Gegen Ende des Spiels fiel jedoch das Durchhalten „richtig schwer“ und erforderte große Anstrengung, denn die Argumente der Bystander*innen wurden zunehmend besser: „Da konnte man nicht allzu viel sagen, weil man weiß, dass sie eh recht hatten“ (G11).

In einigen Online-Rollenspielen kippte die Situation schließlich zu Ungunsten der Angreifer*innen und die Täter*innenrolle konnte nicht mehr wortgewaltig ausgeführt werden: „Ich hab meine Rolle nicht mehr geschafft!“ (G17). Helfer*innen stellten sich gemeinsam gegen die mobbende Person und erreichten damit, dass jene, die sich in der Rolle der Täter*in sahen, nun selbst als Opfer fühlten: „Alle sind auf mich losgegangen“ (G20), „obwohl ich die Böse war, wurde dann ich gehänselt“ (G17). Durch die Gegenrede der Bystander*innen fühlten sich schließlich die Spieler*innen in der Angreifer-Rolle in ihrer Selbstwahrnehmung „schlecht“. Der/die Täter*in hätte es auch verdient sich schlecht zu fühlen, kommentierten die Jugendlichen in ihrer Rolle als Bystander*innen (G20). Sichtlich macht es mehr Spaß, die Rolle des (Mit-)Täters bzw. der (Mit-)Täterin auszuführen als andere Rollen zu spielen, solange man sich dabei überlegen fühlt und sich keinen Gegenreden von Bystander*innen stellen muss. Die Umsetzung der Rolle des Mittäters bzw. der Mittäterin wird als gut gelungen eingeschätzt (G9).

*b) Die Rolle der Bystander*innen:*

Mehrheitlich berichteten Jugendliche in den Diskussionsrunden, dass sie unabhängig von ihren je spezifischen Rollenbeschreibungen die Position als Bystander*innen überwiegend unbefriedigend und schwierig erlebten. Ein Opfer zu unterstützen sei richtig anstrengend, äußerten häufig Diskutant*innen, und ein Mädchen formulierte: „Ich hab’ Kopfschmerzen bekommen!“ (G20).

Begründet werden diese frustrierenden und belastenden Rollenerlebnisse damit, dass

- ⇒ die unterschiedlichen Interventionen (öffentlich und privat) kaum Wirkung zeigten,
- ⇒ die Strategien der Unterstützung nur in Ansätzen umgesetzt wurden,
- ⇒ die vorhandenen Handlungsmöglichkeiten sehr eingeschränkt erlebt wurden,
- ⇒ keine Ideen für gerechte Konfliktlösungen gefunden werden konnten,
- ⇒ die Angriffe auch mit guten Argumenten nicht gestoppt werden konnten,
- ⇒ schließlich der gemeinsame Rückzug aller Bystander*innen als die beste Deeskalationsstrategie gesehen wurde.

Vereinzelt berichten Teilnehmer*innen darüber, dass die „Guten und die Gerechteren“ in ihrem Rollenspiel mehr und mit besseren Argumenten als die Angreifer*innen agierten, jedoch trotz geschickter Taktik die Übergriffe nicht stoppen bzw. die Täter*innen zur Einsicht bewegen konnten (G11). Scheinbar gelingt es

nur ganz selten, in der Bystander*innen-Rolle mit Argumenten oder anderen Maßnahmen wie der Nutzung des Privat-Chats deeskalierend zu wirken. Der Tenor der Diskutant*innen lautet dementsprechend: „Ich wollte sie schützen, aber es ist nicht gelungen“ (G14). Oder wie ein 16-jähriger Bursche formuliert: „Man kann nicht so als 3. Person zwei die sich gegenseitig nicht mögen [beruhigen], weil die haben nicht nur Konflikt im Chat, online, sondern auch früher [...] vor allem das mit dem Bild, da kann nicht ein Dritter dazwischenfunken, das geht nicht, das funktioniert meistens gar nicht! (G18).

Eine wesentliche und positive Erfahrung als Bystander*innen erzählte ein 16-jähriges Mädchen in einer Berufsschule: „Ich find das eigentlich voll arg, weil ich hab mich so in der Position quasi zu dem Opfer gesetzt, also ich hab ihn dann voll in Schutz genommen ... und mich hat's gewundert, dass der Mensch der es [das Opfer] eigentlich reizen sollte oder Menschen mobben sollte, nicht auf mich losgegangen ist sondern immer noch auf dieselbe Person anstatt er gleich zwei nimmt und die dann noch mehr runtergemacht werden.“ (G12). Diese Erwartungshaltung, dass bei Unterstützung des Opfers der*die Täter*in „auch auf mich losgehen“ wird, wird aber nicht nur wie im Zitat widerlegt, sondern auch von einigen Bystander*innen durch die konkreten Erfahrungen im Rollenspiel bestätigt (G14).

Die große Herausforderung der Bystander*innen -Rolle liege vor allem darin, gerechte Lösungen für alle zu finden, d.h. auch den*die Täter*in nicht zu verletzen und „das ist viel schwieriger als die böse Rolle zu spielen“, so die Diskutant*innen einer Workshopgruppe (G21). Die Schwierigkeit für Bystander*innen bestehe in der Solidarisierung mit dem Opfer einerseits und in der Anforderung, gegen die Angriffe der Täter*innen zu agieren, ohne selbst übergriffig zu werden. Der Ausspruch: „Ich blieb anständig“ (G10), verweist sehr treffend auf den persönlichen Anspruch, dem Bystander*innen manchmal im Rollenspiel nicht gerecht werden konnten: „Ich war nett, aber weil mein Bruder fertiggemacht wurde, habe ich dann auch geschimpft!“ (G17).

Jugendliche meinten in einer anderen Diskussionsrunde auf die direkte Frage, was Täter*innen, umstimmen könnte, den Konflikt zu beenden, dass es „Nichts“ gäbe, Jugendliche aus ihrer Täter*innenrolle zu holen, wenn sie nur auf die eigenen Postings fokussiert seien (G21). Als eine Konsequenz dieser Einstellung kann die Strategie des sich nicht Einmischens oder des rechtzeitigen Rückzugs aus der Konfliktsituation bedeuten: „Man will selber keinen Stress haben und hält sich da raus“ (G18), obwohl es bei manchen Bystander*innen ein Unwohlsein zurücklässt: „Es hat mir richtig weh getan nicht helfen zu können“ (G20).

c) Die Rolle des Opfers:

Ein Blick auf die Rollen des Opfers zeigt, dass das Rollenerleben ganz unterschiedlich empfunden wurde: Jugendliche erzählten, dass sie sich in der Rolle als Opfer sehr hilflos, ohnmächtig und handlungsunfähig gefühlt hätten, da sie keine Ideen über mögliche Handlungsoptionen hatten. (G4). Daher wurde es auch als sehr positiv von den Opfern erlebt, dass Bystander*innen statt ihrer posteten bzw. sich für sie

einsetzten: „Gut, dass andere geschrieben haben“ (G4) oder „Es war schön, dass sich jemand für mich eingesetzt hat“ (G19). Es gab aber auch in den Rollenspielen Personen, die sich nie als Opfer fühlten und sich gegen eine Zuschreibung eines Opferstatus gewehrt hatten. Auch wenn Bystander*innen und Angreifer*innen in diesem Fall eine andere Sichtweise hätten, wäre dies für die betroffenen Personen irrelevant bzw. uninteressant (G6). So meinte beispielsweise ein Mädchen in diesem Zusammenhang: „Es juckt mich nicht als Opfer“ (G20) und führte weiter aus, dass es ihr auch keinen Spaß machen würde, auf Beschimpfungen mit eigenem Schimpfen zu reagieren. Ein anderer Jugendlicher in der Rolle des Opfers meinte wiederum, dass auf dem Foto kaum jemand zu erkennen war und fragte ganz generell in die Gruppe: „Welches Foto?“ Es scheint so, dass dieser Person sichtlich entgangen war, dass hier eine Gruppe von Online-Rollenspieler*innen infolge eines geposteten Fake-Bildes samt persönlicher Anschuldigungen mit unterschiedlichen Zielen Postings austauscht (G19).

In den Diskussionsrunden werden auch unangenehme Erlebnisse in der Opferrolle thematisiert. Ein Jugendlicher teilte beispielsweise seine schlimmste Erfahrung in der Gruppe mit: „Wenn die ganze Gruppe auf eine Person geht, das ist das Schlimmste“ (G12). Inhaltsleere Beleidigungen, also Beschimpfungen ohne Grundlagen, werden aus Sicht eines 16-jährigen Mädchens weniger problematisch und schlimm erlebt als beispielsweise die direkte Aufforderung „Heul’ nicht!“. Diese Erfahrung machte jedoch ein Mädchen in ihrer Spielrolle. Deshalb sei es auch besser, „sich ganz rauszuziehen als Opfer“ und selbst nichts zu posten und auch Beiträge von anderen zu ignorieren: „Ich hab’ die ganze Zeit versucht - ich war die, die angegriffen wurde - und ich wollte nicht zu viel Kontra geben, denn es ist nur schlimmer und schlimmer geworden, je mehr Bühne wir den anderen gegeben haben.“ (G11). Schlussfolgernd sei es aus Sicht der Gruppenteilnehmer*innen für Opfer und Bystander*innen sehr wichtig, dass man den Angreifer*innen so wenig Gelegenheit wie möglich bietet, um zu reagieren bzw. „ohne Nachdenken zu posten“ (G11).

Wahrgenommene Wirksamkeit der Handlungen und bestmögliche Interventionen

Im folgenden Teil sollen die impliziten Erkenntnisse zur Wirksamkeit der ausprobierten Handlungen, die teilweise schon im vorhergehenden Kapitel angedeutet wurden, und die Einschätzung von bestmöglichen Interventionen aus Sicht der Rollenspieler*innen dargestellt werden. Auffallend ist, dass sich die Interventionsstrategien und deren Wirksamkeit innerhalb einer Gruppe, sofern sie auch von den Diskutant*innen in die Gruppenreflexionen eingebracht wurden, nur gering unterscheiden. Das würde bedeuten, dass die Spieler*innen ihr Repertoire an Handlungsmöglichkeiten mit einigen Ausnahmen weitgehend unabhängig von den unterschiedlichen Spielgeschichten (Beschimpfung, Nacktfoto, Fakebild zur Verbreitung von falschen Inhalten) einsetzten. Eine Darstellung der wirksamsten Strategien differenziert nach Spielszenen scheint deshalb nicht zielführend zu sein. Vorweg soll noch festgehalten werden, dass die Workshop-teilnehmer*innen nur begrenzte Zeit hatten, in der Gruppe zu erzählen und

viele auch nur wenig über die Wirksamkeit ihrer Handlungen einbringen konnten oder wollten. Der folgende Überblick konzentriert sich auf jene Gesprächsinhalte, die auch explizit in die Diskussionsrunden eingebracht wurden.

Begründungen für das Scheitern der Interventionen

Ganz allgemein kann festgehalten werden, dass die unterschiedlichen eingesetzten Interventionsstrategien fast durchgängig in ihrer Wirksamkeit scheiterten. Auf die Frage, wie wirksam die Rollenspieler*innen ihre Interventionen – im Sinne von Deeskalation und Beruhigung des Konflikts – einschätzen, erhielt das Forschungsteam fast nur negative Rückmeldungen:

„es hat gar nichts gewirkt“, „nicht wirklich gewirkt“, „nein, man konnte nicht stoppen“, „nicht geklappt“, „es hat gar nichts gewirkt, beruhigen und stoppen ist nicht gelungen“, „verteidigen hat nichts gebracht“, „wirksam nicht, aber lustig“, „provozieren schon, aber beruhigen geht nicht“, „es bringt sich nicht mehr“, „beruhigen ging nicht“, „keine Wirkung zu helfen“, „nicht wirksam“, „beschützen hat nicht funktioniert“, „es gibt keine wirkungsvolle Intervention“, „hat nichts genützt zu sagen hör’ auf“, „es war komplett egal, was ich geschrieben habe“, „ich wurde die ganze Zeit ignoriert“, „unmöglich“, „es ist egal, was man schreibt“, „n e i n“, „es hat nichts wirklich geholfen“, „alles abgeprallt“, „hat ein bisserl was gewirkt (grinsen)“, „nur für kurze Momente wirksam, aber dann doch nicht“, u.v.m.

„Ich hatte nicht das Gefühl, dass meine ‚Beruhigungsversuche‘ wirklich gut funktioniert haben“.

Zumeist wird von den Diskutant*innen mit der Einschätzung der (Un-)Wirksamkeit ihrer Handlungen eine – oft auch entschuldigende – Erklärung mitgeliefert. Ein differenzierter Blick auf die Erklärungsversuche der Jugendlichen, warum die Interventionen der Bystander*innen nicht gewirkt haben, zeigt die folgende Bündelung der Aussagen:

⇒ Im Internet wird man nicht ernstgenommen!

Eine mehrfach geäußerte Überzeugung illustriert eine Grundüberzeugung einiger Jugendlicher: Man werde im Internet nicht ernstgenommen und darum könne man auch kaum etwas bewirken. Demzufolge seien auch Beschimpfungen und dergleichen nicht ernst zu nehmen (G7). Diese Einstellung lässt aber zugleich den Umkehrschluss zu, dass Anschuldigungen, Beschimpfungen und ähnliches grundsätzlich nicht ernstgenommen werden sollten, weil im Internet vieles aus Spaß gepostet werde: „Und ich bin der Meinung, ganz ehrlich, man sollte so was n i e ernstnehmen und wenn man das tut dann ist man absolut selber schuld und dass man dann selber so blöd war – das Internet ist nicht deine Familie, deine Freunde, das reale Leben!“ (18-Jähriger, BS, G7)

⇒ Einmischen von außen durch Dritte ist ein No-Go!

Die Diskutant*innen sind davon überzeugt, dass bei einem Konflikt in den sozialen Medien eine Intervention „von außen“ nicht hilfreich ist. Es könne nicht wirksam sein, wenn sich jemand von außen

einmischt und sagt: „Lass das!“ (G1). Ganz im Gegenteil bestünde dabei die Gefahr, dass der Konflikt dadurch weiter eskaliere, denn er werde durch Interventionen nicht wirklich gestoppt, wenn sich Leute einmischten, „es [der Konflikt] wird dann halt zwischen zwei Gruppen geführt“ (G11). Als eine zentrale Voraussetzung für eine Einmischung, wenn überhaupt, wird gesehen, dass sich das Opfer selbst meldet und um Unterstützung bittet. Wenn das nicht gegeben ist, gelingt auch eine wirksame Hilfestellung nicht.

⇒ Postings sind nur begrenzt einsetzbar!

Insbesondere das Posten längerer Textnachrichten erscheint den Diskutant*innen als einschränkend und damit nicht zielführend. Es gebe keine anderen Möglichkeiten als Meldungen wie „schäm’ dich, hör auf“ oder „Stop“ zu schreiben (G10). Differenzierende Texte werden bei Online-Konflikten nicht wahrgenommen, weil „die schreiben weiter so frech!“ (G17). Letztlich sei es egal, was man schreibe, es sei einfach nicht möglich, jemanden im Chat zu einer Entschuldigung zu bewegen: „das ist ein Ding der Unmöglichkeit“ (G18). Es sei auch nicht möglich, aufgeheizte Situationen über das Internet zu beruhigen. Denn „Chatten ist was anderes als reden, deshalb lass ich es“ (G20), im Chat werde nur auf die eigene Position fokussiert und nicht versucht, sich auf die anderen Postings einzufühlen bzw. darauf Bezug zu nehmen. Zudem stellten einzelne Jugendliche fest, dass man als Bystander*innen in einem Online-Konflikt auch deshalb nicht wahrgenommen werde, weil Täter*in und Opfer in ihrer Interaktion scheinbar aufeinander fixiert seien und Postings Dritter ignoriert würden.

⇒ Zu viele Postings erzeugen Chaos!

Erfahrungen der Teilnehmer*innen zeigten, dass ein öffentliches Posten „großes“ bzw. „zu viel Chaos bringe“ und letztlich gar nicht wahrgenommen werde (G8 u. G11). Vielmehr berichten einzelne Spielteilnehmer*innen davon, dass der Online-Konflikt auch durch beruhigende Worte eher stärker geworden sei, was sie zur Erkenntnis gebracht habe, es sei „besser sich nicht einzumischen“ (G4). Erschwerend für eine Deeskalation sei es auch, wenn zu viele Leute in der Gruppe sind, die gerade mit einem Konflikt beschäftigt ist. Es bestehe dann die Gefahr, dass sich alle irgendwie daran beteiligen und deshalb zu viele Postings abgesetzt werden, die nicht gelesen werden, entweder aufgrund der Menge an Postings oder weil „die Leute trauen sich alles zu sagen – gegenüber jeden!“ (G7) und deshalb „[...] bringt’s nicht in Gruppen zu schreiben“ (G20).

⇒ Selbstschutz und Eigenverantwortung ist Aufgabe aller Internet User*innen!

Ähnlich auffallend ist die fast durchgängige Einstellung unter den Jugendlichen, dass jede*r für das eigene Tun – möglicherweise auch das eigene Leid – letztlich selbst verantwortlich sei: „Wenn man sich was zu Herzen nimmt, ist man selber schuld (Gestik mit Hand auf Kopf)“ (G7). Ein Anspruch auf individuelle Souveränität scheint hier im Vordergrund zu stehen. Dieser führt in letzter Konsequenz dazu, dass man erst dann zum „Opfer“ wird bzw. werden kann, wenn man diese Rolle tatsächlich übernimmt.

⇒ Online-Konflikte sind offline zu lösen!

Einhellige Überzeugung der Diskutant*innen ist es, dass das Chatten sich stark vom persönlichen Gespräch mit dem*der Konfliktparter*in unterscheidet, besonders bei großen Schwierigkeiten. Die Schlussfolgerung für viele jugendliche Bystander*innen: „... deshalb lass ich es!“ (G20). Konsequenterweise erübrige sich aus Sicht der Jugendlichen demnach ein couragiertes Eintreten als (Online-) Bystander*innen. Denn in erster Linie müssten sich die direkt Involvierten um eine Lösung ihres Konfliktes kümmern – und das gelinge vielfach nur über direkte Kontaktaufnahme, persönliches Treffen, reden, abklären. Wenn die angreifende Person fremd und Kontaktaufnahme nicht möglich sei, dann bestehe die Lösung vor allem darin, sie oder ihn konsequent zu ignorieren, nicht ernst zu nehmen und weitere Kontaktaufnahmen durch Blockieren oder Ausstieg aus der (WhatsApp)Gruppe zu verhindern.

Erfahrungen von wirkungsvollen Interventionen

Die Rollenspieler*innen berichten neben dem Bedauern über ihre überwiegend wirkungslosen Interventionen auch, dass ihre Handlungen punktuell und kurzfristig positive Wirkung erzielen konnten und nennen dabei folgende relevante Aspekte:

⇒ Parteinahme für das Opfer

Einige Bystander*innen meinen, dass es ihnen gelungen sei, ihre Anteilnahme mit dem Opfer zu zeigen und damit auch die Person zu trösten und mental zu unterstützen. Das gelang sowohl durch Postings im öffentlichen als auch im privaten Chat, wie ein 18-Jähriger mitteilt: „Ich war halt nett, ich wollte der Person helfen und ich denke, dass das – also ich hab’ im Privat-Chat geschrieben und ich glaub’, dass das schon ganz gut war.“ (G11).

⇒ Direkter Kontakt mit (Mit-)Täter*innen im nicht-öffentlichen Raum

Die Rollenspieler*innen erzählen, dass die Kontaktaufnahme mit der angreifenden Person im Privat-Chat sowie ein Telefonanruf die Kommunikation vereinfachte und überzeugende Argumentation ermöglichte, um den*die Täter*in zu stoppen und zu einer Entschuldigung zu bewegen. Konkretes Nachfragen seitens der Moderation in der Gruppendiskussion zeigte auch auf, dass dafür in den Telefongesprächen auch Drohungen eingesetzt wurden, wie beispielsweise Andeutungen über mögliche Folgen für die „kleine Schwester“ (G17) oder Warnungen wie „wenn du nicht sofort aufhörst, dann erzähle ich das ...“ (G14). Diese Drohungen und Einschüchterungsversuche werden als legitimes Mittel betrachtet.

⇒ Gutes Zusammenspiel von Bystander*innen durch Absprachen im Privat-Chat

Im Rollenspiel konnten sich die „Guten“ im Privat-Chat absprechen und koordinieren (G21), was es ihnen auch ermöglichte, sich im öffentlichen Chat gut aufeinander beziehen zu können: „Andere Personen haben auf meine Interventionen reagiert bzw. haben auf mich reagiert.“ (G5). Durch diese Absprachen und Abstimmungsprozesse war es den Bystander*innen möglich, sich mit den Inhalten der Angriffe genauer

zu beschäftigen und gezielt dagegen argumentieren zu können (G11). Zudem wurden Beschimpfungen mit Gegenangriffen und guten Begründungen teilweise entschärft und nicht mehr ernst genommen (G16).

⇒ Gezielter Einsatz von Ablenkungs- und Abwehrstrategien

Die mitgeteilten Erfahrungen der Jugendlichen in den Rollenspielen verweisen darauf, dass es den Bystander*innen durch unterschiedliche Methoden zeitweilig gelungen ist, die Eskalation des Konflikts durch das öffentliche Posten böser Beleidigungen zu entschärfen. So wurden beispielsweise lustige GIFs oder andere Bilder versendet, um punktuell den Ernst rauszunehmen (G21). Mehrmals wurde die Taktik des Zu-Spamens mit „STOPSTOP-STOPSTOP...“ verwendet, um die Beschimpfungen eine Zeitlang zu blockieren (G10). Unterstützer*innen kommunizierten sehr klar und unmissverständlich, dass diese Beschimpfungen unerwünscht sind, und forderten dies im öffentlichen Chat wiederholt mit „Hört auf!“ ein (G6). Ebenso wurde inmitten eines eskalierenden Streits versucht, bewusst mit Fragen wie „Gemma ins Kino?“ oder Fragen zum Thema Essen – weil ein Bystander*innen inzwischen hungrig wurde (G21) – paradox zu intervenieren und damit abzulenken.

⇒ Lautstarke Selbstverteidigung als Opfer

Auch die Rolle des Opfers ist offenbar entscheidend dafür, ob ein Konflikt eskaliert oder im Keim erstickt wird. So meinen einige Diskutant*innen, wenn eine Person ihre zugewiesene Rolle als Opfer zurückweise und sich stark zur Wehr setze (G2) und „Mut hat alleine dagegen zu reden“ (G12), dann könne das auch eine positive Wirkung erzielen. Die Erfahrungen in einem konkreten Rollenspiel zeigen, dass die richtigen Worte in der Gegenrede des Opfers wirken (G4) können. Wenn das sogenannte Opfer auf Angriffe gar nicht reagiere und sich um die Beschimpfungen oder Verleumdungen nicht „kümmert“, haben Täter*innen auch kein Feedback und Angriffe gehe ins Leere. (G11)

Bestmögliche Interventionen aus Sicht der Jugendlichen

Was sind nun die bestmöglichen bzw. wirkungsvollsten Handlungen im Internet aus Sicht der Jugendlichen? Die Diskussionen dieser Fragestellung am Ende der rund 60 Gesprächsrunden offenbarten sehr deutlich, dass die rund 170 Workshopteilnehmer*innen kaum auf eigene Selbstwirksamkeits- bzw. positive Wirkungserfahrungen im Umgang mit Konflikten in Sozialen Medien zurückgreifen können und wenig Handlungsoptionen im Internet sehen.

Als erste Priorität gilt für die beteiligten Jugendlichen die persönliche Kontaktaufnahme. Wichtig sei, dass man mit jenen Personen, die übergriffig und verletzend im Internet agieren, direkt Kontakt aufnehme und persönlich redet oder telefoniert. Auch wenn man sich nicht kennt, also sich einander ‚fremd‘ bzw. nicht befreundet sei, solle man trotzdem versuchen, sich zu treffen oder privat anzurufen – auf keinen Fall jedoch öffentlich lange „herumchatten“ bzw. posten. Das private Anschreiben und die face-to-face Problemlösung scheint aus Sicht der Diskutant*innen die wirkungsvollste Strategie zu sein.

Eine weitere Möglichkeit wird in der Mobilisierung von Freund*innen gesehen. Personen aus dem Freundeskreis zu bitten, sich im öffentlichen Chat einzubringen und zu zeigen, dass eine Gruppe von Menschen hinter der angegriffenen Person steht und sich stark macht für jemanden, wird ebenso als eine mögliche wirksame Strategie gesehen.

Eine weitere unkonventionelle Idee bzw. Erfahrung bringt ein Mädchen ein und bezieht sich hier vor allem auf schambesetzte Situationen (z.B. Nacktfoto, Geheimnis). In solchen Situationen erzähle sie eine eigene Geschichte, die diese Peinlichkeit in Relation setzt und damit auch für das Opfer aushaltbar macht: „Ja, aus meiner Sicht, (...) auch irgendetwas Peinliches von mir zu erzählen ... dann ist die Person nicht die, die sich geniert, sondern auch ich und dann sind wir schon zwei und - weiß nicht - vielleicht machen das andere auch und dann ist es vielleicht eine ganze Gruppe und uns ist das dann auch wurscht und der Mobber ist dann der Dumme!“ (16-Jährige, BS, G12)

Zuletzt wird neben bekannten Strategien wie Blockieren, Melden oder Ausschluss aus der Gruppe auch die Strategie des Ignorierens erwähnt, die als Königsdisziplin bei jugendlichen User*innen weit verbreitet zu sein scheint. Hier geht es darum, negative Kommentare, peinliche Fotos oder Fakes einfach nicht ernst zu nehmen: „einfach ignorieren, wäre das Beste“ (G8) oder wie ein 15-Jähriger meint: „Am besten aus der Gruppe kicken, ... ignorieren ... blockieren“. (G18)

1.3 Zusammenfassung und Schlussfolgerungen

Mit Hilfe von Online Rollenspielen wurde das spontane Counter Speech Verhalten von Jugendlichen in konkreten Online-Interaktionen erfasst. Zu diesem Zweck wurden mit Hilfe von Social Media Fake-Profilen typische Übergriffsszenarien in einem kontrollierten Setting simuliert und dafür ein aufwändiges Umsetzungsverfahren entwickelt, inklusive der Ausarbeitung impulsgebender Szenarien und Rollenbeschreibungen, sowie unter Berücksichtigung sicherheitsrelevanter und ethischer Aspekte. Insgesamt wurden 21 Online-Rollenspiel Workshops mit rund 170 14- bis 18-jährigen Jugendlichen in Wien durchgeführt, die sich nach Geschlecht, Bildung und Migrationshintergrund unterschiedlich zusammensetzten (Erhebungszeitraum: Oktober bis Dezember 2019).

Das Analyseziel war es, festzustellen, wie sich das spontane Interaktionsverhalten von Jugendlichen in der Begegnung mit Online-Übergriffen typischerweise gestaltet und wie verschiedene Counter Speech Strategien von Mädchen und Jungen in ihrer Rolle als Online-Bystander*innen eingesetzt werden.

Die Ergebnisse zeigen zunächst ein breites Repertoire an unterschiedlichen Counter Speech Strategien, deren Großteil auf Täter*innen fokussiert. Dazu zählen etwa Warnungen vor möglichen Konsequenzen, Abwertungen von Täter*innen inklusive aggressiver Gegenattacken, Aufforderungen aufzuhören, Forderungen nach sachlichen Argumenten und Fakten, moralische Appelle, Delegitimierungen von Übergriffen des Täters bzw. der Täterin durch die Verteidigung von Opfern, Deeskalations- und Ablenkungsversuche sowie Versuche der direkten Kontaktaufnahme zur Konfliktklärung. Counter Speech Strategien, die sich gezielt an das Opfer wenden, nehmen diese öffentlich in Schutz und entziehen damit Täter*innen oft gleichzeitig die Legitimationsgrundlage getätigter Übergriffe, bringen die Solidarität mit dem Opfer zum Ausdruck oder sprechen dem Opfer Trost und Beratung zu, wofür vor allem die Privat-Chat-Funktion genutzt wird. Counter Speech Strategien, die sich an alle Beteiligten richten, versuchen entweder, die Situation insgesamt zu deeskalieren, z.B. auch durch Stimmungsauflockerungen, oder schlagen konkrete Möglichkeiten vor, wie einen klaren Schlusspunkt zu setzen, indem z.B. konkrete Entschuldigungen eingefordert werden. Teilweise wird auch ein Schlusspunkt erzwungen, indem z.B. die gesamte Kommunikation durch zum Teil massive Ablenkung oder Zu-Spammen durchbrochen wird. Counter Speech Strategien, die sich konkret an andere Bystander*innen wenden, kommen nur selten vor, indem z.B. andere Counter Speaker*innen durch Likes oder kurze Statements unterstützt werden oder Absprachen im Privat-Chat getroffen werden, um ein gemeinsames Vorgehen zu besprechen oder um Unklarheiten abzuklären.

Die Bandbreite an möglichen Counter Speech Strategien wird von Jugendlichen sehr unterschiedlich eingesetzt. Zum einen ist dies von der Gestaltung der Übergriffe selbst bzw. vom Spielkontext abhängig (z.B. gemäßigte vs. sehr ausfällige Übergriffe). Zum anderen zeigen sich aber auch deutliche geschlechts-

und bildungsbezogene Unterschiede sowie Unterschiede zwischen Jugendlichen mit und ohne Migrationshintergrund.

Online-Rollenspiel-Teilnehmer*innen aus der **AHS-Oberstufe** werden zwar mit deutlichen Übergriffen, aber im Vergleich zu den übrigen Gruppen in einem verhältnismäßig gemäßigten Ton konfrontiert. Männliche Teilnehmer reagieren darauf eher mäßig engagiert, tendenziell eher herablassend, werfen Täter*innen unreifes Verhalten vor und drücken ihre Genervtheit mit sanft disziplinierenden bis leicht aggressiven Aufrufen zur Beendigung der Übergriffe aus oder verwenden Ablenkungsversuche. Burschen mit Migrationshintergrund bezeichnen das Verhalten von Täter*innen auch als unehrenhaft. Ein öffentlicher Einsatz für das Opfer wird nicht sichtbar, dies findet lediglich im Privat-Chat statt. Weibliche Spielerinnen setzen hingegen eher auf das Einmahnen der Wahrung der Gesprächskultur und zivilisierter Umgangsformen und auf das Aufzeigen von Regelverstößen. Mädchen mit Migrationshintergrund bringen moralische Aspekte stärker ein und drücken dabei ihre Empörung über das Verhalten von Täter*innen teilweise auch mit einer sehr expliziten Sprache aus, die aufgrund ihrer Emotionalität auch angriffig wirkt. Weibliche Counter Speakerinnen wirken im Vergleich zu männlichen Counter Speakern engagierter und zeigen ein breiteres und vielfältigeres Interventionsrepertoire. Auch die Strategie der Unterstützung und Solidarisierung mit dem Opfer ist bei Mädchen häufiger zu finden, wenn auch Täter*innen-bezogene Interventionen klar überwiegen.

Die Rollenspiel-Workshops, die in **Handelsschulen / Handelsakademien** umgesetzt wurden, fallen von der Inszenierung her im Vergleich zu AHS-Oberstufen sprachlich deutlich abfälliger aus. Vor diesem Hintergrund sind männlicher Teilnehmer zunächst damit beschäftigt, sich Gehör zu verschaffen, und mahnen dies mit teils aggressiven Kraftausdrücken ein oder versuchen, die Kommunikation mit Ablenkungstaktiken zu durchbrechen. Haben sie Gehör gefunden, zeigen sie sich Täter*innen gegenüber aber auch durchaus diplomatisch und lösungsorientiert. Gegenüber Opfern scheinen am ehesten Burschen mit Migrationshintergrund eine Art Beschützerrolle einzunehmen. Weibliche Teilnehmerinnen mahnen zur Beruhigung und zur Wahrung der Gesprächskultur, um gehört zu werden. Gegenüber Täter*innen drücken sie vor allem ihr Unverständnis für deren Übergriffe aus und machen auch unter Einsatz moralisierender Elemente klar, dass ein solches Verhalten nicht zu akzeptieren sei. Sie fordern die Täter*innen teils energisch auf, diese zu beenden und sich zu entschuldigen. Mädchen mit Migrationshintergrund wirken dabei auch durch ihren Sprachgebrauch angriffiger und aggressiver. Es treten aber nicht alle Mädchen energisch auf, einige scheinen durch die ausfällige Art der Inszenierung auch überfordert zu sein und reagieren kaum oder gar nicht.

In den **Berufsschulen** werden die gespielten Übergriffe noch ausfälliger gestaltet und mit zahlreichen vulgären Ausdrucksweisen gespickt, die vor allem von männlichen Jugendlichen eingebracht werden, oft verstärkt durch derb-belustigende oder sexualisierte GIFs. Männliche Rollenspiel-Teilnehmer reagieren darauf, indem sie entweder versuchen, jegliche Kommunikation durch zum Teil ebenso derb-humorvolle

Ablenkungsmanöver zu unterbrechen, oder indem sie gegenüber Täter*innen mit aggressiven und eindringlichen Gegenattacken und Einschüchterungsversuchen auftreten. Auffallend ist aber auch ein relativ hohes Engagement zur Unterstützung und Stärkung des Opfers, um das Machtungleichgewicht zugunsten des Opfers auszugleichen. Weibliche Spielerinnen reagieren auf die Übergriffe mit teils aggressiven, teils lustlosen oder genervten Beruhigungsversuchen und bauen ihre Interventionsstrategien auf Gegenangriffen gegenüber Täter*innen durch Hinweise auf. Sie verwenden z.B. das Aufzeigen rechtlicher Konsequenzen oder moralisch und normativ aufgeladene Zurechtweisungen bei gleichzeitiger Verteidigung des Opfers. Sie versuchen, sich gegenüber Täter*innen als stark zu präsentieren. Mädchen mit Migrationshintergrund wirken auch hier durch ihren Sprachgebrauch häufig angriffiger und aggressiver.

Bei den **Polytechnischen Schulen** weisen die Online Rollenspiele ebenfalls sehr ausfällige Übergriffe auf, mit einer insgesamt aggressiven und derb-humorvollen Grundstimmung inklusive obszöner Beschimpfungen. Männliche Jugendliche versuchen zwar, an die Einhaltung von Normvorgaben zu appellieren, formulieren dies aber gleichzeitig oft selbst als aggressive Gegenattacke. Ablenkung und Irritation haben hier als Strategien einen besonders hohen Stellenwert, teilweise werden auch GIFs genutzt, die in der Regel der derb-humorvollen Grundstimmung angepasst werden (z.B. mit obszönen Gesten und lachenden Emojis) und insgesamt eine geringe Ernsthaftigkeit signalisieren. Werden gegenüber Täter*innen Interventionen gesetzt, fallen diese besonders von Burschen mit Migrationshintergrund überwiegend aggressiv, beschimpfend und konfrontativ aus. Männliche Teilnehmer ohne Migrationshintergrund wirken hingegen eher überfordert und posten entweder nichts oder scheinbar willkürlich Inhalte zur allgemeinen Erheiterung bzw. zur Ablenkung. Interventionen, die das Opfer unterstützen, finden sich nur vereinzelt. Weibliche Spielerinnen rufen ebenfalls zur Beruhigung und Beendigung der Konflikte auf, auch mit moralischen Argumenten, unterstreichen aber mit Hinweisen auf mögliche rechtliche Konsequenzen die Ernsthaftigkeit dieser Aufrufe. Da sich solche „gemäßigten“ Interventionen nicht durchsetzen können, versuchen insbesondere Mädchen mit Migrationshintergrund, mit aggressiven und beleidigenden Gegenattacken, die an die Ausdrucksweise der Übergriffe angepasst sind, Täter*innen die Stirn zu bieten. Ein solches Vorgehen gelingt aber nur wenigen Mädchen, viele andere sind überfordert und reagieren wenig oder kaum. Da die meisten damit beschäftigt sind, sich Gehör zu verschaffen, gehen auch Unterstützungsversuche für das Opfer unter.

In allen Gruppen sind immer wieder sowohl weibliche, als auch männliche Spielteilnehmer*innen zu finden, die scheinbar desinteressiert wirken, aber offenbar aufgrund der überwiegend ausfälligen Inszenierung überfordert sind, entsprechend zu intervenieren. Insbesondere bei Jugendlichen aus Schulen mit formell geringerer Bildung fällt auf, dass diese oft relativ lange inaktiv bleiben und insgesamt der Eindruck von Hilflosigkeit und Ohnmacht entsteht.

Resümierend lässt sich festhalten, dass Mädchen insgesamt ein vielseitigeres Repertoire an Counter Speech Strategien aufweisen als Burschen, die häufig, anstatt auf konstruktive Lösungsversuche zu setzen, den Weg der Ablenkung und Irritation beschreiten. Ein wesentliches Problem scheint darin zu liegen, dass sich jugendliche Bystander*innen erst einmal Gehör verschaffen müssen, bis konstruktive Reaktionen überhaupt möglich sind. Dies geschieht häufig durch ebenso laute, wenn nicht noch lautere, und ebenso aggressive Gegenattacken und Drohungen. Jugendliche Bystander*innen haben offenbar das Gefühl, nur dann gehört zu werden, wenn sie sich auch entsprechend stark präsentieren können. Dies wird durch den Fokus auf Täter*innen-orientierte Strategien unterstrichen, indem Jugendliche ihr Auftreten und ihre Selbstdarstellung als Counter Speaker*innen auch im Hinblick darauf gestalten, wie sie sich vor Täter*innen und vor dem Umfeld von mitlesenden Peers präsentieren möchten: Während etwa Burschen mit höherer Bildung ihre Stärke eher über ein überlegenes intellektuelles Niveau gegenüber Täter*innen darstellen („du bist unreif“), tun dies Burschen mit geringerem Bildungshintergrund eher über verbal/aggressive bzw. auch über körperliche Stärke (in einer möglichen Offline-Konfrontation). Mädchen mit höherem Bildungshintergrund betonen hingegen nicht unbedingt ihre persönliche Stärke, sondern verweisen eher auf das Einhalten normativer Regeln. Mädchen mit geringerem Bildungshintergrund zeigen sich wesentlich konfrontativer und versuchen, Täter*innen klar die Stirn zu bieten. Die Strategie, die eigene Stärke durch selbstbewusstes, gleichzeitig aber oft auch konfrontatives und aggressives Auftreten und expliziten Sprachgebrauch zu betonen, wird besonders von Jugendlichen mit Migrationshintergrund häufig verwendet.

Der starke Fokus auf Täter*innen-bezogene Strategien erzeugt bei den Rollenspiel-Teilnehmer*innen das Gefühl, mit den eigenen Interventionen letztlich erfolglos zu sein. Opfer-bezogene Strategien werden eher von Mädchen und von Jugendlichen mit Migrationshintergrund verwendet – also von jenen, die in der Regel in Online-Kontexten selbst die meisten Opfererfahrungen aufweisen. Bystander*innen-bezogene Strategien, um z.B. andere in ihrem zivilcouragierten Handeln zu unterstützen oder sich zusammenzuschließen, sind bei allen Gruppen nur selten zu finden. Dennoch lassen sich Hinweise finden, wie eine Mobilisierung anderer Bystander*innen zur Unterstützung gelingen könnte: Bystander*innen fühlen sich in ihrer Rolle in der Regel nicht besonders wohl: es ist unangenehm, unerwartet mit Konflikten konfrontiert zu sein, manche fühlen sich davon genervt oder haben keine Lust, durch diese Rolle in schlechte Laune und unangenehme Situationen zu geraten. Insofern werden jene Counter Speech Strategien eher unterstützt, die auch die Verbesserung der eigenen Befindlichkeit in Aussicht stellen (z.B. schnelle Lösungen zur Wiederherstellung einer angenehmen Gesprächsatmosphäre). Bystander*innen sind aber insbesondere bei eskalierenden Übergriffen auch schnell überfordert. Daher werden jene Counter Speech Strategien eher unterstützt, die eine Orientierung anbieten, wo ein Eingreifen tatsächlich angebracht ist und wo nicht, indem Übergriffe klar als Normverstöße oder als moralische Verwerflichkeit deklariert werden (z.B. „das ist strafbar“, „das ist unterste Schublade“). Darüber hinaus fällt es Bystander*innen leichter zu intervenieren, wenn sie dafür eine klare Anweisung erhalten (z.B. meldet alle) oder wenn sie vorbildhafte Interventionen nachahmen können (z.B. öffentliche Zeichen der Solidarität:

alle schreiben Nachrichten mit Namen des Opfers und positiven Emojis) und wenn damit gleichzeitig eine Solidarisierung mit anderen Bystander*innen entsteht.

In Reflexionsrunden und Diskussionen mit den teilnehmenden Jugendlichen wurden die unterschiedlichen Rollen und Strategien reflektiert. Die Artikulation des eigenen Befindens und die Einschätzung der verwendeten Strategien durch die Teilnehmer*innen unterstreicht die genannten Ergebnisse: Die Jugendlichen fühlten sich in ihrer Bystander*innenrolle in der Regel wirkungslos und ignoriert, das Auftreten als Bystander*innen wurde als anstrengend empfunden. Erfolge konnten aus ihrer Sicht nur punktuell und kurzfristig erzielt werden, etwa durch Ablenkung mit lustigen GIFs. Wesentliche Gründe für ihr Scheitern sahen die Beteiligten darin, dass man im Internet als Bystander*innen nicht ernstgenommen wird und generell eine Einmischung von außen unangebracht erscheint. Zudem seien Kommentare nur begrenzt einsetzbar und könnten direkte Face-to-Face Kontakte nicht ersetzen. Online Konflikte seien demnach nur offline zu lösen. Weiters würden viele Postings auf einfach auch zusätzliches Chaos produzieren.

Zumindest kurzfristige Erfahrungen von Erfolg wurden in Zusammenhang mit der aktiven Unterstützung des Opfers gemacht, indem diesem Anteilnahme und Trost gespendet wurde. Erfolgserfahrungen im Umgang mit Täter*innen wurden am ehesten in der Kontaktaufnahme über die Privat-Chat -Funktion gemacht, nicht aber im öffentlichen Chat. Hier wurden eher gezielte Ablenkungsstrategien als erfolgreich bewertet, wenn dadurch die Kommunikation zumindest kurzfristig umgelenkt oder abgebrochen werden konnte. Als Erfolg wurden auch Absprachen mit anderen Bystandern*innen im Privat-Chat genannt, wenn es dadurch gelang, ein gutes Zusammenspiel zu ermöglichen.

Insgesamt machten die Online Rollenspiel-Teilnehmer*innen in ihrer Rolle als Bystander*innen aber kaum positive Wirkungserfahrungen im Umgang mit Konflikten und betrachteten ihren Handlungsspielraum in Sozialen Medien daher als sehr begrenzt. Aus Sicht der beteiligten Jugendlichen gelingt der Umgang mit Konflikten in Online Umgebungen daher am ehesten durch persönliche Kontaktaufnahme und Mobilisierung von Freund*innen. Als Internet-basierte Strategien bleiben lediglich das Blockieren, Melden, der Ausschluss aus der Gruppe und schließlich das weit verbreitete Ignorieren der Konflikte.

Um Jugendliche für Counter Speech zu motivieren, muss also zunächst geklärt und Jugendlichen vermittelt werden, was mit Counter Speech bewirkt und wie diese erfolgreich gestaltet werden kann. Dabei wäre es auch wichtig, den starken Fokus auf Täter*innen-orientierte Counter Speech Strategien zu verlagern, z.B. indem stärker versucht werden könnte, durch Counter Speech andere jugendlicher Bystander*innen als Unterstützung zu mobilisieren.

2. Inhaltsanalyse von realen User*innen-Reaktionen auf YouTube

Als zweite empirische Erhebungsphase wurde eine Inhaltsanalyse von User*innen-Kommentaren auf YouTube umgesetzt. Ziel dieses methodischen Zugangs war es, öffentlich sichtbare, spontane, „alltägliche“ Reaktionen Jugendlicher im Umgang mit negativen Inhalten auf einer breiteren Ebene zu untersuchen.

2.1 Methodisches Design

2.1.1 Datensammlung und Stichprobendesign

Die erste Herausforderung bei der Auswahl der Daten war es, einen Kanal zu bestimmen, der mit hoher Wahrscheinlichkeit Hate und Counter Speech Interaktionen von Jugendlichen zwischen 14 und 18 Jahren beinhaltet. Hate Speech wurde hier als negative Annotation gegenüber einer Person oder einer Gruppe definiert, Counter Speech als darauf bezogene direkte oder indirekte Gegenrede. Verschiedene Social-Media-Kanäle, die bei Jugendlichen beliebt sind, haben unterschiedliche Vor- und Nachteile: So wurde etwa Twitch, eine Live-Streaming-Plattform, die hauptsächlich zum Streamen und Ansehen von Lets-Play-Videos verwendet wird und von jungen Leuten intensiv verwendet wird, in Erwägung gezogen. Der für Twitch-Nutzer*innen verfügbare Chat ist jedoch durch sehr schnelle Abläufe charakterisiert, mit mehreren Nachrichten pro Sekunde in stark frequentierten Kanälen. Die Kommunikation richtet sich zudem hauptsächlich an die Streamer*innen, die darauf reagieren sollen. Obwohl Twitch, wie die Gaming-Szene im Allgemeinen, für ihre toxischen, oft frauenfeindlichen Kommunikationsinhalte bekannt ist, schien es unwahrscheinlich, dass in einer solchen Umgebung viel Counter Speech aufkommen würde.

Aus mehreren Gründen entschieden wir uns schließlich, uns auf **Daten aus YouTube** zu konzentrieren. YouTube ist bei Jugendlichen äußerst beliebt, und praktisch jede*r Teenager*in konsumiert regelmäßig Videos auf der kostenlosen Online-Videoplattform (siehe dazu Jugend-Internet-Monitor 2021).

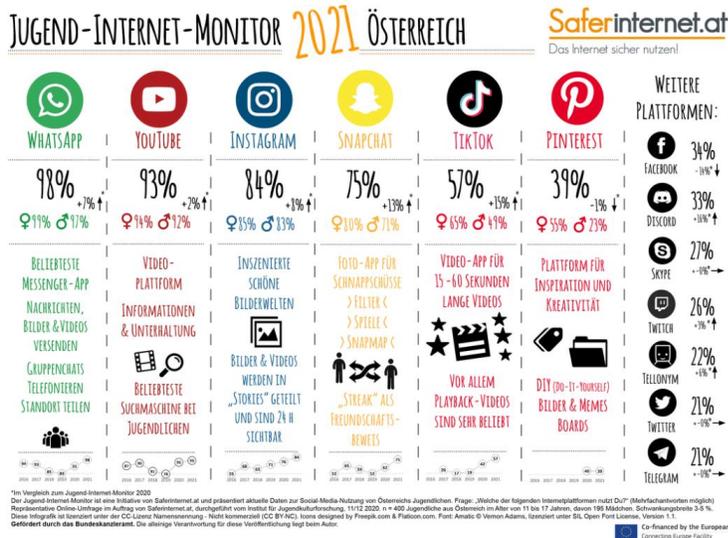


Abbildung 1: Jugend Internet Monitor - saferinternet.at (Letztzugriff 02.02.2022)

Ein weiterer Vorteil von YouTube ist, dass Kanäle und damit Inhalte so kategorisiert werden können, dass Rückschlüsse auf die Zuschauer*innen gezogen werden können. Beispielsweise werden „Let's Play“-Videokanäle viel eher von männlichen Jugendlichen angesehen, während Mode- und Make-up-Kanäle deutlich stärker von weiblichen Nutzerinnen konsumiert werden. Darüber hinaus ermöglicht der Kanal und mehr noch der Titel eines Videos, das Gesamtthema einzugrenzen, auf das sich die Kommentare beziehen, was hilfreich ist, um Kommentarinhalte in einen Kontext zu setzen.

Bei der Konzeption und Umsetzung der zu analysierenden Stichprobe gab es für das Forschungsteam einige zu bewältigende **Herausforderungen**. In einer ersten ursprünglichen Idee wurden in Zusammenarbeit mit Saferinternet.at YouTube Kanäle festgemacht, die in der Untersuchungszielgruppe (14-18 Jahre) häufig genutzt werden. Ziel war es, nach Interaktionsfolgen zu suchen, die sowohl Online-Übergriffe als auch entsprechende Reaktionen beinhalten. Für den Datendownload wurde ein eigener Scraper programmiert. Trotz dieser technischen und methodischen Vorbereitungen erwiesen sich allerdings für die Datenanalyse die zunehmend stärker praktizierten Richtlinien von YouTube gegen Hate Speech und Cybermobbing als hinderlich für das Forschungsziel. Auf YouTube haben Nutzer*innen die Möglichkeit, Videos hochzuladen. Diese Videos können auch von anderen Nutzer*innen kommentiert werden, diese Kommentare können wie die Videos selbst auch kommentiert und geliked sowie disliked werden. Um Hass im Netz entgegenzuwirken, bietet YouTube einerseits einen automatischen Filter für potenziell „unpassende“ Kommentare an, die dann von vornherein nicht sichtbar sind. Kommentare können außerdem von dem*der Kanalbetreiber*in gelöscht werden. Weiters gibt es auf YouTube die Möglichkeit, Moderator*innen hinzuzufügen, die ebenso Kommentare löschen und Nutzer*innen blockieren können, was vor allem bei YouTuber*innen mit hoher Reichweite häufig der Fall ist. Zusätzlich ist es möglich, bestimmte Wörter zu blockieren, sodass keine Kommentare, die diese geblockten Wörter

beinhalten, gepostet werden können. Die Kommentarfunktion kann auch ganz ausgeschaltet werden. Für die geplante Datenanalyse bedeutete dies, dass im Datenmaterial Ursprungskommentare häufig nicht mehr vorhanden und Interaktionsfolgen oft lückenhaft waren. Hinzu kommt, dass User*innen Kommentare abhängig vom Kanal in unterschiedlicher chronologischer Reihenfolge erscheinen.

Diese Herausforderungen erforderten ein **Umdenken in der Stichproben-Strategie**. Folgende strategische Überlegungen wurden unter Einbezug von externen Expert*innen (u.a. auch YouTuber*innen) diskutiert und erprobt: Trotz der lückenhaften Datenlage ist bekannt, dass in YouTube Kommentaren tendenziell viel diskutiert wird, weshalb es hier auch zu vielen Hasskommentaren und auch zu Formen von Online-Zivilcourage kommen kann. Wie viele negative Inhalte in den Kommentaren vorherrschen, hängt aber direkt mit dem Engagement der Kreator*innen selbst zusammen, dh. wie aktiv diese Hassinhalte löschen, um dagegen vorzugehen. Bei größeren Kontroversen, Dramen oder Skandalen (z.B. Outing-Vorwürfe von Miguel Pablo, Video zum Song *wap bap* mit besonders hohen Dislikes) übersteigt das Ausmaß an Hasskommentaren die Kapazitäten stark. **Die Sampling-Strategie wurde daher einerseits stärker an diesen unter Jugendlichen diskutieren aktuellen Kontroversen** ausgerichtet. Andererseits wurden aber auch **Themenkomplexe verfolgt, die in der Regel häufig heftige Diskussionen unter Jugendlichen auslösen**, wie z.B. LGBTQ+ Themen.

Um nun konkrete Kanäle auszuwählen, wurde zunächst eine Liste erstellt, mit dem Ziel, sowohl unterschiedliche inhaltliche Ausrichtungen, als auch ein breites Spektrum an jugendlichen User*innen abzubilden. Ein weiteres Kriterium war, dass der Kanal eine hohe Anzahl an Abonnent*innen und damit genügend Nutzer*innen hat, um auch eine ausreichend hohe Anzahl an Hate und Counter Speech Interaktionen zu finden.

In einem weiteren Schritt wurde bei der Auswahl der Kanäle, die besonders bei Jugendlichen in Österreich und Deutschland beliebt sind, zusätzlich auch eine Liste mit Suchbegriffen erstellt, die auf aktuelle Skandale in den von Jugendlichen frequentierten YouTube-Bereichen verwiesen. Mit Hilfe einer jungen YouTube-Expertin wurden einige Ereignisse identifiziert, die zu heftigen Kontroversen und hoher Sichtbarkeit auf YouTube geführt haben: Beispielsweise ein Skandal um einen mittelmäßig populären deutschen YouTuber, der zunächst sein geplantes Coming-Out in einem Video ankündigte, um eine Woche später zuzugeben, dass alles ein Scherz war, mit dem Ziel, dadurch mehr Follower und Views zu generieren. Die YouTube-Community nahm diesen YouTuber und seine Entschuldigungsversuche nicht allzu gut auf, sodass heftige Debatten und Auseinandersetzungen die Folge waren. Mehrere andere Skandale wie dieser wurden von Suchbegriffen abgedeckt und die Top-Videos, die bei der Eingabe dieser angezeigt wurden, wurden ebenfalls gescraped. Darüber hinaus wurden die Namen mehrerer umstrittener YouTube-Prominenter, die selbst keine beliebten Kanäle hatten, ebenfalls als Suchbegriffe eingesetzt

Die endgültige Auswahl umfasste **Kanäle aus den folgenden Kategorien:**

Musik

Mert - Deutsch Rapper, Mert ist ein YouTuber und Rapper, der vor allem im Frühling 2017 im Gespräch für homophobe Aussagen war, als Reaktion darauf veröffentlichte er ein Statement Video, in dem er angab, homosexuelle Menschen einfach nicht zu akzeptieren. Das trat eine große Kontroverse, sowie eine Welle an Solidarität für LGBTQIA+ Personen los – Shirin David wurde in diesem Kontext ebenso stark kritisiert, da sie mit ihm zusammengearbeitet hatte.

Shirin David - Lifestyle & Beauty Youtuberin, die auch eine Musikkarriere gestartet hat. Durch ihr Äußeres (meistens stark geschminkt etc.) zieht sie sehr viel sexistischen Hass auf sie, weiters stand sie immer wieder in Kritik, mit kontroversen Personen sowohl auf YouTube als auch mit ihrer Musik zusammengearbeitet zu haben (siehe Mert).

Katja Krasavice - Die Sängerin hat ihre Karriere als YouTuberin mit Content voller sexueller Anspielungen begonnen. Dafür wurde sie oft stark kritisiert, da Content wie ihrer eigentlich laut den Guidelines nicht auf YouTube erlaubt ist. Mittlerweile ist sie seit einigen Jahren erfolgreiche Musikerin, auch in ihrer Musik spielt sie oft auf sexuelle Themen an.

Information

LeFloid - Er ist seit über zehn Jahren YouTuber und macht fast genauso lange Videos in seinem Format „LeNews“, in denen er skurrile, absurde und kontroverse Geschehnisse regelmäßig vorstellt. Seine Videos sind durchwegs polemisch, werden von seiner Community aber eher als journalistisch und objektiv aufgefasst. Durch seine Themenwahl kommt es in seinen Kommentaren sehr oft zu Diskussionen zwischen anderen User*innen. Die inhaltliche Ausrichtung seiner Themen kann politisch eher als links eingeordnet werden.

Franziska Schreiber - Sie ist eine Youtuberin, die Videos über Politik macht und aufgrund ihrer Vergangenheit bei der AFD, ihrer Meinung zu „kontroversen Themen“ und der Tatsache, dass sie von funk (öffentlich-rechtliches Programm für Online-Content) unterstützt/finanziert wird kritisiert wird. Die inhaltliche Ausrichtung ihrer Themen kann politisch als eher rechts eingeordnet werden.

Lifestyle & Beauty

BibiBeautyPalace - Sie ist eine der größten Influencer*innen im deutschen Sprachraum und seit Jahren immer wieder im Gespräch, da sie in erster Linie dafür kritisiert wird „Kindern das Geld aus der Tasche zu ziehen“. Die größte Kontroverse rund um sie ist ihr 2017 veröffentlichtes Musikvideo, das mit mittlerweile über 60 Millionen Views sehr lange Zeit das international am meisten gediskutierte Video war. Auch rund um ihren Sohn mit Influencer Julienco und ihre zweite Schwangerschaft gibt es viel Diskurs.

PersiaX - Sie ist selbst ein Mitglied der LGBTQIA+ Community und fällt immer wieder mit ihren Meinungen diesbezüglich und auch zu anderen kontroversen Themen wie beispielsweise dem Islam auf. Durch ihr offenes Leben als transsexuelle Person wird in den Kommentaren viel diskutiert und sie ist häufig mit transphoben Kommentaren konfrontiert.

Miguel Pablo - Der Lifestyle/Prank YouTuber ist immer durch Kontroversen aufgefallen, meistens in Verbindung mit seinen psychischen Problemen, Drogenkonsum oder einem Psychriaufenthalt. Anfang 2020 postete er mehrere Videos über sein Outing als homosexuell, verschiedene Aktivitäten mit einem Partner, nur um dann zu erklären, dass all dies nur gestellt war.

Gaming

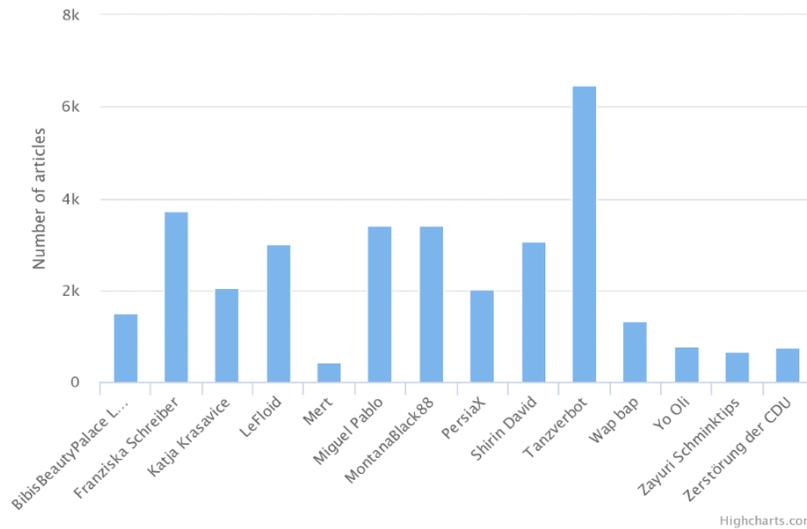
MontanaBlack88 - Monte (Spitzname der Community) ist der größte deutschsprachige Streamer, der vor allem Spiele wie Call of Duty oder FIFA spielt. Zwar macht er mittlerweile den meisten Content auf der Streaming Plattform Twitch, Zusammenschnitte von seinen Gaming Livestreams werden aber nach wie vor auf YouTube hochgeladen. Er spricht in YouTube Videos auch öfter über seine Autos, Glücksspiel oder Frauen – letzteres oft auf sexistische und misogyne Arten und Weise. Er wurde auch schon öfter für gewisse Zeit von der Streamingplattform Twitch wegen unter anderem rassistischen Aussagen gesperrt.

Reaktionsvideos

Tanzverbot - Seine Videos sind großteils Vlogs (Videoblogs) oder Videos, in denen er seine Meinung über gewisse Themen oder auch andere YouTuber*innen kundtut. Tanzverbot nimmt sich hierbei kein Blatt vor den Mund und hat regelmäßig Streit mit anderen Content-Creator*innen. Ein großes Thema ist auch sein Gewicht und Essen.

Die folgende Tabelle zeigt die Verteilung der Kommentar-Blöcke pro YouTube-Kanal/Suchbegriff im gesammelten Datenmaterial.

Tabelle 2: Verteilung der Kommentarblöcke pro YouTube-Kanal/Suchbegriff



Der **Datendownload** erfolgte in zwei Phasen: 12. bis 22. Januar 2020 sowie 26. Februar bis 25. März 2020. Insgesamt wurden 32.852 Kommentarblöcke von 517 Videos heruntergeladen. In einem Bereinigungsschritt wurden 274 Duplikate gelöscht, was eine Gesamtzahl n=32.578 ergibt.

Um dieser enormen Menge an Datenmaterial relevante Passagen mit Hate und Counter Speech Interaktionen für die Inhaltsanalyse auszuwählen, wurden Kommentarblöcke anhand eines **Wörterbuchs (=Sammlung an Suchbegriffen)** vorselektiert, das hauptsächlich jugendtypische und im Netzjargon gebräuchliche Schimpfwörter und gewalthaltige Begriffe enthielt. Dieses Wörterbuch wurde vom Forschungsteam vorrangig auf Basis der Daten aus den Online Rollenspielen erstellt. Um das Wörterbuch zu optimieren, wurde dieses mit kleinen Zufallsstichproben aus dem Datenmaterial solange getestet, bis kein weiteres Verbesserungspotenzial mehr erkennbar war. Die Grammatik und Rechtschreibung von Jugendlichen folgt auf YouTube bzw. generell in digitalen Medien nicht den herkömmlichen Regeln und oft werden Schimpfwörter absichtlich falsch geschrieben, um eine Erkennung und Löschung durch die YouTube-Algorithmen zu vermeiden. Unser Wörterbuchansatz lieferte jedoch überraschend solide Ergebnisse. Auf Grundlage dieses Wörterbuchs wurde ein Searchstring erstellt, welcher die folgenden Wörter enthält:

kahba OR kelek OR hure OR nutte* OR schlampe OR wixxen OR wixx* OR fotze* OR hdf OR fresse OR "halt* maul"~5 OR verhur* OR snowflake OR "nix du"~3 OR "nix da"~3 OR prostituier* OR puff OR schieße OR scheiss* OR verarsch* OR mist OR gschiss* OR dreck* OR stück OR gusch OR goschn OR bitch OR behindert* OR propaganda* OR geschiss* OR nazi* OR rassist* OR nzi OR bullshit OR unfähig* OR anscheiß* OR schiss* OR anscheiss* OR depp* OR pisse* OR arsch* OR brauner OR rassist OR spinn* OR schieß* OR homo* OR "sei einfach"~2 OR Leleck* OR esel* OR schnauze OR likegeil OR "like geil"~2 OR dummkopf OR Idiot OR dulli* OR "halt* maul"~5 OR "Halt Schnauz*"~5 OR bescheuert* OR fresse* OR*

opfer OR "heul doch"~5 OR läster OR hdf OR Kahba OR aq OR amk OR blome OR hajde OR blyat OR hua OR huan OR kek OR hurensohn OR wichser OR hund OR schlampe OR fotze OR fut OR hure OR votze OR "ehre genommen"~6 OR cüsh OR rotz OR ehrenlos* OR "IQ xy"~6 OR "go off"~5 OR "wig snatch*"~4 OR clown OR "white knight"~3 OR simp OR hu OR behindert OR siff OR "male tears" OR "heul doch" OR "erstick" OR "halbe* buch" OR gef OR abartig* OR unnatürlich OR haram OR widerlich OR ekel* OR schwuchtel OR schwanz OR schwabo OR uhrensohn OR nazy OR cring* OR af OR smh OR feminazi OR tunte OR transe OR trap OR homo OR bruh OR penner OR hurentochter OR amkarmy OR "amk army" OR stan OR "hol mal luft" OR "beruhig dich" OR fy OR opfer OR hund OR bitch OR kind OR spiegel OR wack OR whack OR juckt* OR that OR kek OR suck* OR "heul leise"~2 OR pimmel* OR weeb OR ungeziefer OR parasit* OR "ok boomer"~3 OR "mach fertig"~4 OR "bring um"~4*

Das Wörterbuch wurde von zwei Kodierer*innen basierend auf einer Zufallsstichprobe von 150 Kommentarblöcken validiert und die **Intercoder-Reliabilität** berechnet. Unter Verwendung eines **semantischen Triplett-Ansatzes** (Franzosi, 2004 – siehe dazu auch weiter unten) wurde Hate Speech als negative Annotation gegenüber einer Person oder einer Gruppe definiert. Bei der Frage, ob ein Kommentarblock Hassreden enthielt, erreichten die beiden Kodierer*innen eine ausreichende Übereinstimmung (Krippendorffs Alpha: 0,70, 2 Beobachter*innen, 150 Kommentarblöcke). Fälle, in denen sich die Kodierer*innen nicht einig waren, wurden diskutiert und eine Konsensentscheidung getroffen. Damit konnte die Validität des Wörterbuchs, das zum Filtern der Gesamtdaten eingesetzt wurde, bestätigt werden. Das Wörterbuch hatte einen Recall von 1, was bedeutet, dass alle Kommentarblöcke in der Stichprobe, die Hassreden enthielten, als solche markiert wurden, und eine Genauigkeit von 0,48, was bedeutet, dass von den als positiv markierten Dokumenten etwa die Hälfte tatsächlich Hassreden gemäß unserer Definition (semantischer Triplett-Ansatz) enthielt (True Positives: 27, False Positives: 29, True Negatives: 105, False Negatives: 0, Accuracy: .82).

2.1.2 Ausarbeitung des Codebooks und Datenanalyse

Ziel einer quantitativen Inhaltsanalyse ist es, Texte möglichst objektiv (=intersubjektiv nachvollziehbar) verschiedenen Kategorien bzw. Variablen zuzuordnen. In die Kodierung sollte möglichst wenig persönliche Interpretation einfließen, weshalb Entscheidungen immer an einer bestimmten Textstelle und nicht an persönlichen Einstellungen oder Werthaltungen festgemacht werden sollten.

Eine Herausforderung für die vorliegende Studie war es, eine objektive Entscheidungsgrundlage festzulegen, wann Textteile als Hate Speech zu kodieren sind und wann nicht. Das Vorliegen von Hate Speech war Grundvoraussetzung, um in weiterer Folge Counter Speech (= Gegenreaktion in Bezug auf Hate Speech) zu kodieren. Um eine objektive Entscheidungsgrundlage zu gewährleisten, wurden die zu analysierenden YouTube Kommentarblöcke in einem ersten Schritt durch das Anwenden von klaren

grammatischen Regeln in sogenannte **“Semantic Triplets”** (Franzosi 2004) umgewandelt. Diese bestehen aus einem Subjekt, Verb und einem Objekt. Der Wert dieser ursprünglich aus der Linguistik stammende Methode wurde wiederholt für die Sozialwissenschaft aufgezeigt (z.B. Roberts 1997). Der große Vorteil von Semantic Triplets als Kodiereinheit besteht in ihrer inhärenten Reliabilität. Ihre Erstellung folgt objektiven grammatischen Regeln, was wiederum zu einer systematischen Handhabung von Texten führt (Aslanidis 2017).

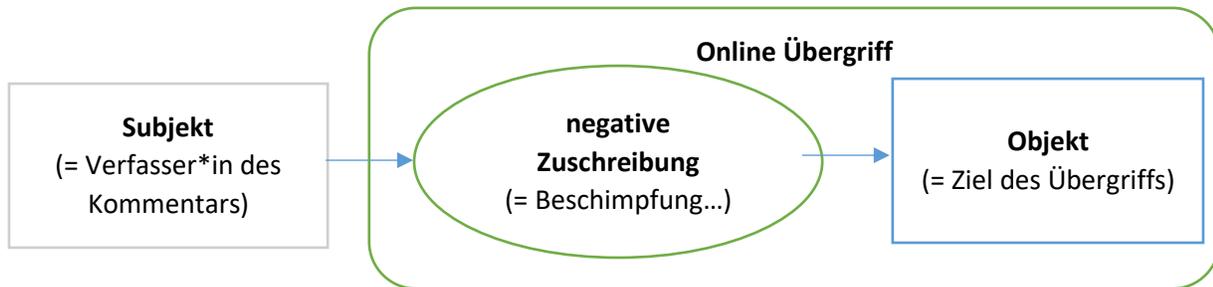


Abbildung 2: Semantic Triplet zur Festlegung der Kodiereinheit

Für die vorliegende Studie waren in erster Linie der Online-Übergriff sowie das Ziel Übergriffs relevant, das Subjekt entspricht der*dem jeweiligen Verfasser*in des Kommentars. Die Verwendung von Semantic Triplets stellt eine große Erleichterung für eine möglichst objektive Kodierung dar: So wurde das **Vorliegen eines Online-Übergriffs folgendermaßen definiert**: Ein Online-Übergriff liegt vor, wenn eine oder mehrere Personen (Subjekt, Hater*innen) eine andere Person oder eine Gruppe (Objekt) beleidigen, angreifen, beschimpfen oder herabwürdigen (negative Attribution, auch Verwendung von negativen Stereotypen und Vorurteilen: z.B. Muslime sind kriminell). Wichtig ist, dass das Ziel der negativen Attribution entweder eine Person (z.B.: der*die YouTuber*in, oder ein*e Kommentar-Verfasser*in) oder eine Gruppe (z.B.: Frauen, Homosexuelle, Ausländer) *sein muss*, damit ein Online-Übergriff vorliegt. Wird negatives/problematisches Gedankengut (z.B. Rassismus, Homophobie, Extremismus, Sexismus) in einem Kommentar reproduziert, dann wird dies als Angriff auf die Gruppe gewertet, auch wenn diese nicht explizit genannt wird.

Beispiel: Zerlegt man die Aussage „Ich finde dich scheisse“ in ihre grammatikalischen Bausteine (Attribution = „Scheisse“; Objekt = „du“), ist sie demnach als Online-Übergriff zu kodieren, weil das Objekt bzw. Ziel der negativen Zuschreibung eine Person ist. Die Aussage „Ich finde diese Aussage scheisse“ jedoch nicht, da das Ziel der negativen Zuschreibung keine Person oder Personengruppe ist, sondern sich die Kritik lediglich auf den Kommentar bezieht.

Diesem Gedanken folgend ist Kritik an sich noch kein Online-Übergriff, auch wenn Schimpfwörter enthalten sind, solange das Ziel nicht eine Person oder Gruppe ist. Ebenso wenig handelt es sich um einen Online-Übergriff, wenn zwar eine Person kritisiert wird, dies jedoch sachlich und ohne persönliche

Beleidigungen geschieht. Beide Elemente, sowohl die Beleidigung als auch das Ziel Person/Gruppe müssen vorhanden sein.

Weiters wurde auch auf die **Relevanz bzw. Kodierbarkeit** geachtet: Manche Kommentarblöcke waren durch beispielsweise ein gelöschter Ursprungskommentar oder mehrere gelöschte Kommentare nicht mehr nachvollziehbar und konnten daher nicht sinnvoll kodiert werden (teilweise war es auch nicht mehr möglich, eindeutig zwischen Übergriff und Gegenreaktion zu unterscheiden). Solche Kommentarblöcke wurden im System als „irrelevant“ markiert und nicht kodiert. Weiters wurden auch Kommentarblock ausgeschlossen, die ausschließlich in einer anderen Sprache (z.B. Türkisch, Englisch) gehalten wurden und keine relevanten deutschsprachigen Kommentare enthielten.

Für die Durchführung der Datenanalyse wurde aufbauend auf diesen Vorüberlegungen zunächst ein **umfassendes Codebook ausgearbeitet**, um einerseits Hate und Counter Speech Elemente in der Interaktion zu identifizieren und Interaktionsverläufe abzubilden sowie andererseits das Set an jugendtypischen Counter Speech Strategien und deren Konsequenzen zu dokumentieren. Dabei wurden folgende Inhalte erfasst:

- Identifikation von Hate-Speech Elementen: liegt HS vor? Anzahl an Hater*innen, Ziel des Übergriffs (z.B. User*in, YouTuber*in, andere Personen[gruppe]), Schweregrad des Übergriffs, inhaltliche Ausrichtung (z.B. Kraftausdrücke, misogynie/ queer-feindliche/ rassistische Aussagen, abwertende Stereotypisierungen, Gewaltandrohung, Verhetzung, Unterstellungen/ Unwahrheiten/ üble Nachrede, persönlichen Angriff gegen Äußerlichkeiten/ Fähigkeiten)
- Identifikation von Counter-Speech Elementen: liegt CS vor? Anzahl an Counter Speaker*innen, wer regiert (Opfer bzw. Opfergruppe zugehörig, Dritte, YouTuber*in), Counter Speech Strategien¹² (Warnung vor Konsequenzen, Shaming & Labeling, Zugehörigkeit & Mitgefühl, Humor, feindlich/ aggressiv/ beschimpfend, Faktencheck, Gewaltandrohung, Ablenkung/ Unterbrechen der Kommunikation, Einfache Aufforderung aufzuhören, Deeskalation, ins Gewissen reden/ mit Fragen zur Rede stellen/ sachlich zurechtweisen)
- Dokumentation der Konsequenzen / des Erfolgs im Interaktionsverlauf: Verhaltensänderungen von Hater*innen (z.B. keine Reaktion, negative/ positive/ gleichbleibende Entwicklung der Tonalität), Verhaltensänderungen der Counter Speaker*innen (z.B. keine Folgeaktion, Standhaftigkeit, negative/ positive/ gleichbleibende Entwicklung der Tonalität), Veränderung des Machtverhältnisses zwischen Hater*innen und Counter Speaker*innen, Bewertung der Komplexität des Interaktionsverlaufs

Das Codebook wurde mit **ausführlichen Dokumentationen inklusive Fallbeispielen versehen**. Die insgesamt vier Kodierer*innen erhielten eine ausführliche gemeinsame **Schulung**, in deren Rahmen

¹² Ausgearbeitet in Anlehnung an Benesch et al. (2016a) und Ergebnisse der Online Rollenspiele.

Kodierungen beispielhaft gemeinsam durchgeführt und diskutiert wurden. **Für die Durchführung der Kodierung** wurde **AmCAT** (Amsterdam Content Analysis Toolkit: <http://autnes.amcat.nl/accounts/register/amCAT>) ausgewählt, eine Software, die im akademischen Kontext häufig für Textanalysen eingesetzt wird.

Um schließlich das **Codebook zu validieren**, wurden jeder*m Kodierer*in jeweils die gleichen 150 Kommentarblöcke (Zufallsauswahl) zugewiesen und anschließend die Intercoder Reliabilität berechnet. Nach erfolgreicher Validierung sollten dann pro Kodierer*in jeweils 500 unterschiedliche Kommentarblöcke zugewiesen werden, um schließlich einen Datensatz von insgesamt 2.000 kodierten Kommentarblöcken zu erhalten. Allerdings zeigten sich für die **Intercoder Reliabilität** sehr schlechte Werte: Pro Variable lagen die Krippendorff's Alpha Werte zwischen 0.57 und 0.0005, d.h. alle Variablen lagen unter dem in der Literatur angegebenen Grenzwert von 0.68.

Folgende **Maßnahmen** wurden daher ergriffen:

- 1) Prüfung, ob einzelne Kodierer*innen den Wert drücken (war nicht der Fall)
- 2) Entfernung bzw. Reformulierung von Variablen mit besonders schlechten Werten, da diese offenbar nicht korrekt kodiert werden konnten
- 3) Adaptierungen und Verbesserungen der Definitionen im Codebook
- 4) Wiederholte Schulung
- 5) Nochmaliges Testen

Dieses Prozedere wurde mehrmals wiederholt. Nach jeder Testung war aber die Intercoder Reliabilität erneut zu niedrig. Die vermuteten Gründe dafür lassen sich wie folgt zusammenfassen:

- große subjektive Bewertungsunterschiede, ab wann z.B. negative Zuschreibungen als Übergriff bewertet werden bzw. wo Hate Speech beginnt (z.B. Einbettung von Kraftausdrücken in humorvolle Kontexte) oder was als Reproduktion von negativem Gedankengut gilt.
- Komplexe und wirre Verläufe in den Interaktionsfolgen, sodass es schwierig bzw. mitunter fast unmöglich ist, den Überblick zu bewahren (insbesondere, welche Personen als Hater*innen, welche als Counter Speaker*innen zu bezeichnen sind; besonders auch bei wechselnden Benutzer*innennamen problematisch)
- Schwierige Zuordnung aufgrund oft mehrerer Übergriffe und verschiedener Arten von Übergriffen
- Fehlendes Kontextwissen, um Kommentare als möglichen Übergriff zu interpretieren
- Unklarheit, worauf sich Kommentare beziehen (z.B. auch aufgrund gelöschter Inhalte)

Solche Probleme sind prinzipiell nicht neu und stellen insbesondere bei automatisierten Hate Speech Detection Tools eine grundlegende Schwierigkeit dar. Überraschend war allerdings das Ausmaß dieser Abweichungen trotz detaillierter Codebook-Definitionen mit Fallbeispielen und Beschreibungen sowie mehrfachen Schulungen der Kodierer*innen. Diese Feststellung veranlasste das Forschungsteam daher,

eben diese **Bewertungsunterschiede in den Analysefokus** zu rücken: Pro Kodierer*in wurden daher nicht 500 unterschiedliche Kommentarblöcke zur Kodierung ausgegeben (was auch aufgrund der zu geringen Intercoder Reliabilität nicht vertretbar gewesen wäre), sondern jede*r Kodierer*in erhielt die 500 gleichen Kommentarblöcke. Ziel war es nun, Faktoren zu analysieren, an denen die Zuordnung festgemacht wird. Die Ergebnisse sind nicht nur für automatisierte Hate Speech Analysen von Relevanz, sondern auch für jugendliche Counter Speaker*innen, die (ohne Schulung) in der Regel vor der gleichen Problematik stehen und nicht einschätzen können, ob es sich dabei um Hate Speech handelt oder nicht. Weiteres Ziel war es außerdem, den Datensatz zu filtern, und zwar sowohl jene Fälle, in denen weitgehend Einigkeit herrscht, als auch jene mit großer Uneinigkeit zu eruieren. Hier sollte dann analysiert werden, warum diese Differenzen bestehen.

Das **finale Codebook** wurde daher nochmals angepasst, verkürzt und vereinfacht und enthielt folgende Inhalte:

- Identifikation von Hate-Speech Elementen: liegt HS vor? inkl. konkrete Markierung des Kommentars, Anzahl an Hater*innen/Hate-Unterstützer*innen, Ziel des Übergriffs (z.B. User*in, andere Personen), Schweregrad des Übergriffs in Abhängigkeit der Stellung in der Interaktionskette – siehe dazu dritter Aufzählungspunkt: z.B. Verstärkung Übergriff, gleichbleibend oder Zurücknahme
- Identifikation von Counter-Speech Elementen: liegt CS vor? inkl. konkrete Markierung des Kommentars, Anzahl an Counter Speaker*innen, wer reagiert (Opfer bzw. Opfergruppe zugehörig, andere Personen), Counter Speech Strategien in Abhängigkeit der Stellung in der Interaktionskette – siehe dazu dritter Aufzählungspunkt: Humor/ Sarkasmus/ Ironie, schlechtes Gewissen/ Moral, Verständnis/ Mitgefühl, aggressive/ feindliche Gegenattacke, sachlich/ faktenbasierte Gegenargumentation/ neutrales Nachfragen, einfache Aufforderung aufzuhören, Deeskalation der Gesamtsituation bzw. keine weitere Folgeaktion, wenn nach erster CS weitere HS folgt
- Dokumentation der Konsequenzen / des Erfolgs im Interaktionsverlauf: Abbildung der Interaktionsketten zwischen Hate und Counter Speaker*innen mit Bewertung der Schwere von HS; CS Strategien; freies Textfeld zu Hinweisen auf Erfolg von Counter Speech

Um diese finalen Anpassungen gut in die Analysestrategien einbringen zu können, erhielten alle Kodierer*innen eine letzte abschließenden Schulung.

2.2 Ergebnisse

2.2.1 Identifikation von Hate Speech

Tabelle 2

Gibt es User*innen, die Hate Speech betreiben oder unterstützen? (n=500)					
	Kodierer*in 1	Kodierer*in 2	Kodierer*in 3	Kodierer*in 4	Mittelwert
Nein bzw. unklar	30,8%	55,8%	32,4%	54,4%	43,4%
Ja, eine Person	34,0%	24,4%	36,8%	33,2%	32,1%
Ja, mehrere Personen	35,2%	19,8%	30,8%	12,4%	24,6%

Wie Tabelle 2 zeigt, gibt es zwischen den Kodierer*innen trotz des ausführlichen Codebooks und mehrmaliger Schulung deutliche Differenzen in den Bewertungen, ob Hate Speech vorliegt oder nicht. Im Durchschnitt wird in etwa 57% der Fälle Hate Speech durch eine*n oder mehrere User*innen wahrgenommen, in etwa 43% der Fälle aber nicht.

Um zu prüfen, in welchen Fällen Übereinstimmung herrscht, wurde in erstem Schritt die Variable nach dem Vorkommen von Hate Speech pro Kodierer*in dichotomisiert in (1 = ja, eine oder mehrere Personen üben Hate Speech aus; 0= nein, es kommt kein Hate Speech vor) und anschließend ein Summenindex gebildet (höchste Übereinstimmung bei Summe 4 bzw. 0). Dabei zeigte sich folgendes Ergebnis:

Tabelle 3

Übereinstimmung in der HS-Kodierung (n=500)	
Übereinstimmung: NIEMAND kodiert HS	21,0%
eine Kodierer*in kodiert HS, andere nicht	13,0%
zwei Kodierer*innen kodieren HS, andere nicht	16,0%
drei Kodierer*innen kodieren HS, andere nicht	18,4%
Übereinstimmung: ALLE kodieren HS	31,6%

In nur knapp über der Hälfte aller Fälle (52,6%) waren sich alle vier Kodierer*innen darüber einig, dass im kodierten Kommentarblock Hate Speech vorkommt bzw. nicht vorkommt: d.h. konkret waren sich 21% der Fälle alle einig, dass im Kommentarblock kein Hate Speech vorkommt und in 31,6%, dass Hate Speech vorkommt. Damit bleibt in 47,4% der Fälle unklar, ob HS vorkommt oder nicht, wobei in 16% der Fälle besonders große Uneinigkeit herrscht, da hier jeweils zwei Kodierer*innen Textstellen als Hate Speech identifizieren und die anderen beiden nicht. Erwartungsgemäß fällt auch die Intercoderreliabilität relativ schlecht aus: Krippendorf Alpha, $\alpha = 0,46$, was deutlich unter dem erwünschten Grenzwert von 0,68 liegt.

Durch den Ausschluss jener Fälle mit besonders großer Uneinigkeit (16% der Fälle) konnte die Intercoderreliabilität zwar erhöht werden, liegt aber mit $\alpha = 0,62$ nach wie vor unter den Erwartungen.

Da ein Kommentarblock in der Regel aus mehreren Interaktionsfolgen besteht, war es neben der Feststellung, ob Hate Speech im zu analysierenden Kommentarblock vorkommt oder nicht, eine weitere Herausforderung zu bestimmen, welcher konkrete Kommentar als erster Hate Speech Kommentar wahrgenommen wurde. Dies war vor allem auch deshalb entscheidend, um zu erkennen, welche User*innen überhaupt als Hater*innen wahrgenommen wurden (zum Teil war dies unklar, da auch hasserfüllte Gegenattacken vorkamen) bzw. hatte dies auch Folgen für die weitere Bewertung und Kategorisierung (z.B. Bewertung der Schwere des Hate Speech Kommentars). Die Kodierer*innen wurden daher zu aufgefordert, den ersten Hate Kommentar mit einem entsprechenden Marker zu versehen.

Um auch hier zu prüfen, in welchen Fällen Übereinstimmung in der Markierung des Hate Speech Kommentars herrscht, wurde auf Basis jener Fälle 31,6% (=158 Fälle), wo sich alle vier Kodierer*innen einig waren, dass im Kommentarblock Hate Speech vorkommt, wieder ein Summenindex (höchste Übereinstimmung bei Sume 4 bzw. 0) mit folgendem Ergebnis gebildet:

Tabelle 4

Übereinstimmung Markierung ersten HS Kommentar (n=158)	
alle vier markieren unterschiedliche HS Kommentare	2,5%
zwei kodieren den gleichen HS Kommentar, zwei jeweils andere	13,9%
jeweils zwei kodieren den gleichen HS Kommentar, unentschieden	12,0%
drei kodieren den gleichen HS Kommentar, eine anders	33,5%
alle vier markieren den gleichen HS Kommentar	38,0%

Bei jenen Fällen, wo alle vier Kodierer*innen angaben, dass Hate Speech vorliegt (n=158) stimmte der Marker nur in 38,0% (=60 Fälle) komplett überein, d.h. in 62% der Fälle waren sich die Kodierer*innen uneinig, welcher der erste Hate Speech Kommentar ist, wenn auch in etwa einem weiteren Drittel (33,5%) zumindest die Mehrheit (drei von vier Kodierer*innen) übereinstimmte (Krippendorf Alpha, $\alpha = 0,51$).

In einem weiteren Schritt wurden nun diese Fälle qualitativ analysiert, und zwar **zunächst jene 60 Fälle, bei denen sich alle vier Kodierer*innen über das Vorkommen und die Markierung von Hates Speech einig waren**. Ziel war es herauszufinden, wie sich diese Kommentarinhalte charakterisieren lassen bzw. woran hier Hate Speech eindeutig festgemacht wird. Die nachfolgend dargestellten Ergebnisse werden mit Beispielen aus dem Datenmaterial unterlegt. Um eine Identifizierung der User*innen zu verhindern, wurden die Nicknamen im gesamten Bericht anonymisiert. Von jedem User*innennamen wurden nur die ersten drei Zeichen beibehalten und durch "XXXX" erweitert. Diese Vorgehensweise ermöglicht es, direkte Bezugnahmen in Kommentaren zu erkennen, ohne die Anonymität zu verletzen.

Die analysierten Fälle enthalten überwiegend Schimpfwörter oder Kraftausdrücke oder/ und beinhalten Diskriminierungsformen, indem sie Transfeindlichkeit, Homophobie, Rassismus oder Sexismus reproduzieren, wobei zum Teil auch beides vorkommt (Reproduktion von -ismen und Beschimpfungen).

Die Kodierer*innen waren sich bei der Deklaration als Hate Speech vor allem dann einig, wenn Beschimpfungen mit Kraftausdrücken vorkamen. Häufig handelte es sich dabei um relativ kurz gehaltene negative Kommentare, dazu zählen zum Beispiel:

- Aggressive und abwertende Aufforderungen, sich nicht mehr zu äußern: „*Halt die Fresse/Halts Maul*“ mit verschiedenen Variationen und Abkürzungen, wie z.B. „*hdf*“, „*Verpiss dich*“ (in digitalen Räumen gleichbedeutend mit „Halts Maul“).

Beispiele

*ZheXXXX 1 year ago @MissBoneshaker *halt *dein *maul *fotze*

MP DXXXX 1 year ago Hdf

tobXXXX. 1 year ago HALT DIE FRESSE DU HEUHLER

- Schimpfwörter mit klarer Signalwirkung, wie z.B. „Hurensohn/-tochter“: diese schließen Abweichungen in der Schreibweise wie „*hs*“, „*hu*****sohn*“ oder „*HU**NSOHN*“ mit ein.

Beispiele

CihXXXX 11 months ago Unlustig du hurentochter

*RicXXXX 1 year ago was ein hur*****sohn*

Ein YXXXXr 4 months ago @halXXXX du Hurensohn

- Herabwürdigende Beleidigungen, die in verschiedenen Formen Ekel zum Ausdruck bringen („Ekelhaft/Widerlich“): Solche Kommentare sind teilweise auch etwas länger ausgeführt, um die Herabwürdigung einer Person zu verstärken, und sind oft sexistisch geprägt.

Beispiele

OdiXXXX 5 months ago Ekelhaft diese widerliche Plastik-Hoe ... Da bekomme ich Sackläuse

SalXXXX 5 months ago @ineXXXX deine mom ekelt mich beim f!ck3n an

SveXXXX 9 months ago Sry, das ist unsachlich, aber der Typ ist richtig ekelig. Das manifestierte Wort „igitt“

- Herabwürdigende Beleidigungen, die einer Person den Intellekt/ die geistigen Fähigkeiten absprechen: hier sind am häufigsten auch direkte Bezüge entweder zum Video oder zu anderen Kommentaren zu finden, d.h. solche Kommentare sind manchmal auch Ausdruck einer Meinungsverschiedenheit.

Beispiele

OSTXXXX 9 months ago Wie dumm seid ihr eigentlich alle, glaubt ihr ernsthaft Deutschland zerstört die Welt alleine und das die Welt „gerettet“ wird wenn niemand die CDU/CSU wählt

HypXXXX 9 months ago Diese dummen Pisser vor dem Autohaus (2.2) . alter ich könnte so ausrasten, alleine wenn ich die hier nur im Video sehe

Die genannten Formen von Beleidigungen bzw. die Benutzung von Schimpfwörtern, wie auch z.B. „Idiot“, „Lappen“, „Depp“ etc. sind in der Gesellschaft als klar negative Inhalte deklariert und daher auch leicht als solche einzuordnen. Schwieriger wird eine solche Kategorisierung als negativ bei diskriminierenden Inhalten: Am häufigsten vorzufinden sind hier Transfeindlichkeit und Sexismus, wobei drei Viertel aller sexistischen Kommentare zusätzlich eine Beleidigung beinhalten und ihre Einordnung als negativ damit wieder einfacher vorzunehmen ist.

Beispiele

SalXXXX 5 months ago @ineXXXX deine mom ekelt mich beim f!ck3n an

PatXXXX 1 month ago Brüste ok, aber Gesicht sieht inzwischen aus als ob mein arsch drüber gefahren ist

zSwXXXX btw 5 months ago @DeaXXXX geh an aids sterben lol

Was als Diskriminierung aufgefasst wird und was nicht, wird auch durch verschiedene persönliche oder bildungsbezogene Hintergründe, Ideologien und Einstellungen beeinflusst und kann daher nicht in jedem Fall eindeutig einzuordnet werden. Dies betrifft – trotz umfassender Schulungen – auch die Kodierung durch das wissenschaftliche Forschungsteam und zeigt sich somit auch in der **Analyse jener Fälle (n=98), wo sich die Kodierer*innen in ihren Bewertungen uneinig waren**. Dazu zählen:

- In der Gesellschaft noch nicht klar eingeordnete Diskriminierungsformen, wie zum Beispiel Transfeindlichkeit: Solche Formen sind nicht immer leicht erkennbar bzw. werden verschiedene transfeindliche Aussagen als unterschiedlich schwer bewertet.

Beispiele

The XXXX 8 months ago Wieso zur Hölle dachte ich seit dem ersten Video dass das ein Kerl ist?... ich merk erst jetzt das es eine frau ist IOOOO!

BusXXXX 8 months ago (edited) @The XXXX sei froh dass du den ganzen stress von vor 2-3 jahren um diesen menschen (fühle mich nicht wohl "frau" zu sagen) nicht mitbekommen hast, das war cringe des todes mit ner rotweinfahne

- Diskriminierungsformen, die in „Grundsatzdiskussionen“ eingebettet sind: z.B. islamfeindliche und xenophobe Ressentiments. Oft finden sich solche negativen Inhalte in sehr langen Blöcken mit vielen Kommentaren von vielen unterschiedlichen User*innen, wodurch es schwierig wird, den Überblick zu behalten. Zudem fehlen oft auch signalisierende Schimpfwörter, sodass der Eindruck entsteht, dass prinzipiell „nur diskutiert“ wird, dabei aber sehr häufig diskriminierendes Gedankengut reproduziert wird.

Beispiele

„Gerade weil der Islam in vielen Sachen nicht konform geht mit anderen Religionen und Rechten ist es doch mehr als fragwürdig wie man einfach gutgläubig sagen kann, dass sie ihre Moscheen bekommen sollen“

„Es ist halt nicht nur, dass sie eine andere Ansichtswiese haben sondern in vielen Punkten gegen die deutschen Grundgesetze verstoßen“

„Bei der wohl menschenverachtendsten Ideologie des Planeten, dem Islam Man muss dieses geistesranke Weltbild sich nicht aufzwingen lassen“

- Diskriminierungsformen, die je nach ideologischem Standpunkt unterschiedlich bewertet werden: Auch hier sind negative Inhalte oft innerhalb langer, politischer Diskussionen zu finden. Vor allem dann, wenn rechte Standpunkte vertreten werden, ist es oft schwierig, diese einheitlich zu kodieren.

Beispiele

SmeabsenSchaf 1 year ago Ja immer ist "rechts" das schlimme, scheis auf Terror von der Antifa Die Antifa ist auch schlimm Sie meinen, dass sie gegen Diskriminierung sind, aber sind gegen Deutsche Sie meinen, dass sie gegen Faschismus sind, betreiben aber selber Faschismus Sie verstecken sich hinter gut klingenden Sachen, wollen im Grunde aber nur Chaos und Zerstörung

Die Ergebnisse verdeutlichen, wie schwierig es ist, Hate Speech tatsächlich als solche zu identifizieren, vor allem dann, wenn „hinweisgebende“ signalisierende Schimpfwörter und Kraftausdrücke fehlen. Vor diesem Hintergrund wäre auch zu diskutieren, ob das gezielte, automatisierte Löschen solcher Kraftausdrücke im Hinblick auf spontanes Counter Speech Verhalten nicht sogar kontraproduktiv sein kann, da eindeutige „Markierungen“ als Hate Speech fehlen. Gerade die Reproduktion von -ismen und negativem Gedankengut, die sich oft in komplexen Zusammenhängen darstellt, braucht für deren Bewertung eine notwendige Sensibilisierung, um diese als toxische Inhalte zu erkennen und sie auf breiter gesellschaftlicher Ebene als Hassrede deklarieren und erkennen zu können. Um dies bei Jugendlichen zu erreichen, ist insbesondere das Bildungswesen gefordert.

2.2.2 Identifikation von Counter Speech

Ob Counter Speech vorliegt oder nicht wurde nur dann kodiert, wenn zuvor angegeben wurde, dass Hate Speech vorliegt. Wie die Ausführungen im vorigen Kapitel zeigen, fällt dies aber sehr unterschiedlich aus und ist ausgesprochen schwierig zu identifizieren. In jenen 158 Fällen, wo sich alle vier Kodierer*innen einig waren, dass HS vorliegt (unabhängig von der Markierung des konkreten ersten Hate Speech Kommentars), wurde nun geprüft, inwieweit sich die Kodierer*innen einig waren, dass Counter Speech vorliegt oder nicht. Dazu wurde das Vorkommen von Counter Speech wieder pro Kodierer*in dichotomisiert in (1 = ja, eine oder mehrere Personen üben Counter Speech aus; 0= nein, es kommt keine Counter Speech vor) und anschließend ein Summenindex gebildet (höchste Übereinstimmung bei Summe 4 bzw. 0).

Tabelle 5

Übereinstimmung in der CS-Kodierung (n=158)	
Übereinstimmung: NIEMAND kodiert CS	14,6%
eine Kodierer*in kodiert CS, andere nicht	7,0%
zwei Kodierer*innen kodieren CS, andere nicht	10,8%
drei Kodierer*innen kodieren CS, andere nicht	20,3%
Übereinstimmung: ALLE kodieren CS	47,5%

Wie Tabelle 5 zeigt, waren sich die Kodierer*innen in insgesamt 62,1% der Fälle über die Bewertung einig: In 14,6% der Fälle haben alle angegeben, dass – trotz Hate Speech - keine Counter Speech vorliegt und in 47,5% der Fälle waren sich die Kodierer*innen einig, dass Counter Speech vorkommt. In 38,1% der Fälle waren war die Bewertung uneinig, wobei sich in 20,3% der Fälle die Mehrheit, das heißt drei von vier Kodierer*innen einig waren (Krippendorf Alpha, $\alpha = 0,52$). Im Vergleich zur Identifikation von Hate Speech fällt hier also das Verhältnis von Einigkeit und Uneinigkeit etwas besser aus.

In einem nächsten Schritt wurde auch hier geprüft, ob die Kodierer*innen bezogen auf jene 47,5% der Fälle (n=75), wo sich alle einig waren, dass Counter Speech vorliegt, auch darüber übereinstimmen, welcher konkrete Kommentar als erster Counter Speech Kommentar zu markieren ist. Dazu wurde wieder ein Summenindex (höchste Übereinstimmung bei Sume 4 bzw. 0) gebildet:

Tabelle 6

Übereinstimmung Markierung erster CS Kommentar (n=75)	
alle vier markieren unterschiedliche CS Kommentare	0,0%
zwei kodieren den gleichen CS Kommentar, zwei jeweils andere	12,0%
jeweils zwei kodieren den gleichen CS Kommentar, unentschieden	8,0%
drei kodieren den gleichen CS Kommentar, eine anders	30,7%
alle vier markieren den gleichen CS Kommentar	49,3%

In knapp die Hälfte der Fälle (49,3%) waren sich alle Kodierer*innen darüber einig, welcher Kommentar den ersten Counter Speech Kommentar darstellt, bei weiteren 30,7% waren sich zumindest drei der vier Kodierer*innen einig. Im Vergleich zur Markierung des ersten Hate Speech Kommentars ist auch hier die Übereinstimmung etwas höher (Krippendorf Alpha, $\alpha = 0,59$).

Auch hier wurden die Fälle qualitativ analysiert, und zwar **sowohl jene 49,3% der Fälle (n=37), die von allen vier Kodierer*innen als Counter Speech identifiziert wurden, als auch alle Hate Speech / Counter Speech Paare (n=30), bei denen sich alle Kodierer*innen einig waren.**

Die Ergebnisse zeigen, dass sich die Kodierer*innen an folgenden Charakteristiken orientierten, um Kommentare als Counter Speech zu deklarieren, wobei wieder zahlreiche Überschneidungen zu finden sind:

- **Direkter Bezug zu/ direkte Antwort auf Hate Speech:** Der*die Hater*in wird im Counter Speech Kommentar getagged, das heißt der Anfang jedes solchen Counter Speech Kommentars ist der User*innenname der Person, welche den ersten Hate Speech Kommentar verfasst hat. Dies kommt in der überwiegenden Mehrheit der Counter Speech Kommentare vor.

Beispiele:

Hate Speech	Counter Speech
badXXXX 11 months ago ich hoffe das bei dem wixxer was dran ist und er für einpaar jährchen hinter schwedischen gardinen darf dann noch einpaar mal die seife aufheben lassen und von hinten nehmen im knast bis er irgendwann in der ecke liegt dieser spaso -	MikXXXX 11 months ago badXXXX chill
MonXXXX 2 months ago Vergleich Jesus nicht mit so einem gay *****	ShoXXXX 2 months ago @MonXXXX Mach das Maul nicht so weit auf Du möchtestgern Internetrambo
KatXXXX 1 week ago Halt doch deine fresse	JuXXXX 1 week ago KatXXXX bisschen salzig unterwegs heute

- **Direkte Folgekommentare:** Zusätzlich wurden mehr als die Hälfte der Counterspeech Kommentare in direkter Folge zum Hatespeech Kommentar verfasst, aber nicht getagged, das heißt es gab keine weiteren Kommentare dazwischen.

Beispiele:

Hate Speech	Counter Speech
SebXXXX 1 year ago blas ihm doch ein	OguXXXX 1 year ago Deine Mutter bläst ihm einen, du Hund
AilXXXX 7 months ago TANZVERBOT IST SCHEIßE DU BIST BESSER	DavXXXX 7 months ago Nein

- Ebenfalls wird in der überwiegenden Mehrheit der Fälle die **inhaltliche Ausrichtung** des Hate Speech Kommentars auch im Counter Speech Kommentar noch einmal aufgegriffen. Häufig wird dies in Form von Gegenattacke formuliert.

Beispiele:

Hate Speech	Counter Speech
SveXXXX 9 months ago Sry, das ist unsachlich, aber der Typ ist richtig ekelig, Das manifestierte Wort „igitt“	PauXXXX 7 months ago Vielleicht bist du auch ekelig, weil du so gut bescheidet weiss
SerXXXX 3 years ago Starshine Xo iBlali ist ein Nazi	WilXXXX 3 years ago SerXXXX dein Vater ist nazi
KomXXXX 1 year ago Sorry aber das ist der misseste w**er der Welt. Das ist komplett abartig sowas machst du doch nicht. Da kriegt man ja die voll griese	CarXXXX 1 year ago Bei deiner Rechtschreibung bekommt man auch die Krise <-- So schreibt man das übrigens
CihXXXX 11 months ago Unlustig du hurentochter	BudXXXX 11 months ago CihXXXX unlustig du HURENSOHN

Die inhaltliche Ausrichtung wird aber auch für die Formulierung von Gegenargumenten aufgegriffen, insbesondere dann, wenn die Reproduktion von -ismen vorgegangen wird.

Beispiele:

Hate Speech	Counter Speech
KraXXXX 11 months ago Ganz ehrlich Leute Hand aufs Herz, so eine Person darf man nicht als Frau bezeichnen Wie fernab der Realität lebt ihr den bitte Was hat sie noch mit einer Frau zu tun Nichts rein gar nichts Könnte genau so eine transe sein	BlīXXXX 11 months ago @KraXXXX Sie ist eine Frau da sie weiblichen Geschlechts ist In welcher Welt lebst du bitte, dass du immer noch nach einem traurigen Klichee von Menschen urteilst und alle gleich sein müssen
MalXXXX 2 months ago PerXXXX (Anm: YouTuberin) du bist keine Frau du hast ein kleinen Penis und heißt Jonas	MelXXXX 2 months ago MalXXXX doch sie ist rechtlich eine Frau in ihrem Ausweis steht Lynn und weiblich also hört auf so eine scheiße zu labern wenn ihr keine Ahnung habt

- **Klare Benennung als Hate Speech:** In solchen Fällen wird entweder der*die Absender*in als Hater*in, der Inhalt als hasserfüllt deklariert, oder eine Handlung als „haten“ bezeichnet.

Beispiele:

Hate Speech	Counter Speech
PanXXXX 1 year ago HarXXXX du weißt schon das das eine Meinung ist ich bin gegen schwule aber wie sie nicht foltern oder was auch immer	YuzXXXX 1 year ago @PanXXXX Homophobie ist keine Meinung, sondern Hass. Du magst Schwule nicht Okay, kein Ding, verstehe ich nicht, aber ist halt deine Meinung. Aber gegen die Existenz einer Personengruppe zu sein, dass ist nicht tolerierbar
OliXXXX 1 year ago Jetzt mal ganz ehrlich ist doch klar das das gefragt wirddu bist echt like geil so was zu schreiben du möchtest doch nur das Bibi und juli dein Video liken	UniXXXX 1 year ago @OliXXXX hält die fresse Warum gibt es immer hater für solche REALEN Kommentare Du bist hier due likegeile sau
HazXXXX 1 month ago Du bist der Grund warum es Kondome gibt du semmel red mal anständig	:pXXXX 1 month ago Er hat sogar ein Reaction Video aufgenommen wo er zeigt dassner den Spaß von Abk versteht und nichts ernstgenommen, also was hatest du

Um Counter Speech identifizieren zu können, braucht es demnach immer den gemeinsamen Kontext mit Hate Speech. Da Counter Speech häufig in Form von (wenig wünschenswerten) Gegenattacken formuliert ist (ca. die Hälfte der analysierten Fälle), wäre dies ohne die gleichzeitige Identifizierung des Hate Speech Kommentars praktisch nicht möglich. Eine solche Identifizierung gelingt aber bei klaren und kürzeren Kommentarblöcken deutlich leichter als bei komplexen Interaktionsverläufen. Diese Problematik wurde bereits im vorangegangenen Abschnitt thematisiert und zeigt sich vor allem bei langen Kommentarblöcken, in denen „Grundsatzdiskussionen“ beispielsweise zu Homosexualität geführt werden, oder oft latent diskriminierendes Gedankengut reproduziert wird. In solchen Fällen greifen selbst die oben genannten Charakteristiken zur Zuordnung nicht mehr, wie das folgende Beispiel zeigt:

Beispiel: unklare Zuordnung, was Hate und Counter Speech ist
AndXXXX 3 years ago Wir wollen keine Homophobie, löscht dich
CSHXXXX 3 years ago Mr Löscht du dich

2.2.3 Rollenzuordnung von Opfern und Counter Speaker*innen

Aufgabe der Kodierer*innen war es auch festzuhalten, wem diese sowohl die Opfer- als auch die Counter Speech Rolle zuordneten, wobei hier bezüglich des Opfers zwischen User*innen und dritten Personen (z.B. YouTuber*in, Content-Creator*innen, Prominente, andere Personengruppen) unterschieden wurde und bei Counter Speaker*innen, ob diese gleichzeitig auch Opfer des Übergriffs (bzw. sich der Opfergruppe zugehörig fühlen) sind oder nicht. Dabei waren in beiden Fällen auch Mehrfachnennungen zulässig, beispielsweise wenn sich ein Hasskommentar sowohl an ein*e User*in, als auch an ein*e YouTuber*in richtete. Wie die nachfolgenden Tabellen zeigen, gab es auch hier zum Teil Uneinigkeiten in der Zuordnung der Rollen, vor allem dann, wenn keine Einigkeit bestand, welches der erste Hate Kommentar war.

Tabelle 7

Richtet sich Hate Speech persönlich gegen andere User*innen?		
	alle einige, dass HS vorkommt: n=158	alle einige, welcher der erste HS Kommentar ist n=60
ja - alle einig	31,4%	45,0%
nein - alle einig	33,3%	40,0%
uneinig	35,3%	15,0%

Auch wenn sich alle Kodierer*innen in der Markierung des Hate Speech Kommentars einig sind, gibt es immer noch in 15% der Fälle Uneinigkeit darüber, ob sich der Hate Speech Kommentar persönlich gegen andere User*innen richtet oder nicht.

Tabelle 8

Richtet sich Hate Speech gegen YouTuber*in, andere Content-Creator*innen, Prominente, bestimmte Personengruppen?		
	alle einige, dass HS vorkommt: n=158	alle einige, welcher der erste HS Kommentar ist n=60
ja - alle einig	41,7%	45,0%
nein - alle einig	30,1%	43,3%
uneinig	28,2%	11,7%

Auch bei der Zuordnung, ob dritte Personen Opfer von Hate Speech sind, gibt es immer noch bei 11,7% der Fälle Unklarheiten, obwohl sich alle Kodierer*innen über die Markierung des Hate Speech Kommentars einig sind.

Bezogen auf die Ausmaße an Betroffenheit scheint Hate Speech gegenüber User*innen etwa im selben Ausmaß gängig zu sein wie gegen dritte Personen (jeweils 45%).

Tabelle 9

Ist / sind Counter Speaker*innen auch Opfer von Hate Speech bzw. der Opfer-Gruppe zugehörig?		
	alle einige, dass CS vorkommt: n=75	alle einig, welcher der erste CS Kommentar ist n=37
ja - alle einig	12,0%	13,5%
nein - alle einig	52,0%	56,8%
uneinig	36,0%	29,7%

Diese Unsicherheiten sind bei der Zuordnung der Counter Speaker*innen deutlich größer: in 36,0% (bei Einigkeit, dass Counter Speech grundsätzlich vorkommt) bzw. in 29,7% der Fälle (bei Einigkeit darüber, was der erste Counter Speech Kommentar konkret ist) sind sich die Kodierer*innen uneinig darüber, ob die Counter Speaker*innen gleichzeitig auch Opfer sind.

Tabelle 10

Ist / sind Counter Speaker*innen dritte Personen (nicht Opfer)?		
	alle einige, dass CS vorkommt: n=75	alle einig, welches der erste CS Kommentar ist n=37
ja - alle einig	69,3%	73,0%
nein - alle einig	4,0%	5,4%
uneinig	26,9%	21,6%

Ähnlich zeigt sich das auch bei der Bewertung darüber, ob die Counter Speaker*innen andere Personen sind, die nicht zur Opfergruppe zählen: Hier gibt es in 26,9% (bei Einigkeit, dass Counter Speech vorkommt) bzw. in 21,6% der Fälle (bei Einigkeit über die Markierung des ersten Counter Speech Kommentars) Unsicherheiten bezüglich der Kodierung.

Prinzipiell deuten die Zahlen darauf hin, dass Counter Speaker*innen mehrheitlich Personen sind, die nicht gleichzeitig auch Opfer der Übergriffe sind. Ein möglicher Grund dafür könnte sein, dass es in vielen Fällen schwierig ist sicherzugehen, dass die Counter Speaker*innen auch Teil der Opfergruppe sind. Wenn der*die Hater*in sich auf eine*n konkrete*n User*in bezieht und diese dann als Counter Speaker*in auftritt, ist eine klare Zuordnung möglich – dies ist bei weniger gerichteten Hasskommentaren wesentlich schwieriger. Häufig ist unklar, ob eine Person beispielsweise als Counter Speaker*in auf einen rassistischen/ homophoben Kommentar reagiert, weil sie selbst Teil dieser betroffenen Gruppe ist oder ob sie als nicht betroffene Person reagiert. In vereinzelten Fällen wurde diese Information von den Counter Speaker*innen in ihre Kommentare verwoben, indem sie sich ganz klar als Teil der Gruppe bezeichneten. In jenen Fällen, in denen nicht eindeutig war, gegen wen sich der Hate-Kommentar richtet, konnte auch nicht mit Sicherheit eruiert werden, ob es sich bei den Counter Speaker*innen um direkt betroffene Personen handelt.

2.2.4 Darstellung der Komplexität von Hate Speech – Counter Speech Interaktionen

Die Kodierer*innen wurden gebeten, die Interaktionsfolgen der analysierten Kommentarblöcke entsprechenden Interaktionsketten zuzuordnen. Damit sollten einerseits die Komplexität der Interaktionsverläufe dokumentiert werden, andererseits wurden diese Interaktionsketten in weiterer Folge herangezogen, um die den weiteren Verlauf bzw. die Entwicklung von Hate Speech – Counter Speech Interaktionen zu dokumentieren.

Tabelle 11

Komplexität von Hate Speech - Counter Speech Interaktionen (n=jene Fälle, wo HS identifiziert wurde)					
	Kodierer*in 1 (n=346)	Kodierer*in 2 (n=221)	Kodierer*in 3 (n=338)	Kodierer*in 4 (n=228)	Mittelwert
nur HS	31,5%	27,6%	37,0%	44,7%	35,2%
HS - CS	21,1%	20,8%	20,1%	21,9%	21,0%
HS- CS - HC	12,1%	14,9%	13,0%	10,5%	12,6%
HS- CS - HC - CS	9,8%	9,0%	8,0%	10,1%	9,2%
längere Interaktionen	25,4%	27,6%	21,9%	11,8%	21,7%

Durch die Uneinigkeit bei der Identifikation von Hate Speech und Counter Speech wurden die Interaktionsketten zum Teil unterschiedlich wahrgenommen. Die durchschnittliche Bewertung zeigt aber, dass in etwas mehr als einem Drittel der Fälle (35,2%) lediglich Hate Speech ohne weitere Counter Speech Interaktionen vorkommt und in knapp zwei Drittel der Fälle (64,8%) zusätzlich auch Counter Speech identifiziert wird – wie weiter oben gezeigt, zählen dazu aber auch ebenso hasserfüllte Gegenattacken. In durchschnittlich 30,9% der Fälle sind längere Interaktionsfolgen von zumindest vier Gliedern (HS – CS – HS – CS oder länger) zu finden, was auf längere Auseinandersetzungen zwischen Hater*innen und Counter Speaker*innen schließen lässt.

Bei längeren Interaktionsketten wurde außerdem geprüft, ob diese mit einem Hate Speech oder Counter Speech Kommentar enden. Wie Tabelle 12 zeigt, enden diese überwiegend mit Hate Speech Kommentaren (56,5% der Fälle).

Tabelle 12

Abschluss von längeren HS/CS Interaktionen (n=Interaktionsketten länger als HS-CS-HS-CS)					
	Kodierer*in 1 (n=90)	Kodierer*in 2 (n=58)	Kodierer*in 3 (n=75)	Kodierer*in 4 (n=28)	Mittelwert
HS letzter Kommentar	55,6%	43,1%	56,0%	71,4%	56,5%
CS letzter Kommentar	44,4%	56,9%	44,0%	28,6%	43,5%

2.2.5 Bewertung der Schwere von Hate Speech

Nach der Identifizierung von Hate Speech war es eine weitere Aufgabe der Kodierer*innen, Hate Speech Kommentare nach ihrer Schwere zu bewerten, wobei hier zwischen „leicht bis mittel schwer“ und „schwer“ einzustufen war. Als Bewertungsgrundlage wurde im Codebook ein gemeinsames Prozedere ausgearbeitet und mit Beispielen versehen. Wie folgende Tabelle zeigt, gab es zwischen den Kodierer*innen dennoch sehr große Unstimmigkeiten:

Tabelle 13

Als wie schwer ist der erste Hate Speech Kommentar einzustufen?		
	alle einig, dass HS vorkommt: n=158	alle einig, welcher der erste HS Kommentar ist n=60
Alle (4 mal) HS leicht	10,9%	10,0%
3 mal HS leicht, 1 mal schwer	41,0%	45,0%
2 mal HS leicht, 2 mal HS schwer	24,4%	16,7%
3 mal HS schwer, 1 mal HS leicht	12,2%	15,0%
Alle (4 mal) HS schwer	11,5%	13,3%

In jenen Fällen (n=158), in denen Einigkeit bestand, dass Hate Speech im Kommentarblock vorkommt, waren sich insgesamt nur in 22,4% alle Kodierer*innen über die Bewertung der Schwere einig (10,9% leicht, 11,5% schwer). Sehr ähnlich zeigt sich das Verhältnis in jenen Fällen, wo sich die Kodierer*innen nicht nur einig waren, dass Hate Speech vorkam, sondern auch welches der erste HS Kommentar war (n=60): auch hier waren sich die Kodierer*innen in nur 23,3% über die Bewertung einig (10,0% leicht, 13,3% schwer). In allen anderen Fällen gab es immer zumindest eine*n Kodierer*in, welche*r die Bewertung anderes einschätzte. Insgesamt scheint das Verhältnis von leichten und schweren Übergriffen zugunsten von leichten Übergriffen auszufallen: In 51,9% der Fälle, bei denen sich die Kodierer*innen einig waren, dass HS vorkam bzw. 55,0% der Fälle, bei denen außerdem Einigkeit über Markierung von HS herrschte, kodierte zumindest die Mehrheit der Kodierer*innen einen leichten und in 23,7% (Einigkeit HS kommt vor) bzw. 28,3% (Einigkeit über Markierung von HS) einen schweren Übergriff.

Um festzustellen, ob für die Kodierer*innen neben den Definitionen im Codebook noch weitere Charakteristika eine Rolle spielen, was als leicht oder schwer zu bewerten ist, wurde die inhaltliche Ausrichtung der Hate Speech Kommentare qualitativ analysiert.

Prinzipiell enthalten alle Kommentare Beleidigungen, Beschimpfungen oder reproduzieren -ismen bzw. negatives Gedankengut. Für die Einstufung als leicht oder schwer scheinen folgende Merkmale eine Rolle zu spielen:

- Frequenz von Beleidigungen bzw. in Kombination mit der Reproduktion von -ismen innerhalb eines Kommentars: Je mehr negative Inhalte vorkommen, desto eher wird etwas als schwerer Hate Speech Kommentar eingestuft:

Beispiele für einstimmige Einstufung als leicht:

*Jd 1XXXX 3 weeks ago @DarXXXX er hat es doch nicht verstanden was rewi gesagt hat du lelleck
OliXXXX 1 year ago Jetzt mal ganz ehrlich ist doch klar das das gefragt wirddu bist echt
like geil so was zu schreiben du möchtest doch nur das Bibi und juli dein Video liken*

Beispiel für uneinige Einstufung als leicht oder schwer:

*SveXXXX 9 months ago Sry, das ist unsachlich, aber der Typ ist richtig ekelig. Das manifestierte Wort
„igitt“*

Beispiele für einstimmige Einstufung als schwer:

*Dvd XXXX 10 months ago mizXXXXX du HURENSOHN was für Toll die sollen verrecken
OdiXXXX 5 months ago Ekelhaft diese widerliche Plastik-Hoe ... Da bekomme ich Sackläuse*

In diesem Zusammenhang kann auch das Verfassen des Kommentars in Capslock (alles in Großbuchstaben) ausschlaggebend sein: Eine solche Schreibweise wird oft als Schreien und damit als wesentlich aggressiver aufgefasst, was sich auch auf die Bewertung der Schwere auswirken kann, wie folgende Beispiele zeigen.

Beispiele:

*MP DXXXX 1 year ago Hdf
KatXXXX 1 week ago Halt doch deine fresse
tobXXXX. 1 year ago HALT DIE FRESSE DU HEUCHLER*

- Inhaltliche Ausrichtung von Beschimpfungen: Werden einfache, „gängige“ Schimpfwörter verwendet, wie sie oft auch im Kontext von Diskussionen entstehen, so werden Kommentare eher als leicht eingestuft als bei Ausdrucksweisen, die zusätzlich herabwürdigen, bedrohlich wirken oder etwas Schlechtes (z.B. den Tod) wünschen und damit klar eine Normübertretung darstellen.

Beispiele für einstimmige Einstufung als leicht:

*halXXXX 7 months ago Der kek hat in seinem Leben noch nie richtig gearbeitet und erzählt was von
Stress, muss los*

HypXXXX 9 months ago Diese dummen Pisser vor dem Autohaus . alter ich könnte so ausrasten, alleine wenn ich die hier nur im Video sehe.

Beispiele für uneinige Einstufung als leicht oder schwer:

CoDXXXX 6 months ago junge was laberst du, wie kann man nur so personen kennen du 11 jährige jungfrau du

*RicXXXX 1 year ago was ein hur*****sohn*

Beispiel für einstimmige Einstufung als schwer:

zSwXXXX btw 5 months ago @DeaXXXX geh an aids sterben lol

- Klare Deklaration als negatives Gedankengut: Wie bereits ausgeführt, werden nicht alle Reproduktionen von -ismen oder negativem Gedankengut von den Kodierer*innen klar erkannt bzw. werden diese oft als unterschiedlich schwer eingestuft. Wie die folgenden Beispiele zeigen, wird z.B. nationalsozialistisches Gedankengut klarer als schwer eingestuft als Transfeindlichkeit, was oft zu uneinheitlichen Bewertungen führt:

Beispiel für Uneinigkeit in der Bewertung ob schwer oder leicht:

PanXXXX 1 year ago HarXXXX du weißt schon das das eine Meinung ist ich bin gegen schwule aber wie sie nicht foltern oder was auch immer

Beispiel für einstimmige Einstufung als schwer:

badXXXX 11 months ago ich hoffe das bei dem wixxer was dran ist und er für einpaar jährchen hinter schwedischen gardinen darf. dann noch einpaar mal die seife aufheben lassen und von hinten nehmen im knast bis er irgendwann in der ecke liegt dieser spaso

Auch bei der Bewertung der Schwere des letzten Hate Speech Kommentars in längeren Interaktionsketten gibt es Unstimmigkeiten zwischen den Kodierer*innen (Tabelle 14). Auf inhaltlicher Ebene zeigt sich insgesamt aber, dass Hate Speech im Laufe der Interaktionen verstärkt oder zumindest bestätigt wird und positive Entwicklungen im Hinblick auf einsichtes Verhalten nur marginal wahrgenommen werden.

Tabelle 14

Bewertung des letzten Hate Speech Kommentars (n=Interaktionsketten länger als HS-CS-HS-CS)					
	Kodierer*in 1 (n=90)	Kodierer*in 2 (n=58)	Kodierer*in 3 (n=75)	Kodierer*in 4 (n=28)	Mittelwert
mit schwerem Übergriff verstärkt	53,3%	19,0%	13,5%	10,7%	24,1%
mit harmlosem / mittlerem Übergriff	20,0%	37,9%	59,5%	39,3%	39,2%
ohne weiteren Übergriff verstärkt	14,4%	39,7%	14,9%	50,0%	29,8%
distanzierend / einsichtig /	12,2%	3,4%	12,2%	0,0%	7,0%

Gerade bei der Bewertung der Schwere von Hate Speech ist davon auszugehen, dass bei User*innen noch stärkere Differenzen in der Bewertung zu finden sind als bei den Kodierer*innen, denen im Rahmen der Schulung ähnliche Bewertungsmaßstäbe vermittelt werden sollten. Als wie schwerwiegend Hate Speech bewertet wird, kann sich aber auch auf die Motivation von User*innen, Counter Speech zu betreiben, auswirken. Umso wichtiger wäre gerade für Jugendliche eine entsprechende Sensibilisierung der Bewertung von Hate Speech.

2.2.6 Kategorisierung und Bewertung von Counter Speech Strategien

Um die inhaltliche Ausrichtung von Counter Speech festzumachen, wurden die Kodierer*innen gebeten, diese entsprechenden Kategorien zuzuordnen (Mehrfachnennungen möglich, wurde für jedes CS-Glied in der Interaktionskette bewertet). Damit sollten auch Veränderungen in Abhängigkeit der Merkmale der Hate Speech Kommentare sowie des weiteren Interaktionsverlaufs dokumentiert werden.

Tabelle 15

Kategorisierung von Counter Speech Strategien (n = CS identifiziert; Mehrfachantworten)					
	Kodierer*in 1 (n=237)	Kodierer*in 2 (n=160)	Kodierer*in 3 (n=213)	Kodierer*in 4 (n=124)	Mittelwert
Humor / Sarkasmus / Ironie	17,3%	15,6%	15,5%	20,2%	17,2%
ins Gewissen reden / Mora	42,2%	3,1%	5,2%	12,9%	15,9%
Verständnis für Hater*in	8,4%	0,6%	3,3%	1,6%	3,5%
aggressiv / feindlich / Gegenattacke	46,6%	38,1%	51,2%	58,1%	48,5%
sachlich / Fakten / neutral nachfragen	51,9%	57,5%	47,9%	39,5%	49,0%
Deeskalation / auffordern aufzuhören	11,4%	10,6%	7,5%	6,5%	9,0%

Durchschnittlich bewerteten die Kodierer*innen die identifizierten Counter Speech Kommentare in jeweils knapp der Hälfte der Fälle als aggressiv (48,5%) oder/und sachlich (49,0%). Andere Strategien kommen in einem verhältnismäßig geringen Ausmaß vor, erwähnenswert sind humorvolle Reaktionen (17,2%). Zwischen den einzelnen Kodierer*innen variieren die Bewertungen erwartungsgemäß aber wieder zum Teil beträchtlich.

Tabelle 16

Counter Speech Strategien als Folge von schwerer Hate Speech (n = HS schwer; Mehrfachantworten)					
	Kodierer*in 1 (n=200)	Kodierer*in 2 (n=49)	Kodierer*in 3 (n=30)	Kodierer*in 4 (n=35)	Mittelwert
kein Counter Speech	29,8%	27,9%	52,5%	41,7%	38,0%
Humor / Sarkasmus / Ironie	15,0%	8,2%	20,0%	22,9%	16,5%
ins Gewissen reden / Mora	43,5%	2,0%	10,0%	17,1%	18,2%
Verständnis für Hater*in	8,5%	0,0%	6,7%	2,9%	4,5%
aggressiv / feindlich / Gegenattacke	43,5%	46,9%	50,0%	57,1%	49,4%
sachlich / Fakten / neutral nachfragen	55,0%	55,1%	50,0%	45,7%	51,5%
Deeskalation / auffordern aufzuhören	12,5%	16,3%	6,7%	5,7%	10,3%

Wie Tabelle 16 zeigt, ändert sich dieses Verhältnis kaum, wenn Counter Speech als Reaktion auf schwere Hate Speech Kommentare betrachtet werden: Überwiegend bewerten die Kodierer*innen die Counter Speech Reaktionen entweder als aggressiv (49,4%) oder sachlich (51,5%). Kodierer*in 1 kodiert in einem relativ hohen Ausmaß auch „ins Gewissen reden“ (43,5%) als relevante Strategie, was allerdings von der Ausrichtung her ebenfalls in einem Spektrum von sachlich bis aggressiv gestaltet werden kann. Auffällig ist, dass in der durchschnittlichen Bewertung in 38% der Fälle nicht auf schwere Hate Speech reagiert wird bzw. kein Counter Speech von den Kodierer*innen identifiziert wurde.

Tabelle 17

Counter Speech Strategien als Folge von leichter Hate Speech (n = HS leicht; Mehrfachantworten)					
	Kodierer*in 1 (n=37)	Kodierer*in 2 (n=111)	Kodierer*in 3 (n=182)	Kodierer*in 4 (n=89)	Mittelwert
kein Counter Speech	39,3%	27,0%	34,1%	46,4%	36,7%
Humor / Sarkasmus / Ironie	29,7%	18,9%	14,8%	19,1%	20,6%
ins Gewissen reden / Mora	35,1%	3,6%	4,4%	11,2%	13,6%
Verständnis für Hater*in	8,1%	0,9%	2,7%	1,1%	3,2%
aggressiv / feindlich / Gegenattacke	62,2%	34,2%	51,1%	58,4%	51,5%
sachlich / Fakten / neutral nachfragen	35,1%	58,6%	47,8%	37,1%	44,7%
Deeskalation / auffordern aufzuhören	5,4%	8,1%	7,7%	6,7%	7,0%

Wird nun Counter Speech in Reaktion auf leichte Hate Speech Kommentare untersucht, zeigen sich nur leichte Veränderungen, indem der Anteil an durch die Kodierer*innen als sachlich bewertete Reaktionen durchschnittlichen etwas zurückgeht (44,7%), dafür aber aggressive Reaktionen (51,5%), aber auch

humorvolle Reaktionen (20,6%) tendenziell ansteigen. Eine mögliche Erklärung für diese leichten Veränderungen könnte sein, dass es bei leichteren Übergriffen (und damit einer vermeintlich harmlosen Person in der Rolle als Hater*in) auch für Counter Speaker*innen eine geringere Hemmschwelle gibt, mit aggressiven oder sich lustig machenden Gegenattacken zu reagieren.

Der Anteil jener, die bei leichter Hate Speech nicht reagiert haben, beträgt auch hier wieder mehr als ein Drittel (36,7%) bzw. wurde in diesen Fällen keine Counter Speech Reaktion erkannt.

In einem nächsten Schritt wurde nach Hinweisen gesucht, inwieweit unterschiedliche Counter Speech Strategien auch unterschiedliche Hate Speech Reaktionen bedingen, wobei hier aufgrund der stark variierenden Bewertungen der Kodierer*innen, bzw. der unterschiedlichen und je nach Bewertung zum Teil geringen Fallzahlen lediglich ein deskriptiver Überblick gegeben wird, der aber eine vorsichtige Einschätzung erster Tendenzen ermöglicht.

In der überwiegenden Mehrheit von bis zu zwei Drittel der Fälle folgt eine Reaktion von den Hater*innen, die den Übergriff bestätigt oder verstärkt. Auffallend ist, dass diese Hate Speech Folgeaktionen bei humorvollen Counter Speech Strategien offenbar weniger übergriffig ausfallen (30,7%) als z.B. im Vergleich zu aggressiven (43,3%) oder sachlichen (45,6%) Counter Speech Strategien. Der Anteil an Hate-Speech Folgeaktionen mit weiteren Übergriffen ist aber überall höher als der Anteil jener, wo Hate Speech ohne weitere Übergriffe verstärkt wird. Positive Folgeaktionen, wie einsichtiges oder sich von Hate Speech distanzierendes Verhalten oder sogar eine Entschuldigung kommen erwartungsgemäß selten vor und sind am ehesten bei solchen Counter Speech Strategien zu finden, die auch eine entsprechend positive Gesprächsatmosphäre zulassen, wie z.B. Humor (4,9%), Verständnis (6,7%) und Deeskalation (5,9%).

Tabelle 18

Counter Speech Strategien und Hate Speech Folgeaktionen (durchschnittliche Bewertungen der Kodierer*innen; Mehrfachantworten)				
	keine Reaktion von Hater*in	HS verstärkt / bestätigt - ohne Übergriff	HS verstärkt / bestätigt - mit Übergriff	Entschuldigung, Einsicht, Distanzierung
Humor (K1n=41, K2n=25, K3n=33, K4n=25)	39,5%	25,0%	30,7%	4,9%
ins Gewissen reden (K1n=100, K2n=5, K3n=11, K4n=16)	34,1%	18,2%	45,5%	2,3%
Verständnis (K1n=20, K2n=1, K3n=7, K4n=2)	40,0%	10,0%	43,3%	6,7%
aggressiv (K1n=110, K2n=61, K3n=108, K4n=72)	34,5%	20,5%	43,3%	1,7%
sachlich (K1n=123, K2n=92, K3n=102, K4n=49)	30,9%	20,8%	45,6%	2,8%
Deeskalation (K1n=27, K2n=17, K3n=16, K4n=8)	27,9%	26,5%	39,7%	5,9%

Um zu prüfen, ob sich die inhaltliche Ausrichtung der Counter Speech Strategien im Interaktionsverlauf ändert, wurden auch die folgenden bzw. zweiten Counter Speech Strategien entsprechenden Kategorien zugeordnet.

Tabelle 19

Kategorisierung zweiter Counter Speech Strategien (n = CS identifiziert; Mehrfachantworten)					
	Kodierer*in 1 (n=122)	Kodierer*in 2 (n=79)	Kodierer*in 3 (n=97)	Kodierer*in 4 (n=48)	Mittelwert
Humor / Sarkasmus / Ironie	21,3%	19,0%	13,4%	25,0%	19,7%
ins Gewissen reden / Mora	44,3%	5,1%	7,2%	8,3%	16,2%
Verständnis für Hater*in	7,4%	5,1%	2,1%	0,0%	3,7%
aggressiv / feindlich / Gegenattacke	38,5%	34,2%	40,2%	52,1%	41,3%
sachlich / Fakten / neutral nachfragen	50,8%	65,8%	54,6%	43,8%	53,8%
Deeskalation / auffordern aufzuhören	4,9%	6,3%	4,1%	0,0%	3,8%

Im Vergleich zur Kategorisierung der ersten Counter Speech Strategien zeigen sich hier kaum Veränderungen, außer dass der Anteil an aggressiven Counter Speech Strategien (erstes CS, Tabelle 15 48,5% ; zweites CS, Tabelle 19 41,3%) etwas rückläufig ist zugunsten sachlicher Counter Speech Strategien (erstes CS, Tabelle 15 49,0% ; zweites CS, Tabelle 19 53,8%). Dies könnte dadurch erklärt werden, dass bei längeren Interaktionsketten oft auch auf sachlicherer Ebene diskutiert wird, während aggressive Reaktionen eher nur in kurzen Interaktionsfolgen zu finden sind.

Tabelle 20

Zweite CS Strategien als Folge von HS MIT weiterem Übergriff (n = HS schwer; Mehrfachantworten)					
	Kodierer*in 1 (n=90)	Kodierer*in 2 (n=42)	Kodierer*in 3 (n=75)	Kodierer*in 4 (n=25)	Mittelwert
kein Counter Speech	25,0%	28,8%	27,2%	26,5%	26,9%
Humor / Sarkasmus / Ironie	20,0%	16,7%	10,7%	24,0%	17,9%
ins Gewissen reden / Mora	47,8%	7,1%	2,7%	12,0%	17,4%
Verständnis für Hater*in	6,7%	4,8%	2,7%	0,0%	3,6%
aggressiv / feindlich / Gegenattacke	38,9%	35,7%	40,0%	60,0%	43,7%
sachlich / Fakten / neutral nachfragen	56,7%	59,5%	57,3%	36,0%	52,4%
Deeskalation / auffordern aufzuhören	4,4%	9,5%	4,0%	0,0%	4,5%

Auch hier zeigen sich kaum Veränderungen in den Bewertungen im Vergleich zum Vorgehen bei der ersten Counter Speech Strategie gegen schwere Hate Speech Kommentare. Auffallend ist lediglich, dass der Anteil an inaktiven Counter Speaker*innen etwas geringer ausfällt (erstes CS, Tabelle 16 38,0% ; zweites CS, Tabelle 20 26,9%) und auch hier wieder der Anteil an aggressiven Counter Speech Strategien rückläufig ist (erstes CS, Tabelle 16 49,4% ; zweites CS Tabelle 20 43,7%). Es scheint so, dass Counter Speaker*innen, die bereit sind, sich auf längere Interaktionen einzulassen, eher sachlichere Strategien (52,4%) bevorzugen.

Tabelle 21

Zweite CS Strategien als Folge von HS OHNE weiterem Übergriff (n = HS leicht; Mehrfachantworten)					
	Kodierer*in 1 (n=25)	Kodierer*in 2 (n=34)	Kodierer*in 3 (n=18)	Kodierer*in 4 (n=23)	Mittelwert
kein Counter Speech	26,5%	34,6%	41,9%	41,0%	36,0%
Humor / Sarkasmus / Ironie	28,0%	23,5%	22,2%	26,1%	25,0%
ins Gewissen reden / Mora	40,0%	2,9%	27,8%	4,3%	18,8%
Verständnis für Hater*in	8,0%	5,9%	0,0%	0,0%	3,5%
aggressiv / feindlich / Gegenattacke	40,0%	29,4%	44,4%	43,5%	39,3%
sachlich / Fakten / neutral nachfragen	36,0%	76,5%	44,4%	52,2%	52,3%
Deeskalation / auf-fordern aufzuhören	4,0%	2,9%	5,6%	0,0%	3,1%

Eine solche Veränderung im Anteil als aggressiv bewerteten Counter Speech Strategien zeigt sich noch deutlicher beim Vorgehen gegen leichtere Formen von Hate Speech: Im Vergleich zu den ersten eingesetzten Counter Speech Strategien ist nun der Anteil an als aggressiv eingestuftem Counter Speech Strategien deutlich gesunken (erstes CS, Tabelle 17 51,5%; zweites CS, Tabelle 21 39,3%) und der Anteil der als sachlich kodierten Counter Speech Strategien angestiegen (erstes CS Tabelle 17 44,7%; zweites CS, Tabelle 21 52,3%).

Die nun nachfolgende Tabelle gibt erste Hinweise darauf, inwieweit die inhaltliche Ausrichtung der Counter Speech Strategien konsistent geblieben ist, wobei die Ergebnisse aufgrund der teilweise sehr geringen Fallzahlen und der Differenzen in der Kategorisierung wieder nur vorsichtige Tendenzen aufzeigen.

Tabelle 22

		Zweite Counter Speech Strategie (Mehrfachantworten)						
		Nichts	Humor	ins Gewissen reden	Verständnis	aggressiv	sachlich	Deeskalation
Erste Counter Speech Strategie	Humor (K1n=16, K2n=12, K3n=12, K4n=8)	35,6%	34,4%	16,7%	10,9%	48,5%	46,9%	2,1%
	ins Gewissen reden (K1n=51, K2n=1, K3n=3, K4n=5)	44,0%	13,7%	22,3%	2,0%	46,3%	22,3%	0,0%
	Verständnis (K1n=12, K2n=1, K3n=2, K4n=0)	14,4%	2,1%	8,3%	14,6%	31,3%	41,7%	12,5%
	aggressiv (K1n=53, K2n=32, K3n=52, K4n=28)	28,0%	20,8%	13,5%	2,4%	57,5%	41,6%	4,4%
	sachlich (K1n=63, K2n=49, K3n=52, K4n=22)	26,7%	19,8%	14,2%	3,7%	32,6%	66,5%	3,7%
	Deeskalation (K1n=12, K2n=17, K3n=16, K4n=1)	48,9%	7,0%	19,5%	2,8%	39,6%	52,8%	9,0%

Auffällig ist, dass es bei der inhaltlichen Ausrichtung der Strategien teilweise große Inkonsistenzen gibt, nur bei aggressiven und sachlichen Counter Speech Strategien bleibt die inhaltliche Ausrichtung relativ konsistent: Das heißt, 57,5% der Counter Speaker*innen, die bereits beim ersten Counter Speech Kommentar aggressiv reagiert haben, behalten ein solches Vorgehen bei. Gleichzeitig bleiben rund zwei Drittel (66,5%) jener, die von Beginn an sachlich als Counter Speaker*innen aufgetreten sind, auch beim zweiten Counter Speech Kommentar bei dieser Strategie. Bei allen anderen Strategien scheint es eine Verlagerung zu geben, entweder in Richtung aggressiv oder sachlich, oder aber es wird entschieden, keine weiteren Counter Speech Kommentare mehr zu posten. Eine solche Entwicklung zeigt sich deutlich auch bei jenen, die im ersten Counter Speech Kommentar humorvoll reagiert haben: Hier reagiert nun knapp die Hälfte beim zweiten Counter Speech Kommentar aggressiv (48,5%) bzw. sachlich (46,9%), mehr als ein Drittel (35,6%) postet keinen weiteren Counter Speech Kommentar und lediglich ein weiteres Drittel (34,4%) behält die humorvolle Strategie bei.

Die Ergebnisse verweisen darauf, dass aggressive und sachliche Counter Speech Strategien letztlich längerfristig bzw. mehrfach umsetzbar sind, während andere Strategien offenbar eher für kurzfristige Effekte eingesetzt werden. Dies zeigt sich auch in der Bewertung des letzten Counter Speech Kommentars in längeren Interaktionsketten (Tabelle 23): Auch hier überwiegt in der mittleren Bewertung der Anteil an sachlichen Counter Speech Strategien (51,4%), gefolgt von aggressiven Interventionsformen (33,4%).

Tabelle 23

Bewertung des letzten Counter Speech Kommentars (n=Interaktionsketten länger als HS-CS-HS-CS)					
	Kodierer*in 1 (n=71)	Kodierer*in 2 (n=57)	Kodierer*in 3 (n=52)	Kodierer*in 4 (n=19)	Mittelwert
Humor	15,5%	15,8%	5,8%	10,5%	11,9%
ins Gewissen reden	36,6%	3,5%	1,9%	26,3%	17,1%
Verständnis	8,5%	3,5%	1,9%	0,0%	3,5%
aggressiv	35,2%	21,1%	40,4%	36,8%	33,4%
sachlich	40,8%	68,4%	59,6%	36,8%	51,4%
Deeskalation	9,9%	3,5%	3,8%	0,0%	4,3%

2.2.7 Erfolg von Counter Speech

Auch bei der Bewertung darüber, ob Counter Speech im analysierten Kommentarblock letztlich als erfolgreich oder nicht erfolgreich bewertet wird, gib es zwischen den Kodierer*innen große Differenzen. Die Bewertung wurde hier bewusst offen gestaltet, das heißt die Kodierer*innen mussten angeben, ob es im gesamten Kommentarblock Hinweise auf Erfolg von Counter Speech gibt und in einem freien Textfeld darstellen, wie sich dies bemerkbar macht.

Tabelle 24

Bewertung des Erfolgs von Counter Speech (n=37; einstimmige Markierung des ersten CS Kommentars)	
Übereinstimmung in der Erfolgsbewertung	32,4%
Keine Übereinstimmung in der Erfolgsbewertung	67,6%

Selbst bei Übereinstimmung über die Markierung des ersten Counter Speech Kommentars sind sich die Kodierer*innen nur in einem Drittel der Fälle (32,4%, 12 Fälle) darüber einig, ob Counter Speech als erfolgreich zu bewerten ist. Von diesen 12 Fällen bewerteten die Kodierer*innen 11 Fälle einstimmig als nicht erfolgreich. In nur einem einzigen Fall waren sie sich einig, dass Counter Speech erfolgreich war.

Die Kodierer*innen wurden gebeten, in einem offenen Textfeld darzustellen, warum Counter Speech in ihrer Bewertung als erfolgreich oder nicht erfolgreich beurteilt wird. Die folgenden Tabellen 25 und 26 fassen die Analyseergebnisse zusammen. Die Kodierer*innen orientierten sich stark an formalen Charakteristiken für die Beurteilung des Erfolgs oder Nicht-Erfolgs von Counter Speech: Konkret wird vor allem das Ende von Interaktionsfolgen betrachtet, das heißt die Zuordnung, ob das letzte Element in der Interaktionskette als Hate oder Counter Speech kategorisiert werden kann. Jene Person, die das letzte Wort hat, wird als erfolgreicher eingestuft. Insgesamt wird dies am häufigsten als Begründung für die Kodierung genannt. Als zweites formales Merkmal wird das wahrgenommene Verhältnis des Ausmaßes von Hate Speech und Counter Speech herangezogen: Wird in einem Kommentarblock insgesamt mehr

Counter Speech als Hate Speech wahrgenommen, wird dies eher als Erfolg gewertet – vor allem dann, wenn die Interaktionsfolge auch mit Counter Speech endet. Umkehrt wird eine Überzahl an Hate Speech als Misserfolg gewertet, auch dann, wenn zum Beispiel einzelne Counter Speaker*innen von mehrer Hater*innen attackiert werden.

Neben diesen formalen Charakteristiken werden auch inhaltliche Dimensionen in Betracht gezogen: Als erfolgreich wird Counter Speech dann beurteilt, wenn sich die Tonalität insgesamt zum Positiven wendet bzw. Counter Speaker*innen in der Tonalität positiv bleiben (und Hater*innen in ihrer Aggressivität nachlassen). Entsprechend nicht erfolgreich ist Counter Speech, wenn es nicht gelingt, die Tonalität bzw. die Übergriffe der Hater*innen in eine positivere Richtung zu lenken oder wenn die Stimmung insgesamt sogar noch negativer wird. Als weiteres inhaltliches Merkmal für erfolgreiche Counter Speech werden Signale der Einsicht oder Entschuldigungen von Hater*innen angeführt, wobei solche relativ selten vorzufinden sind. Im Gegensatz dazu wird Counter Speech als nicht erfolgreich bewertet, wenn Counter Speaker*innen im Vergleich zu Hater*innen inhaltlich schwach erscheinen oder der Eindruck entsteht, dass diese von Hater*innen nicht ernst genommen werden. Aber auch, wenn diese im Verhältnis zur wahrgenommenen Schwere der getätigten Übergriffe nur zögerlich oder gar nicht reagieren.

Tabelle 25

Erfolg	kein Erfolg
<p>Nach Counter Speech folgt kein Hate Speech <u>Begründungs-Beispiele:</u></p> <ul style="list-style-type: none"> nach CS kein HS mehr kein weiterer HS, CS geht nicht auf HS Beschimpfungen ein 	<p>Nach Hate Speech folgt kein Counter Speech <u>Begründungs-Beispiele:</u></p> <ul style="list-style-type: none"> kein CS nach HS (betrifft 60% aller Begründungen) nach weiterem HS keine Reaktion von CS keine Reaktion, obwohl es ein schwerer Übergriff ist
<p>Es gibt insgesamt (viel) mehr Counter Speech als Hate Speech <u>Begründungs-Beispiele:</u></p> <ul style="list-style-type: none"> nach viel mehr CS als HS keine Antwort von einzelner HS mehr viel CS folgt auf HS - keine Reaktion von HS kein HS mehr nach dem viele CS kommentiert haben 	<p>Mehrere Hate-Kommentare gegen wenige CS-Kommentare <u>Begründungs-Beispiele:</u></p> <ul style="list-style-type: none"> viele unterschiedliche User*innen kommentieren, aber nur eine Person, die sich gegen den Ursprungskommentar stellt CS wird mehrfach attackiert
<p>Tonalität von Counter Speaker*innen bleibt positiv, Stimmung wird im Allgemeinen positiver <u>Begründungs-Beispiele:</u></p> <ul style="list-style-type: none"> Die CS bleiben alle ruhig und sachlich, niemand unterstützt den*die Hater*in, wodurch er*sie nach einiger Zeit auch nicht mehr zurück schreibt Der letzte Kommentar ist ein sachlicher und freundlicher Counter, auf den dann nichts mehr folgt 	<p>Allgemein negative Tonalität, Hater*in bleibt oder wird noch negativer <u>Begründungs-Beispiele:</u></p> <ul style="list-style-type: none"> Beschimpfungen werden trotz CS weitergeführt HS antwortet zwar nicht mehr, aber die Stimmung ist im Verlauf aggressiver geworden die CS sind immer sachlich und nicht übergriffig, aber HS bleiben gleich und verändern ihre Umgangsform nicht
<p>Hater*innen zeigen Einsicht, distanzieren / entschuldigen sich. <u>Begründungs-Beispiele:</u></p> <ul style="list-style-type: none"> Originale HS nimmt sich zurück, distanziert sich von den restlichen HS-Kommentaren und weist diese in Schranken. Hater*in bedankt sich schlussendlich für das Gespräch und rudert freundlich zurück: <i>"HighLifeGermany 1 year ago @Isi Danke fürs Gespräch ich werde versuchen es zu verstehen :D dir noch alles gute"</i> HS kann CS ernstnehmen, CS bedankt sich 	<p>Schwache Counter Speaker*innen und starke Hater*innen <u>Begründungs-Beispiele:</u></p> <ul style="list-style-type: none"> HS nimmt CS nicht ernst zu wenig Gegenreaktion bei problematischen Aussagen (z.B. Opfer von Vergewaltigungen seien selbst schuld) insgesamt sehr viele Kommentare aber Hater*innen sind stärker und bleiben länger dran CS nimmt sich eher zurück, HS postet großzügig islamophobes Gedankengut

Besonders unklar ist die Bewertung über den Erfolg oder Nicht-Erfolg (Tabelle 26) von Counter Speech dann, wenn Konflikte nicht gelöst, sondern eher durch Ablenkung „versanden“ oder insgesamt kein klares Ende der Interaktionen erkennbar ist. Ebenso war die Bewertung nicht eindeutig, wenn sich Hate Speech auf andere Themen verlagert oder Counter Speaker*innen ihre Rollen wechseln und in späterer Folge selbst zu Hater*innen werden.

Tabelle 26

Unklar, ob erfolgreich oder nicht
<p>Auseinandersetzungen „versanden“; Ablenkung und keine Lösung des Konflikts <u>Begründungs-Beispiele:</u></p> <ul style="list-style-type: none"> • Die Diskussion versandet sehr schnell dadurch, dass dann viele User*innen ihre persönliche Leidensgeschichte erzählen. • Es wird nicht direkt auf den Hate reagiert, aber es wird das Thema gewechselt wodurch direktes Reagieren vielleicht weniger notwendig ist.
<p>Verlagerung von Hate Speech <u>Begründungs-Beispiel:</u></p> <ul style="list-style-type: none"> • HS nimmt sich zurück und gibt zu, dass der Kommentar unwahr ist, jedoch begeht Hater*in dann einen Übergriff gegen andere
<p>Rollenwechsel: Counter Speaker*innen werden zu Hater*innen <u>Begründungs-Beispiele:</u></p> <ul style="list-style-type: none"> • Es gibt in der Diskussion einen Shift und Counterspeaker*innen werden selbst mit neuen Themen zu Hater*innen
<p>Kein erkennbares Ende <u>Begründungs-Beispiele:</u></p> <ul style="list-style-type: none"> • Endloses Hin- und Herdiskutierenn • sinnloses Einanderbeleidigen

Die Ergebnisse zeigen insgesamt, wie wichtig ein hohes Ausmaß und Sichtbarkeit von Counter Speech ist, um diese als erfolgreich zu bewerten. Gleichzeitig spielt auch eine positive Tonalität eine entscheidende Rolle. Um das Potenzial von Counter Speech zu fördern, wäre es wichtig, Jugendliche dazu zu mobilisieren, möglichst häufig als Counter Speaker*innen mit inhaltlich positiver Ausrichtung aufzutreten, auch am Ende von längeren Interaktionsketten, um letztlich einen „erfolgreichen“ Abschluss zu signalisieren.

2.3 Zusammenfassung und Schlussfolgerungen

Ziel der Inhaltsanalyse von User*innen-Kommentaren auf YouTube war es, reales und öffentlich sichtbares Counter Speech Verhalten Jugendlicher zu untersuchen. Dabei mussten zum einen Herausforderungen in Zusammenhang mit der Datensammlung (z.B. Probleme mit lückenhaften Daten aufgrund automatisierter Filter durch YouTube; Auswahl entsprechender Kanäle/ Themen mit Zielgruppe Jugendliche) gelöst werden. Zum anderen musste die Fülle an Datenmaterial (Download von 32.852 Kommentarblöcken von 517 Videos mittels Scraper) auf Basis eines eigens entwickelten Wörterbuchs (inklusive jugendtypischer und im Netzjargon gebräuchlicher Schimpfwörter und gewalthaltiger Begriffe; Berücksichtigung unterschiedlicher Schreibweisen) selektiert werden. Für die Datenanalyse wurde weiters ein umfassendes Codebook ausgearbeitet, um Hate und Counter Speech Interaktionen zu identifizieren und abzubilden sowie die Bandbreite an jugendtypischen Counter Speech Strategien und deren Konsequenzen zu dokumentieren. Im Zuge der notwendigen Validierung des Codebooks inklusive entsprechender Schulung der Kodierer*innen zeigten sich große Probleme mit der Intercoder-Reliabilität, die auch nach mehrfachen Überarbeitungsschritten unterdurchschnittliche und damit nicht vertretbare Werte aufwies, was auf große Uneinigkeiten in den Bewertungen der einzelnen Kodierer*innen zurückzuführen ist. Allein für die Frage, ob Hate Speech vorliegt oder nicht, war der Wert der Intercoder-Reliabilität stets deutlich unter dem Grenzwert von 0.68.

Probleme bei der Identifizierung von Hate Speech sind prinzipiell nicht neu und sind insbesondere bei automatisierten Hate Speech Detection Tools oft sehr problematisch. Überraschend war aber das Ausmaß der Abweichungen zwischen den Kodierer*innen, trotz mehrfacher Schulungen und detaillierter Codebook-Definitionen mit Fallbeispielen. Diese Tatsache veranlasste das Forschungsteam, diese Bewertungsunterschiede in den Analysefokus zu rücken. Die Ergebnisse sind nicht nur für automatisierte Hate Speech Detection von Relevanz, sondern auch für jugendliche Counter Speaker*innen, welche ebenso Einschätzungsschwierigkeiten haben, ob es sich um Hate Speech handelt oder nicht bzw. ob folglich Counter Speech gefragt ist oder nicht.

Die Ergebnisse verdeutlichen erwartungsgemäß die großen Differenzen zwischen den Kodierer*innen in den Bewertungen, ob Hate Speech bzw. Counter Speech vorkommt bzw. wie sich die Interaktionen gestalten. Die Analysen zeigen, dass Kommentare nur dann relativ einheitlich als Hate Speech identifiziert wurden, wenn diese durch z.B. in der Gesellschaft gängige Beschimpfungen oder Kraftausdrücke auch klar als solche deklariert werden und daher auch leicht als solche einzuordnen sind. In der Regel sind das relativ kurz gehaltene, negative Kommentare. Deutlich schwieriger wird die Zuordnung, wenn mittels Hate Speech Diskriminierungsformen wie Transfeindlichkeit, Homophobie, Rassismus oder Sexismus reproduziert werden. Dies ist zum einen auf die unterschiedliche Bewertung von Diskriminierungsformen aufgrund individueller Rahmenbedingungen und Einstellungen zurückzuführen, zum anderen sind

bestimmte Diskriminierungsformen, wie zum Beispiel Transfeindlichkeit, noch nicht klar im gesellschaftlichen Wertegefüge eingeordnet, was besonders dann schwierig wird, wenn signalisierende Schimpfwörter als klare „Markierung“ fehlen und der Eindruck entsteht, dass prinzipiell „nur diskutiert“ wird. Die Reproduktion von -ismen und negativem Gedankengut stellt sich zudem häufig in komplexen Kommunikationsverläufen dar, sodass für deren Wahrnehmung und Bewertung eine entsprechende Sensibilisierung notwendig ist.

Die großen Unterschiede in der Identifizierung von Hate Speech bedingen folglich auch Unterschiede in der Identifikation von Counter Speech, da diese in der Regel nur im gemeinsamen Kontext mit Hate Speech erkennbar ist: Entsprechende Erkennungsmerkmale sind dann etwa das Taggen von Hater*innen (d.h. User*innenname des Haters bzw. der Haterin steht am Anfang des Counter Speech Kommentars), unmittelbare Folgekommentare, aber auch das Aufgreifen der inhaltlichen Ausrichtung des Hate Speech Kommentars (z.B. gleiche Wortwahl, gleiches Schimpfwort) oder auch die klare Benennung als Hate Speech. Etwa die Hälfte der analysierten Counter Speech Kommentare waren in Form von (wenig wünschenswerten) Gegenattacken formuliert, d.h. eine Zuordnung als Counter Speech wäre ohne den entsprechenden Hate Speech Kommentar nicht möglich gewesen.

Unklarheiten gab es folglich auch bei der Rollenzuordnung der Personen, wer Opfer, Täter*in oder Counter Speaker*in ist bzw. inwieweit Counter Speech durch das Opfer selbst oder durch dritte Personen erfolgt, wobei die Ergebnisse prinzipiell darauf hindeuten, dass Counter Speaker*innen mehrheitlich Personen sind, die nicht gleichzeitig auch Opfer sind.

Auch die Interaktionsverläufe zwischen Hater*innen und Counter Speaker*innen werden durch die Kodierer*innen unterschiedlich wahrgenommen und bewertet, wobei in der durchschnittlichen Bewertung in etwa zwei Drittel jener Fälle, wo Hate Speech identifiziert wurde, auch Counter Speech wahrgenommen wurde – allerdings sehr oft in Form ebenso hasserfüllter Gegenattacken. In der Regel verlaufen die Interaktionen zwischen Hater*innen und Counter Speaker*innen eher kürzer. Längere Auseinandersetzungen finden sich in durchschnittlich etwas mehr als 30% der Fälle, wobei diese überwiegend mit einem abschließendem Hate Kommentar enden bzw. wird Hate Speech im Laufe der Interaktionen verstärkt oder zumindest bestätigt wird. Positive Entwicklungen werden nur selten identifiziert.

Besonders große Uneinigkeiten gab es auch hinsichtlich der Bewertung der Schwere von Hate Speech. Durchschnittlich waren sich die Kodierer*innen nur in einem Fünftel der Fälle über die Bewertung der Schwere einig, wobei insgesamt mehr leichte als schwere Übergriffe identifiziert wurden. Was als leicht oder schwer eingestuft wird, wird an der Häufigkeit negativer Inhalte innerhalb eines Kommentars (z.B. Mehrfachbeleidigungen), an der Intensität der eingesetzten Beschimpfungen (z.B. Demütigung, Tod wünschen) bzw. an klar als negativ deklariertem Gedankengut (z.B. nationalsozialistisch) festgemacht.

Die inhaltliche Ausrichtung von Counter Speech wurde ebenfalls unterschiedlich bewertet. Durchschnittlich wurde aber sowohl bei leichter als auch bei schwerer Hate Speech von den Kodierer*innen vor allem aggressive, oder aber auch sachliche Strategien wahrgenommen, wobei leichteren Formen von Hate Speech in der durchschnittlichen Bewertung noch etwas häufiger mit aggressiver Counter Speech beantwortet wurden, möglicherweise aufgrund der geringeren Hemmschwelle („harmlose/r Hater*in). Andere inhaltliche Ausrichtungen wurden verhältnismäßig selten gefunden, am ehesten noch humorvolle/ sarkastische Reaktionen, wobei humorvolle Strategien hier wieder eher bei leichten Übergriffen identifiziert wurden. Sowohl bei leichter als auch bei schwerer Hate Speech wurde durchschnittlich in deutlich mehr als einem Drittel der Fälle nicht reagiert.

In durchschnittlich zwei Drittel der Fälle wird Counter Speech wieder mit Hate Speech beantwortet, wobei diese Hate Speech Folgereaktionen bei humorvollen Counter Speech Strategien von den Kodierer*innen tendenziell etwas weniger übergriffig wahrgenommen wurden als z.B. im Vergleich zu aggressiven oder sachlichen Counter Speech Strategien. Positive Entwicklungen fanden sich am ehesten bei solchen Counter Speech Strategien, die auch eine entsprechend positive Gesprächsatmosphäre zulassen, wie z.B. Humor, Verständnis, oder deeskalierende Strategien, der Anteil an positiven Entwicklungen ist aber insgesamt sehr gering (durchschnittlich bis max. knapp 7%).

Die Analyse der Interaktionsverläufe zeigt weiteres, dass bei längeren Interaktionen eher versucht wird, Counter Speech auf sachlicher Ebene zu betreiben, während sich aggressive Formen eher in kürzeren Interaktionen finden. Bei aggressiven und sachlichen Counter Speech Strategien wird diese inhaltliche Ausrichtung auch relativ konsistent im Interaktionsverlauf beibehalten, bei anderen Strategien scheint sich das aber im Laufe der Zeit entweder in Richtung aggressiv oder sachlich zu verlagern, oder aber es wird entschieden, nicht mehr darauf zu reagieren.

Schließlich wurde auch die Bewertung des Erfolgs von Counter Speech von den Kodierer*innen sehr uneinheitlich wahrgenommen, wobei sich diese überwiegend darüber einig waren, dass die jeweilige Counter Speech in den kodierten Kommentarblöcken nicht erfolgreich war. In der Erfolgsbewertung orientierten sich diese vor allem stark an formalen Kriterien, wie etwa, ob der letzte Kommentar der Interaktionskette ein Hate oder Counter Speech Kommentar war oder ob im Verhältnis der Anteil an Counter Speech höher war als jener von Hate Speech. Positive Signale wie etwa die Änderungen in der Tonalität oder einsichtiges Verhalten wurden ebenfalls für die Erfolgsbewertung herangezogen, kamen aber nur selten vor. Erfolgreiche Counter Speech wurde zum Teil auch mit einem schwachen oder unzureichenden Auftreten von Counter Speaker*innen argumentiert. Erfolglosigkeit ergab sich aber auch, wenn Konflikte „versandeten“ oder sich verlagerten, anstatt gelöst zu werden.

Die gewonnenen Erkenntnisse sind auch in Zusammenhang mit der Identifizierung von Hate Speech äußerst gewinnbringend: Bisher verfügbare Tools berücksichtigen in der Regel nicht explizit die

sprachlichen Besonderheiten der Jugendsprache. Das im Projekt genierte Wörterbuch könnte daher eine wesentliche Bereicherung für computergestützte Tools sein. Zudem konzentrieren sich die meisten Untersuchungen auf Twitter (auch aufgrund einfacherer Sampling-Strategien mit Hilfe verfügbarer Hashtags). Andere Plattformen wie YouTube werden deutlich seltener in Hinblick auf Hate Speech untersucht.

Insgesamt war das Forschungsteam – wie auch schon bei den Online Rollenspielen – stark mit den **sprachlichen Herausforderungen im Datenmaterial** konfrontiert. Gerade im Internet werden jugendsprachliche Ausdrucksweisen mit dem spezifischen Netzjargon (z.B. Gamersprache) vermischt und durch Emoticons, Emojis, Emotes, Internetmemes, Hashtags, Abkürzung vernetzt. Das macht es für erwachsene Forscher*innen oft unmöglich, Textpassagen zu interpretieren. Dies führte dazu, einerseits jüngere Mitglieder in das Forschungsteam zu integrieren und andererseits den Kontakt zu Jugendlichen als Expert*innen ihrer Lebenswelt zu intensivieren und damit den notwendigen Austausch zwischen Forscher*innen und Jugendlichen auf gleicher Ebene in den Mittelpunkt zu rücken.

Die Ergebnisse zeigen, dass es für eine Mobilisierung zu Counter Speech auch einer hohen Sensibilisierung in der Wahrnehmung von Hate Speech bedarf. Gleichzeitig wird deutlich, wie wichtig die Erhöhung der Sichtbarkeit von Counter Speech (möglichst abseits von aggressiven Gegenattacken) ist, um diese als erfolgreich zu bewerten. Neben einer notwendigen Sensibilisierung, auch in Hinblick auf die Bewertung der Schwere von Hate Speech, insbesondere bei weniger offensichtlichen Formen der Diskriminierung, wäre es also wichtig, Jugendliche dazu zu mobilisieren, möglichst häufig als Counter Speaker*innen mit inhaltlich positiver Ausrichtung aufzutreten. Dabei könnten auch bestehende Counter Speech Kommentare unterstützt werden oder gut gemeinte, aber aggressiv formulierte Gegenattacken relativiert werden. Hilfreich wäre auch, Hate Speech klar als solche zu benennen und damit für andere erkenntlich zu machen, oder aber auch Interaktionsketten bewusst mit positiven Kommentaren abzuschließen. Solche Signale des Erfolgs könnten die Bewertung der Sinnhaftigkeit von Counter Speech insgesamt verbessern.

3. Online Aktionsforschung zur Erarbeitung von Counter Speech Vorschlägen

Eine weitere methodisch Zielsetzung des Projekts war es, aufbauend auf den gewonnenen Erkenntnissen (Online Rollenspiele, Inhaltsanalyse) in enger Zusammenarbeit mit den Praxispartner*innen jugendgerechte Counter Speech Vorschläge für den praxisrelevanten Einsatz auszuarbeiten. Um eine jugendtypische Darstellungsweise zu gewährleisten, wurden drei workshopartige Arbeitsgruppen angedacht, um konkrete, beispielhafte Text- und/oder Bildsequenzen umzusetzen. Durch die Lukrierung einer Zusatzfinanzierung der Stadt Wien im Call „Digitaler Humanismus“ (Projekttitle: MAKING Cyber Heroes, Laufzeit Jänner 2020 bis September 2021) konnte diese Phase deutlich intensiviert, wodurch Jugendliche selbst viel stärker in die Entwicklung von jugendgerechten Counter Speech Strategien miteinbezogen werden konnten. Die geplanten Workshops konnten zu einer umfassenden, partizipativen Digitalen Aktionsforschung ausgebaut werden, indem Forscher*innen, Projektpartner*innen gemeinsam mit Jugendlichen auf gleicher Augenhöhe zusammengearbeitet haben. Zusätzlich wurden dafür auch Workshops mit Zivilcourage-Trainer*innen von Saferinternet.at, Mauthausen Komitee Österreich und ZARA durchgeführt, die durch ihre professionelle Vermittlungsarbeit an Jugendlichen wertvolle Beiträge leisteten, die ebenfalls in die Durchführung der Aktionsforschung miteinfließen. Der vorliegende Berichtsteil entspricht einer komprimierten Fassung des „MAKING Cyber Heroes“-Endberichts für die Stadt Wien.

Counter Speech ist auf vielfache Weise gestaltbar und umfasst jegliche Form digitaler Artefakte (z.B. Videobotschaften, humoristische oder Widersprüche aufzeigende Memes / GIFs, faktenbasierte Gegendarstellungen, aktive Gegenrede durch Einzelpersonen). Gerade für Jugendliche bieten die vielfältigen medialen Möglichkeiten in sozialen Netzwerken grundsätzlich hohes Potenzial, nicht nur die Effizienz von Online Zivilcourage zu stärken, sondern insgesamt die Attraktivität von Online Zivilcourage für Jugendliche zu erhöhen. Bisher liegen allerdings keine empirischen Erkenntnisse darüber vor, inwieweit es Jugendlichen auch gelingt, dieses Potenzial nutzbar zu machen. Im Zuge der Online Aktionsforschung wurde gemeinsam erarbeitet, (1) wie Counter Speech durch Jugendliche unter Berücksichtigung vorhandener Kompetenzen in und mit digitalen Medien gestaltet werden kann, (2) wie dabei die medialen bzw. plattformspezifischen Möglichkeiten nutzbar gemacht werden können und (3) inwieweit es Jugendlichen gelingen kann, erfolgreiche Interventionsstrategien umzusetzen. Ziel der Studie war eine gemeinsame, aktive Auseinandersetzung mit (praktizierbarer) Online Zivilcourage, indem jugendgerechte und praxistaugliche Online-Interventionen im Sinne zivilcouragierter Gegenrede entwickelt und gleichzeitig die Anwendbarkeit verschiedener Online Zivilcourage Gestaltungsmöglichkeiten im Umgang mit Hate Speech und anderer negativer Rhetorik erforscht wurden.

Aus den Erkenntnissen werden praxisrelevante Empfehlungen abgeleitet, um Counter Speech durch Jugendliche nachhaltig zu fördern.

3.1 Methodisches Design

Um die jugendgerechte Umsetzung von Counter Speech und die Sichtweisen der Jugendlichen selbst zu berücksichtigen, wurde eine partizipative Online-Aktionsforschung mittels Zoom umgesetzt und die teilnehmenden Jugendlichen aktiv in den Forschungsprozess miteinbezogen. Gearbeitet wurde mit **vier Gruppen von je 5 bis 11 14- bis 22-jährigen Jugendlichen** (insgesamt 31 Teilnehmer*innen, davon 17 weiblich, 14 männlich, großteils aus Wien, vereinzelt aus Bundesländern) mit unterschiedlichem Bildungshintergrund. Pro Gruppe gab es vier Treffen, d.h. insgesamt wurden 16 digitale Aktionsforschungsrunden durchgeführt (Untersuchungszeitraum Februar 2020 bis Mai 2021). Zwischen den einzelnen Treffen erhielten die Jugendlichen kurze Arbeitsaufträge (z.B. Sammeln von Screenshots mit Online-Übergriffen und Konterreaktionen).

Aktionsforschung eignet sich besonders für die inhaltliche Auseinandersetzung mit sozialen Problemen und daraus abgeleiteten Interventionen in der Praxis, indem Personen der betroffenen Zielgruppe aktiv am Forschungsprozess mitbeteiligt werden. Methoden der partizipativen Aktionsforschung haben sich gerade in der Arbeit mit Jugendlichen besonders bewährt (Wöhler et al., 2017). Die digitale Umsetzung, die vor allem durch die Corona-Pandemie initiiert wurde, eröffnete neue und attraktive Möglichkeiten, die gemeinsamen Phasen der Forschung und Aktion zu gestalten: z.B. gemeinsames Vorgehen gegen reale Hate-Kommentare inklusive Projektion der Interaktionen in Echtzeit (Bildschirm teilen Funktion, WhatsApp-Web); gleichzeitige Nutzung sprach- und textbasierter Diskussionsmöglichkeiten (z.B. Präsentation von Memes, Posten konkreter ausformulierter Kommentare im Chat, mündliche Erläuterungen); Simulation von Alltags-Situationen im Internet (z.B. spontanes Reagieren auf unerwartete Hate-Kommentare). Außerdem wurden vorbereitete Videos (z.B. YouTube Videos zum Thema Hate Speech), Powerpoint-Slides mit Counter Speech-Vorschlägen inkl. animierter GIFs präsentiert. Alle digitalen Workshops wurden aufgezeichnet und als Analysegrundlage verwendet.

3.1.2 Datenbasis und Methodik der Digitalen Aktionsforschung

Bei der Rekrutierung der Aktionsforschungs-Teilnehmer*innen wurde nach motivierten und freiwillig teilnehmenden Jugendlichen gesucht, die prinzipielles Interesse an der Thematik zeigten und bereit waren, mehrmals gemeinsam mit dem Forschungsteam zusammenzuarbeiten. Dabei wurde versucht, sowohl in geschlechts- als auch in bildungsbezogener Hinsicht ein ausgeglichenes Verhältnis zu schaffen.

Die **Zusammensetzung der einzelnen Gruppen** wird nachfolgend kurz dargestellt:

Gruppe 1: durchschnittlich 16/17-Jährige, überwiegend Mädchen, Bildung: durchmischt, eher höherer Bildungshintergrund; Nutzung aller gängigen sozialen Medien (z.B. WhatsApp, Instagram, Snapchat, TikTok,

YouTube, vereinzelt auch Facebook). Die Gruppe wurde aus dem Umfeld von Saferinternet.at rekrutiert, daher zeigten sich die Jugendlichen als besonders engagiert, Internet-affin und sensibilisiert für heikle Themen im Internet.

Gruppe 2: durchschnittlich 16/17-Jährige, überwiegend Mädchen, höhere Bildung; Nutzung aller gängigen sozialen Medien, aber nicht besonders intensiv. Die Jugendlichen waren größtenteils als Peer-Mediator*innen an ihrer Schule tätig und daher für Probleme Jugendlicher sensibilisiert.

Gruppe 3: durchschnittlich 17/18-Jährige, überwiegend Burschen, mittlere Bildung; intensive Nutzung aller gängigen sozialen Medien und verstärkt auch Netzwerke der Gaming-Szene (z.B. Twitch) oder alternative Netzwerke wie Reddit oder Twitter. Die Jugendlichen absolvierten eine Lehre in Wien mit dem Ausbildungsschwerpunkt E-Commerce, zeigten hohes Interesse an der Thematik des Projekts und waren in ihrem Internet-Alltag als aktive Nutzer*innen immer wieder damit konfrontiert.

Gruppe 4: durchschnittlich 18/19-Jährige, überwiegend Burschen, niedrige Bildung (absolvierten eine überbetriebliche Lehrlingsausbildung); sehr unterschiedliches Nutzungsspektrum sozialer Medien: aktive (z.B. eine Gamerin) bis weitgehend passive Konsumation von digitalen Inhalten (z.B. nur Musik hören); dennoch auch Nutzung gängiger sozialer Medien, wie z.B. WhatsApp, Snapchat oder Instagram wenn auch zum Teil eingeschränkter (z.B. früher auf Instagram aktiv, jetzt nicht mehr). Unter den aktiveren Nutzer*innen in der Gruppe auch Opfererfahrungen mit Hate-Kommentaren.

3.1.2 Konzeption der partizipativen, digitalen Aktionsforschung

Ziel der partizipativen Aktionsforschung war eine Auseinandersetzung mit (praktizierbarer) Online Zivilcourage gemeinsam mit Jugendlichen. Damit wurden Jugendliche einerseits zu Forschungspartner*innen, gleichzeitig waren sie selbst Teil der betroffenen Zielgruppe im Umgang mit Hate Speech bzw. von Counter Speech Maßnahmen. Jugendliche sollten in Zusammenarbeit mit dem Forschungsteam jugendgerechte und praxistaugliche Online-Interventionen im Sinne zivilcouragierter Gegenrede entwickeln und gleichzeitig die Anwendbarkeit verschiedener Online Zivilcourage Gestaltungsmöglichkeiten im Umgang mit Hate Speech und Cyberbullying erforschen.

Bezeichnend für partizipative Aktionsforschung sind wechselnde **Phasen der Forschung und Aktion**. In jedem Schritt werden die gemeinsam gewonnenen Erkenntnisse analysiert und reflektiert. Um die Zusammenarbeit mit den Jugendlichen entsprechend vorzubereiten, wurde für jedes Treffen im Rahmen der Aktionsforschung ein eigenes Workshop-Konzept erarbeitet um unterschiedliche Phasen der Forschung und Aktion anzuregen. Umgesetzt wurde z.B. die Konfrontation der Jugendlichen mit WhatsApp-Chatprotokollen aus den Online Rollenspielen mit dem Ziel, Optimierungspotenziale zu erkennen; Auseinandersetzungen mit ausgewählten YouTube und TikTok Videos zum Thema Umgang mit Hate Speech, um eine gemeinsame Analyse aus Sicht Jugendlicher hinsichtlich ihrer Wirkung auf jugendliche Peers anzuregen; Umsetzung einer „Speed Counter Speech“-Übung, um Situationen zu

simulieren, wo unmittelbare, weitgehend unvorbereitetes und schnelle Counter Speech Interventionen erforderlich ist oder aktive Vorgehen gegen echte Hate-Kommentare mit Hilfe von Fake-Profilen, um reales Auftreten als Counter-Speaker*in zu simulieren. Im Folgenden werden nun relevante Ergebnisse zusammengefasst:

3.2 Ergebnisse

3.2.1 Erfahrungen Jugendlicher mit Counter Speech im Internet-Alltag

Die Aktionsforschungs-Teilnehmer*innen haben insgesamt **wenig eigene Erfahrungen im Praktizieren von Counter Speech**, die sich in der Regel auf den Einsatz für befreundete oder bekannte Personen beschränken. Sich für fremde Personen in öffentlichen Kontexten, wie z.B. TikTok oder Instagram einzusetzen, wird sogar von manchen eher negativ konnotiert: *„Dann komm ich rüber wie so ein Fangirl 'Aahh ich muss eine Person gegen 20 Millionen kämpfen und ich werde gewinnen' (ganz hohe, piepsige Stimme)“* (G4, weiblich, min 41:40). Gleichzeitig widerspricht es vielfach den eigenen Moralvorstellungen, wenn auch fremde Personen massiven öffentlichen Attacken ausgesetzt sind.

Die **Wahrnehmung von Counter Speech durch andere Personen ist ebenfalls eher gering ausgeprägt und oft auch tendenziell negativ konnotiert**: Um auf Counter Speech in Reaktion auf Hate Speech aufmerksam zu werden, ist das Lesen von entsprechenden Kommentaren Voraussetzung – auf Plattformen wie YouTube wird das von Jugendlichen aber sehr unterschiedlich gehandhabt: Viele lesen solche Kommentare nicht. Andere, die die Kommentare lesen, sehen die Argumentationsweisen der unterschiedlichen Gegenreaktionen eher als Unterhaltung. In der Wahrnehmung der Jugendlichen ist Counter Speech als Reaktion auf Hate Speech zwar auffindbar, aber deutlicher seltener als Hate Speech zu finden. Counter Speech wird insgesamt als mühsamer und langwieriger Interaktionsprozess betrachtet, der sich oft als zu komplex und überfordernd gestaltet, um den Verlauf überblicken zu können mit wenig Aussicht auf Erfolg. Counter Speech wird dann eher als Zeitverschwendung eingeschätzt und Übergriffe ignoriert (*„Also ich les mir das durch, also nicht alles, sondern das, was mich interessiert. Aber ich glaub nicht, dass ich die dann kommentier, ich glaub, ich würd dann einfach weiterscrollen, weil es ist dann auch irgendwie Zeitverschwendung meistens“* G3, männlich, min 10:17). Manche Hasskommentare werden einfach auch als „dumm“ wahrgenommen und entsprechende Counter Speech Interventionen als unnötig eingestuft (*„Zum Beispiel einer schreibt ein Kommentar schreibt der andere 'haha wie hast du das und das geschrieben' und keine Ahnung ja so ur unnötige Sachen wirklich“* G4, weiblich, min 28:50). Hinderlich ist zudem das auch das fehlende Kontextwissen bezüglich des Opfers, da sehr oft nicht klar ist, für wen man sich eigentlich einsetzt, insbesondere wenn es um fremde Personen geht. Gerade bei fremden Personen wird Counter Speech dann eher als Einmischen denn als Intervenieren wahrgenommen. Darüber hinaus thematisieren die beteiligten Jugendlichen eine prinzipielle Angst vor Gegenreaktionen, wenn das eigene

Kommentar bzw. Posting von anderen – nicht unbedingt nur von Hater*innen - nicht gut angenommen wird („*Man wird halt bis aufs Kleinste sozusagen kritisiert. Deswegen gibt es ganz viele Leute die einfach gar nichts kommentieren, gar nichts liken*“ G4, weiblich, min 35:52).

Counter Speech ist für Jugendliche auch prinzipiell kein vordergründig wichtiges Thema: „*Ich persönlich such jetzt nicht im Internet nach Hate-Kommentaren, auf die ich antworten kann. Ich meine, wenn ich auf Instagram bin, dann denke ich nicht als erstes an Hate-Kommentare*“ (G2, weiblich, min 01:20). Für Jugendliche ist es mittlerweile auch zur Normalität geworden, dass man man Hate-Kommentaren im Internet begegnet und diese in der Regel ignoriert, was auch insgesamt als beste Counter Speech Strategie eingeschätzt wird: „*Die beste Lösung ist nicht dagegen anzukämpfen, sondern es [=Hate Speech] einfach zu ignorieren ... Das Internet ist anonym und ein komischer Ort, deshalb kann man da nichts dagegen machen*“ (G3, männlich, min 59:41). Gleichzeitig wird es aber auch als wichtig betrachtet, zu informieren, dass Hass keine Meinung ist.

Hinzu kommt, dass sich die **Definition von Counter Speech als zivilcouragierte Handlung gegen Hate Speech und andere negative Inhalte im Internet nicht unbedingt mit jenen Vorstellungen deckt, die Jugendliche damit verbinden:** Counter Speech in größeren öffentlichen Kontexten umfasst neben Beiträgen in Kommentarspalten aus Sicht Jugendlicher zum Beispiel auch Reaktionen auf den provozierenden oder negativen Content von Influencer*innen, etwa in Form von Reaktionsvideos anderer Influencer*innen (oft mittels „Duetten“, d.h. das provozierende Video wird hergezeigt und kommentiert) oder in Form von zwar aggressiven, aber durchaus „verständlichen“ Gegenattacken durch User*innen. Auch die „Cancel Culture“ wird mit Counter Speech in Verbindung gebracht, d.h. wenn (bekannte) Personen aufgrund ihres (vermeintlichen) Fehlverhaltens auf allen Plattformen attackiert, ausgeschlossen bzw. zensuriert werden. Counter Speech wird auch mit Auseinandersetzungen bei kontroversen Themen (z.B. Veganismus, Homosexualität) assoziiert, die teils zwar als faktenbasierte Gegenrede wahrgenommen werden, teils aber auch einfach als eher negativ konnotierte, unreflektierte Meinungsabgaben verstanden werden. Auch Gegenreaktionen von Fan-Gruppen werden unter dem Begriff Counter Speech subsummiert, als jene, die als Unterstützer*innen *immer* hinter ihren favorisierten Celebrities stehen und sich für diese einsetzen. Die Vorstellungen von Counter Speech decken sich am ehesten dann, wenn es um Interventionen für Freund*innen geht oder wenn eigene Opfererfahrungen vorhanden sind. Counter Speech wird dann eher als ein emotionsgeladenes, intuitives Handeln gesehen („Wenn es einem am Herzen liegt, dann macht man das intuitiv, dann ist einem das egal, was mit einem selber passiert“ G2, weiblich, min 01:16).

Counter Speech wird zudem nicht unbedingt als Interventionen durch Dritte verstanden, sondern oft als eine Reaktion des betroffenen Opfers selbst. Dass wird häufig bei Videospielen wahrgenommen, wo betroffenen Opfer oft auf aggressive Art und Weise Spielunfähigkeit vorgeworfen wird und diese sich dagegen zur Wehr setzen, aber auch z.B. in WhatsApp-Gruppen, wo das Opfer selbst eine Gegenreaktion

zeigt. Gelungener Konter wird dann mit einer Demonstration der Überlegenheit durch das Opfer selbst gleichgesetzt, wobei gelungene Reaktionen von Opfern insgesamt auch mehr Anerkennung als Reaktionen durch Dritte erhalten. Counter Speech durch dritte Personen wird oft eher als Ausdruck der eigenen Genervtheit denn als Gegenrede wahrgenommen (*„Es ist dann eher so, dass die Leute genervt sind und dann sagen, hey, hört auf zu spammen“* G1, weiblich, min 19:08). Oft wird die Situation dann so gelöst, dass die betroffenen Personen als Reaktion aus der Gruppe geworfen werden bzw. einfach eine neue Gruppe gegründet wird oder die Gruppe auf stumm geschaltet wird. Alternativ wird in solchen Gruppen aber oft auch mit deeskalierenden, meist humorvollen Memes oder GIFs oder „Insidern“ interveniert, um von Konflikten abzulenken – zu einer konstruktiven Konfliktlösung wird dabei aber nicht beigetragen, sondern Übergriffe eher „ausgelagert“ oder zur Seite geschoben.

Die geringen Erfahrungen im Praktizieren von Counter Speech, die insgesamt geringe Attraktivität in der Wahrnehmung von Counter Speech durch Dritte, sowie die gleichzeitig klare moralischer Verurteilung von Hate Speech stellt für die Jugendliche insgesamt eine hohe Herausforderung dar, geeignete Counter Speech Strategien zu entwickeln.

3.2.2 Jugendliche Gestaltungskompetenzen und Interventions-Strategien in privaten Kontexten (z.B. WhatsApp-Gruppen)

Im Zuge des ersten Aktionsforschung-Workshop wurde den Jugendlichen ein Auszug aus einem Online Rollenspiel (Chatprotokoll) präsentiert: Aufgabe war es, die Akteur*innen im Spiel zu charakterisieren und eine Handlungsbewertung vorzunehmen. Anschließend wurde den Jugendlichen die Aufgabe gestellt, sich zu überlegen, wie die Counter Speaker*in im Online Rollenspiel anders / besser / wirkungsvoller reagieren hätte können, wobei auch kreative Interventionsmöglichkeiten z.B. mittels GIFs oder Memes möglich waren. Ziel war es, dass die jugendlichen Aktionsforschungsteilnehmer*innen konkrete Konter-Kommentare / -Postings ausformulieren, so wie das auch real gepostet werden könnte.

Für die Umsetzung wurde die Chatfunktion der genutzten Videokonferenz-Plattform verwendet oder es wurde direkt in die forschungseigene WhatsApp-Gruppe gepostet. Die Kommentare / Postings wurden mittels WhatsApp-Web projiziert (bzw. mit der „Bildschirm teilen“-Funktion) und anschließend gemeinsam besprochen. Während dieser Diskussion gab es auch Input durch die Forscher*innen, um unterschiedliche strategische Ausrichtungen in den Blick zu nehmen, z.B. wie es gelingen könnte, die Situation insgesamt zu deeskalieren / den Übergriff zu „unterbrechen“; den/die Täter*in wirkungsvoll zu konfrontieren; das Opfer wirkungsvoll zu unterstützen; andere mitlesende Peers zur Unterstützung zu mobilisieren.

Der vorgelegte Rollenspiel-Auszug (inszeniert wurde Szene 2: von jemandem wird ein Nacktvideo/-bild verbreitet, siehe dazu entsprechendes Kapitel zu den Online Rollenspielen in diesem Berichtsband) wurde prinzipiell als realistisch eingeschätzt, zum Teil waren ähnliche Szenarien die Aktionsforschungsteilnehmer*innen auch bekannt, die Art der Umsetzung (die Inszenierung wurde sehr ausfällig mit vielen obszönen Ausdrucksweisen gestaltet) wurde aber eher jüngeren Jugendlichen (z.B. 12-15 Jahre) zugeschrieben.

In einer ersten Reaktion zeigten sich die Jugendlichen eher überfordert mit der Vorstellung, selbst als Counter Speaker*in in so einer Situation intervenieren zu müssen. Es wurde befürchtet, selbst zum Opfer zu werden, da die Täter*innen als aggressiv eingestuft wurden (*„Ich wüsste nicht, was ich in dem Moment machen soll. Ob ich lieber jemanden anrufen soll [z.B. Rat auf Draht] oder auf auf die einreden sollte“* (G2, weiblich, min 18:18). Die Bemühungen einer Counter Speaker*in im Rollenspiel-Auszug wurde prinzipiell als erfolglos eingestuft. Der Grund sei, dass diese zu nett, zu vorsichtig, zu leise (findet kein Gehör) und zu schwach (*„bitte hört auf“*) dagegen vorgehe, und somit nicht ernst genommen werde und es nicht schaffe, die Situation zu entschärfen. *„Ja die versinkt halt einfach in dem Hate, weil die wird gar nicht so beachtet von den anderen, auch gar nicht geantwortet, gar nichts. Sie antwortet zwar auf die anderen Sachen aber sie wird so überfahren sag ich mal“* (G4, weiblich, min 56:30).

Die Jugendlichen wurden anschließend um die **Entwicklung konkreter Verbesserungsvorschläge bzw. Counter Speech Strategien bezogen auf das Rollenspiel gebeten**, die nun kurz vorgestellt werden:

Strategie: Autoritäre und direkte Zurechtweisung

Die Jugendlichen treten auf eine direkte Art und Weise, die auch an erwachsene Autoritätspersonen auf mit dem Ziel, ernst genommen zu werden. Das verlangt gleichzeitig ein hohes Maß an Selbstsicherheit und vermittelt aber gleichzeitig auch Überheblichkeit. Der Vorschlag wurde zudem unter der Annahme geschrieben, dass es sich hier um vermeintlich jüngere Täter*innen handelt. In der Regel werden direkte Handlungsaufforderung formuliert, z.B. mit der Aufforderung negative Inhalte zu löschen, oft in Kombination mit einer moralischen Zurechtweisung, oder auch Hinweise auf mögliche strafrechtliche oder anderen Folgen. Häufig wird dies auch durch passende Memes unterstützt, um z.B. zu signalisieren, dass der Übergriff als „unnötig“ wahrgenommen wird, gleichzeitig wird dadurch die Situation aber auch insgesamt etwas aufgelockert. Vorgeschlagen wird in diesem Zusammenhang auch ein öffentliches Bloßstellen oder Anprangern, z.B. mittels Instagram Story, wobei dieser Vorschlag nicht von allen Jugendlichen unterstützt wird. Ein anderer Zugang zur Stärkung der eigenen Autorität wird auch durch Bezeichnungen von Täter*innen als „kleines Kind“ versucht (*„Jemanden im Internet als kleines Kind zu beleidigen wirkt im Internet ziemlich viel. Da wirkt man so dominanter, ok du bist ein kleines Kind und ich bin groß und ich hab dir was zu sagen quasi“* G3, männlich, min 01:17:24). Insbesondere bei Jugendlichen mit höherer Bildung wird dies aber eher als unreifes Auftreten interpretiert: *„wird sind nicht im Kindergarten“* (G1, männlich, min 01:07:08).

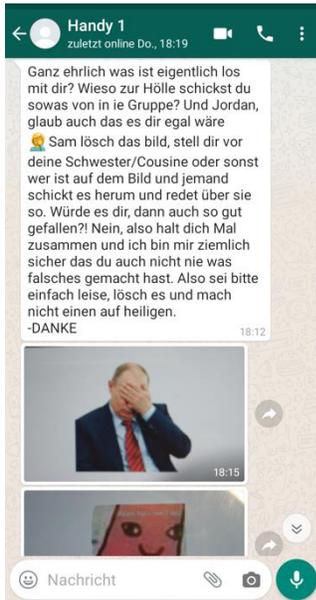


Abbildung 1: Beispiel aus G1
(Anmerkung: mehr als ein Jahr vor Ausbruch des Ukraine Kriegs entstanden)



Abbildung 2: Beispiel aus G3

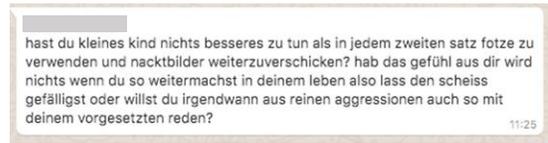


Abbildung 3: Beispiel aus G3

Strategie: Angebot einer Klärung oder Appell an die Moral und abschließende Drohung

Auch hier ist es strategisches Ziel, als Counter Speaker*in ernstgenommen zu werden. Wieder zeigt sich ein Muster, das auch von erwachsenen Autoritätspersonen stammen könnte: zunächst ein Angebot der gemeinsamen Klärung bzw. ein Appell an die Moral, gefolgt von der Ankündigung möglicher Konsequenzen. Der Kontakt zur Täterin / zum Täter wird hier über den Privat-Chat versucht, um mehr auf die Person eingehen zu können bzw. diese besser zu erreichen. Öffentliche Konfrontationen gelingen aus Sicht Jugendlicher nicht, es würde in der Öffentlichkeit eher der Druck für Täter*innen steigen, in ihrer Rolle zu bleiben: Zudem könnten insbesondere öffentlich gepostete moralisch aufgeladene Inhalte aus Sicht von Jugendlichen unpassend wirken bzw. das eigene Auftreten schwächen („Ich glaube, wenn ich da so einen spießigen Text in eine Gruppe schreiben würde, dass das erst mal so uncool wirken würde“ G3, männlich, min 01:21:29).



Abbildung 4: An Täter*innen adressiert, Beispiel aus G1

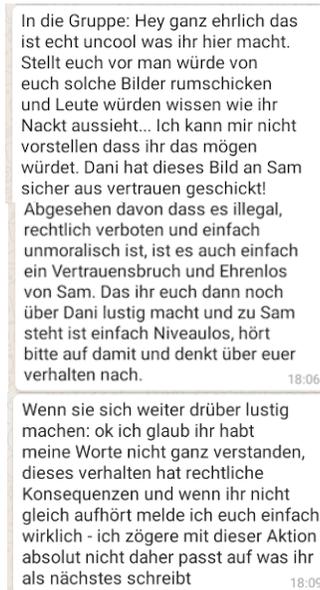


Abbildung 5: An Mitlesende adressiert, Beispiel aus G1

Strategie: Direkte Konfrontation bzw. Gegenangriff an Täter*innen

Auch hier steht im Vordergrund, sich als Counter Speaker*in als stark, selbstsicher und damit ernst zu nehmend zu präsentieren. Ziel sei es hier, Hater*innen herabzuwürdigen, sich über diese lustig zu machen und bloßzustellen, was auch durch GIFs gestützt werden kann. Bewirkt werden soll aus jugendlicher Sicht, dass sich Täter*innen schlecht fühlen und beginnen ihr Handeln zu reflektieren. Ein einfaches „hör auf“ würde aus Sicht der Jugendlichen nicht funktionieren.

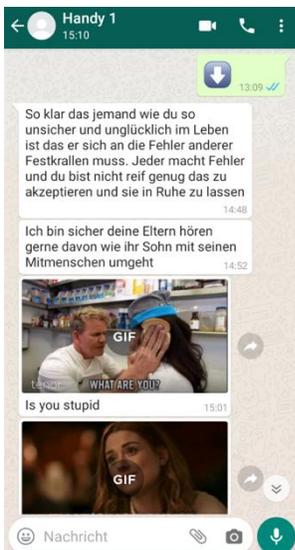
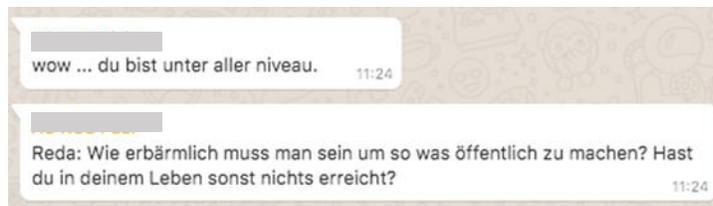


Abbildung 6: Beispiele aus G2



Strategie: Einschüchterung mit Fakten zu rechtlichen Konsequenzen

Mit dem Posten eines konkreten Gesetzestextes sollen Hater*innen durch die glaubhafte Ankündigung rechtlicher Konsequenzen eingeschüchtert werden. Interessant sind die unterschiedlichen Zugänge, mit denen diese Idee verknüpft wird (siehe untenstehende Beispiele): im ersten Fall, in dem ein Ausschnitt des eigenen Gesichtes signalisiert, dass hier tatsächlich eine Person dahintersteht, die die Umsetzung ernst meint. Im zweiten Fall wird der Gesetzestext mit einer moralischen und emotional aufgeladenen Begründung kombiniert, die unterstreicht, warum die Drohung einer Anzeige bzw. die Weiterleitung an eine erwachsene Autoritätsperson umgesetzt werden wird; eine weitere Kontaktaufnahme zur / zum Counter Speaker*in, um diese/n vielleicht doch noch davon abzuhalten, wird mit „macht’s gut ich trete aus“ ausgeschlossen. Im dritten Fall wird der Gesetzestext mit einem Appel an die Moral unterstützt, wobei dieser stärker im Vordergrund steht als die die mögliche Konsequenz einer Strafe, die quasi lediglich als Möglichkeit in den Raum gestellt wird („außerdem ist es strafbar“). Allerdings verlangt ein solcher Zugang entsprechende intellektuelle Fähigkeiten und Wissen, um auf solche Fakten hinzuweisen.

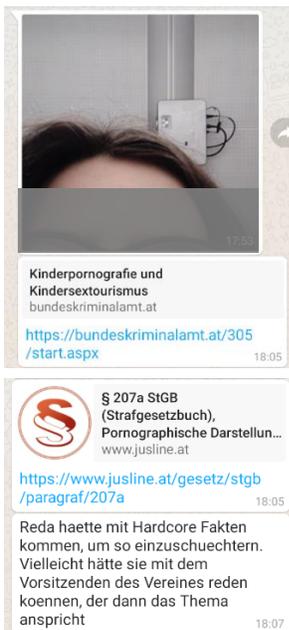


Abbildung 7: Beispiel aus G1



Abbildung 8: Beispiel aus G2

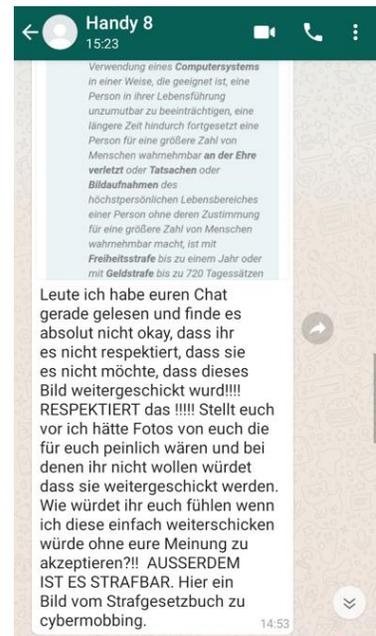


Abbildung 9: Beispiel aus G2

Im Rahmen der Aktionsforschung wurden entwickelte Strategien auch anderen Gruppen präsentiert mit der Bitte um entsprechende Bewertung. Dabei kamen für einen solchen Zugang auch Einwände: Jugendliche mit höherem Bildungshintergrund meinten, dass zwar der Inhalt der Paragraphen verständlich sei, aber nicht klar wäre, ob sich die betroffene Person dies dann auch durchliest. Jene mit geringem Bildungshintergrund meinten dazu, dass das viele trotzdem nicht ernstnehmen würden, da die meisten wüssten, dass man auch aufgrund der Anonymität nicht sofort eine Anzeige machen würde.

Strategie: Mit Memes Aufmerksamkeit schaffen und dann die eigentliche Botschaft übermitteln

Durch das vorangestellte Meme passend zur Situation soll auf humorvolle Art und Weise Aufmerksamkeit erzeugt werden, bevor die eigentliche Botschaft, z.B. mit der Aufforderung, damit aufzuhören, übermittelt wird. Das Meme vermittelt aber für die Jugendlichen auch auf eine subtile Art und Weise so etwas wie: „seid nicht böse auf mich!“ und lockert die Situation insgesamt auf. Memes oder GIFs werden generell als geeignetes Mittel gesehen, um Text zu unterstützen, sind aber nicht immer passend bzw. muss je nach Inhalt ein passendes Meme gefunden werden: z.B. bei Gesetzestexten würde man eher etwas passiv Aggressives, Angstmachendes einsetzen; Memes / GIFs müssen nach Ansicht der Jugendlichen mit Bedacht eingesetzt werden, da sie eher mit guter Laune verbunden werden und dadurch nicht vermittelt werden soll, dass die Situation nicht so schlimm sei. Insofern eignen sich GIFs nach Ansicht der Jugendlichen eher dazu, die Situation insgesamt zum Deeskalieren zu bringen, aber weniger, um Täter*innen oder Opfern etwas mitzuteilen. Oft sei die Situation einfach zu ernst, als dass man so etwas noch schicken könnte.



Abbildung 10: Beispiel aus G1



Abbildung 11: Beispiel aus G2



Abbildung 12: Beispiel aus G1

Strategie: Ausschließlicher Einsatz bildbasierter Reaktionen, z.B. mit GIFs oder Memes

Jugendlichen macht es prinzipiell Spaß, nach passenden Memes zu suchen, die als Reaktion auf Hate Speech einsetzbar wären. Allerdings wird das nicht in allen Situationen als passend wahrgenommen, da diese in der Regel humorvoll gestaltet sind und Humor als Strategie insgesamt als sehr ambivalent erlebt wird. Passend sei das eher „bei kleinen Nervereien“, z.B. wenn jemand ständig nach Hausübungen fragt. Bei Nacktbildern würde das nicht mehr funktionieren, da sich das Opfer angegriffen oder nicht ernst genommen fühlen würde und Täter*innen nicht mit Konsequenzen konfrontiert wären. Voraussetzung sei außerdem, dass auch die mitlesenden Peers entsprechenden Humor aufweisen, sonst könne alles schlimmer werden. Zudem müsse man als Counter Speaker*in entweder selbst kreativ sein oder zum richtigen Zeitpunkt passende humorvolle Kommentare formulieren können bzw. Memes oder GIFs finden – vor allem Jugendliche mit geringem Bildungshintergrund meinen dazu, dass ihnen oft nicht viel dazu einfallen würde oder das einfach in der Situation nicht gelingt – auch im Zuge der Aktionsforschung kam

von dieser Gruppe an Jugendlichen kaum Vorschläge in dieser Richtung. Der Einsatz von Emojis wird ebenso von manchen skeptisch gesehen, da es hier ebenfalls Spielraum für Missverständnisse geben würde, vor allem bei sarkastischen oder ernst gemeinten Inhalten.

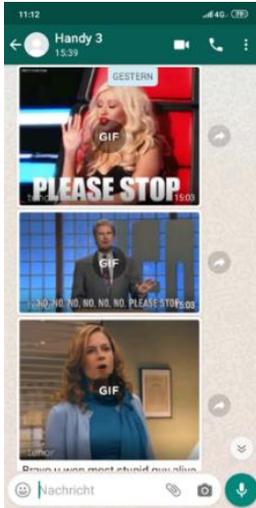


Abbildung 13: Beispiel aus G2



Abbildung 14: Beispiel aus G2



Abbildung 15: Beispiel aus G2

Strategie: auf Englisch (oder anderen Fremdsprache) kontern

Ein weiterer interessanter Zugang, den einzelne Jugendliche mit höherem Bildungshintergrund (meist besonders internetaffin) praktizierten ist es auf Englisch zu kommentieren, was allerdings auch eine sprachliche Versiertheit voraussetzt: In einer Fremdsprache zu schreiben stellt eine größere emotionale und kognitive Distanz her und macht es leichter, unangenehme oder schwierige Inhalte zu transportieren bzw. intervenieren – andererseits sind auch Kraftausdrücke und obszöne Begriffe leichter auszusprechen.



Abbildung 16: Beispiel aus G3 - Englisch

Strategie: Schlechtes Gewissen erzeugen

Diese Strategie wird hauptsächlich von Jugendlichen mit niedriger Bildung vorgeschlagen. Die Argumentationsweise orientiert sich dabei an der Opferperspektive. Ziel ist es für die Jugendlichen, Täter*innen ihr falsches Verhalten zu erklären bzw. ein Verständnis zu schaffen: „*Ich glaub Menschen können eine Situation besser verstehen, wenn sie sich vorstellen wie sie wären in so einer Situation oder wenn sie das schon erlebt haben oder gefühlt haben*“ (G4, weiblich, min 01:04:47). In der konkreten Umsetzung wird versucht, Konsequenzen für das Opfer zu verdeutlicht werden, sein/ihr Verhalten aus der Perspektive nahestehender Personen (z.B. Eltern) zu betrachten bzw. Übergriffe auf die Person des/der Täter*in selbst oder einer ihm/ihr nahestehenden Person zu projizieren.

"Durch dass du sein/ihr vertrauen missbraucht hast, wird es schwer sein für sie/ihn das jemals zu überwinden. Es kann auch noch weiter, viel schlimmere Folgen haben, wie das sie sich das Leben nimmt und dann hast du ein Menschenleben auf dem Gewissen, schonmal darüber Nachgedacht? "

Ich würd ihn mal fragen ob er denkt das er was in seinem Leben erreicht hat und seine Eltern stolz auf ihn sind wenn er schon so tief fällt

Abbildung 17: Beispiel aus G3

Stell dir vor das is deine Tochter 🐱 11:24

"Wie würdest du es finden wenn so ein Foto von dir in Umlauf gebracht wird, stell dir vor wie sie/er jetzt in der Schule gemobbt wird und wenn es noch auf anderen Plattformen weitergeleitet wurde wie sie da jetzt als Opfer dasteht, du willst bestimmt auch nicht als Mobbingopfer dastehen nur weil du einer Person vertraut hast."

Stell dir vor das wäre deine Freundin, könntest du so mit gutem Gewissen leben? Was willst du damit erreichen? anscheinend haben dir deine Eltern nicht genug Aufmerksamkeit geschenkt

Abbildung 18: Beispiel aus G3

Strategie: Opfer unterstützen

Auch diese Strategie wird vorrangig von Jugendlichen mit geringer Bildung vorgeschlagen. Generell ist es nach Ansicht der Jugendlichen besser, emotionalen Support für das Opfer eher privat zu übermitteln: Es könnte für das Opfer sonst noch unangenehmer werden, wenn öffentlich eine große Sache daraus gemacht wird. „Also die meisten wollen nicht die Opfer sein, von denen dann jeder lernt oder weiß. Deswegen kann man damit nicht so öffentlich gehen. Außer halt die Person ist damit einverstanden, dass man damit öffentlich geht und so eine Geschichte erzählt“ (G4, weiblich, 01:06:49). Generell würde es Opfern am meisten helfen, einfach darüber sprechen zu können und zu wissen, dass Leute hinter einem stehen. Die Unterstützung von Opfern kann ganz unterschiedlich gestaltet sein: z.B. eher indirekt mittels Memes, oder auch durch direkte Vermittlung von Trost oder auch Verhaltens-Tipps.



Abbildung 19: Beispiel aus G3

Ich hätten dani privat geschrieben:
Verlass die Gruppe mach screenshots !
mach dir keine Sorgen das passiert sehr viele frauen du bist nicht allein und früher oder später werden diese Leute es sehr breuen
Wenn du jemanden zum reden brauchst kannst du mich jederzeit kontaktieren

Abbildung 20: Beispiel aus G4

Strategie: Einfache Aufforderung aufzuhören

Ebenso ausschließlich von Jugendlichen mit geringer Bildung wurden von manchen auch Vorschläge gemacht, die sich kaum von der Vorgehensweise der fiktive Counter Speaker*in im Rollenspiel unterscheiden. Dies wurde damit begründet, dass ihnen nichts Besseres einfallen würde oder dass sie keine Meinung dazu hätten. Ihnen sei klar, dass solche Strategien nicht viel bringen würden, aber man könne es versuchen. Einigen Jugendlichen fehlte es auch an notwendigen Kommunikationskompetenzen bzw. Kompetenzen im Auffinden von Argumenten.



Abbildung 21: Beispiel aus G3



Abbildung 22: Beispiel aus G4

Strategie: Mitlesende Peers zur Unterstützung mobilisieren

Insgesamt waren alle von den Jugendlichen eingebrachten Vorschläge hauptsächlich täter*innenorientiert ausgerichtet. Von den Forscher*innen wurde daher angeregt, zu überlegen, ob und wie auch eine Mobilisierung mitlesender Peers zur Unterstützung des eigenen Counter Speech Verhaltens gelingen könnte. Die Jugendlichen meinen dazu, dass das vor allem im Privat-Chat gelingen könne, da aufgrund des Gruppendrucks vorgefasste Meinungen nicht geändert werden: *„Bei so Gruppenchats steht jeder auf seiner Position. Entweder sie sind auf der Seite von diesem Typen oder von diesem Mädels. Und ich schätze mal, es wird ziemlich schwer sein, die zu überzeugen, die Position zu wechseln, die sind meistens sehr stur.“* (G3, männlich, min 01:34:13). Insgesamt haben die Jugendlichen kaum Erfahrungen, sich im Vorgehen gegen Übergriffe zusammenzuschließen, am ehesten noch, wenn man sich privat ausgetauscht hat und anschließend eine eigene Gruppe gebildet hat. In der gemeinsamen Diskussion wurden von den Jugendlichen folgende Vorschläge genannt: einen besonders gelungenen Kommentar (z.B. besonders schlagfertig bzw. lustig oder sehr klug, etwas Lustiges mit Memes, GIFs) posten, damit das andere unterstützen, mit einem Hate-Posting eine Instagram-Story machen und dann andere auffordern dies zu kommentieren bzw. einen Link zum Hate Kommentar kopieren und in Communities posten, die unterstützen könnten. Die konkrete Unterstützung würde sich am ehesten mit einem „Like“ bemerkbar machen oder mit kurzen Kommentaren wie z.B. „danke“ oder „endlich mal vernünftige Menschen“. Die Forscher*innen stellten auch die Nutzung eines Hashtags in den Raum, was aber von Jugendlichen nur wenig aufgenommen wurde. Sich zusammenzuschließen und als Gruppe gegen Hate vorzugehen ist zumindest für einen Teil der Jugendlichen vorstellbar, allein gegen Hate vorzugehen sei jedenfalls nicht zielführend. Tatsächlich wahrgenommen hätten sie das bisher aber noch nicht in ihrem Internet-Alltag, außer, dass es vorkommen könne, dass Counter-Kommentare mehr Likes erhalten als Hate-Kommentare. Das würde aber eher bei aufgeladenen Themen wie Rassismus etc. vorkommen, weniger bei einzelnen Personen, die angegriffen werden.

Strategie: Ignorieren und (später) mit Ausschluss sanktionieren

Hater*innen zu ignorieren wird zwar auch von den Jugendlichen nicht unbedingt mit Counter Speech gleichgesetzt, ist aber die übliche Praxis. Der Ausschluss von Personen z.B. aus der Gruppe, um unangebrachtes Verhalten zu sanktionieren, wird aber als mögliche Gegenreaktion verstanden, ebenso wie jemanden zu blockieren oder stumm zu schalten. Die jugendlichen Aktionsforschungsteilnehmer*innen sehen das prinzipiell als das vergleichsweise effektivste Mittel (*„Wenn die Person keine Einsicht zeigt, dann würde ich die mal komplett aus der Gruppe ausgrenzen und komplett links liegen lassen, vielleicht dass sie mal darüber nachdenken, was sie macht Ich würd sie jetzt*

nicht raushauen, aber mal komplett ignorieren. Wenn sie aber nicht aufhört, dann aber einfach rauswerfen und solange draußen lassen, bis sie den Fehler mal eingestanden hat.“ G3, männlich, min 01:15:00).

Die Ergebnisse zeigen, dass zwar von Jugendlichen immer noch die Strategie des Ignorierens bzw. Ausschließens von Täter*innen als am effektivsten wahrgenommen wird, sich aber für aktive Counter Speech drei Hauptstrategien ableiten lassen, die klar Täter*innen-orientiert ausgerichtet sind und aus der Sicht von Jugendlichen als am wirkungsvollsten empfunden werden: (1) Sich selbst als ernst zu nehmender Counterpart präsentieren und Stärke beweisen, z.B. durch autoritär vermittelte Handlungsaufforderungen, Drohungen, Einschüchterungsversuchen und Zurechtweisungen oder / und einer direkten Konfrontation mittels Gegenangriff, (2) sich Gehör schaffen durch Taktiken, die Aufmerksamkeit schafft (z.B. mittels Memes/GIFs), um dann entweder gezielt Botschaften zu vermitteln oder zumindest zu deeskalieren, (3) indirekte Wege der Vermittlung wählen (z.B. auf Englisch kommunizieren oder GIFs/Memes einsetzen), um die Unsicherheit im eigenen Auftreten als Counter Speaker*in insbesondere bei der Vermittlung von „uncoolen“ weil z.B. moralischen Inhalten zu umgehen, Jugendliche mit geringem Bildungshintergrund sehen aber auch die direkte Konfrontation von Täter*innen mit der moralischen Verwerflichkeit als möglichen erfolversprechenden Weg. Strategien zu Mobilisierung mitlesende Peers zur Unterstützung werden nicht mitgedacht, könnten aber nach Einschätzung der Jugendlichen wirksam sein, allerdings fehlt es dazu an konkreten Erfahrungen und Kompetenzen. Vergleichsweise wenig effektiv werden Strategien eingeschätzt, die mit einfachen Mittel zum Aufhören aufrufen, allerdings mangelt es diesen Personen auch an alternativen Counter Speech Ideen.

Die Ergebnisse verweisen außerdem auf bildungsspezifische Differenzen in der konkreten Umsetzung von Counter Speech. Während Jugendliche mit höherem Bildungshintergrund eher versuchen, ihre intellektuelle Überlegenheit und Bevormundung auszuspielen, versuchen Jugendliche mit geringerem Bildungshintergrund hingegen eher aus der Opferperspektive zu argumentieren und versuchen sich in moralisierenden Zugängen, die oft aber auch in Gegenattacken münden.

3.2.3 Jugendliche Gestaltungskompetenzen und Interventions-Strategien in öffentlichen Kontexten (z.B. TikTok, Instagram, YouTube)

Im Zuge der Aktionsforschung wurde „Speed Counter Speech“ als weitere Übung entwickelt und in die Aktionsforschung integriert. Ziel sollte es sein, reale Settings in öffentlichen Kontexten (z.B. auf YouTube, Instagram, TikTok) zu simulieren, bei denen man in der Regel unvorbereitet auf Hate Kommentare trifft und meist auf rein textbasierter Ebene reagieren muss. Dabei war es Aufgabe der Jugendlichen, jeweils eine Minute lang spontan auf vom Forschungsteam ausgewählte Hate-Kommentare (reale Hate-Kommentare, die im Zuge einer Inhaltsanalyse auf Social Media gepostet wurden – ohne Absender*innen-Information) zu reagieren. Dabei konnte rein auf der Textebene und nur mit Einsatz von Emojis reagiert

werden (ohne GIFs, Memes), was auch den Einschränkungen. Diese wurden von den Forscher*innen in den Chat (MSTeams) gepostet. Unmittelbar anschließend wurde das eigene Vorgehen gegen diese inszenierten Online-Übergriffe reflektiert, der gesamte Chatverlauf noch einmal gemeinsam analysiert und mögliche Verbesserungsvorschläge diskutiert.

Im Folgenden werden die jeweiligen Counter Speech Reaktionen auf unterschiedliche Formen von Hate Speech zusammengefasst:

Aggressive Beschimpfung und Abwertung

Beispiel eines geposteten Hate-Kommenars:



10:50
Halt doch dein Maul xD Du Stück Dreck...so etwas wie dich wollen wir nicht

Abbildung 21

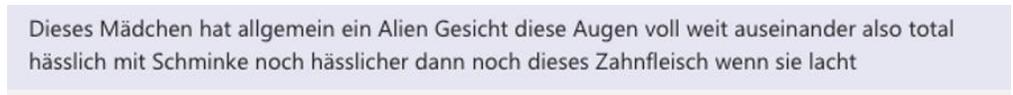
Beispielhafte Counter Speech Reaktion der Jugendlichen:

- Gegenattacke: „Halt du selber MAUL du stück SGeiße wenn ich dich sehe dann prügele ich dich bist du K.O. bit“
- Retourkutschen: „Wer hat gesagt das wir dich wollen?“
- Sich lustig machendes Kurzkomentar: „Lol xdd“

Auffällig ist, dass vor allem Jugendliche mit geringem Bildungshintergrund zu aggressiveren Gegenattacken neigen. Jene mit höherem Bildungshintergrund versuchen eher, sich über Hater*innen lustig zu machen. Solcher Art Hate-Kommentare bieten kaum Spielraum für Gegenargumente, reagiert wird hier eher aus der Emotion heraus. Auf das Opfer selbst wird nicht Bezug genommen – allerdings lässt das Hate-Kommentar auch offen, wer als Opfer konkret angesprochen wird.

Hate-Kommentare gegen das Aussehen von Personen

Beispiel eines geposteten Hate-Kommenars:



Dieses Mädchen hat allgemein ein Alien Gesicht diese Augen voll weit auseinander also total hässlich mit Schminke noch hässlicher dann noch dieses Zahnfleisch wenn sie lacht

Abbildung 22

Beispielhafte Counter Speech Reaktion der Jugendlichen:

- Opfer verteidigen: „Für Ihre Gesichtsform kann sie nichts genauso wenig für das wie Ihre Augen platziert sind!“; „Jeder Mensch ist auf seine eigene Art und weise schön“

- Moralisierende Zurechtweisung: „Ich glaube nicht, dass du das Recht hast so zu Urteilen. Jeder Mensch hat Mensch hat Macken & jeder Mensch ist auf seine Weise schön.“; „Das ist nicht dein Problem wie sie aussieht!“
- Gegenattacke: „Wen interessiert deine Meinung, bist wahrscheinlich noch hässlicher“
- Ironische Retourkutsche: „Willst du jetzt damit sagen, dass sie besser aussieht als du?“
- Inhaltsleeres Kurzkomentar: „OMG“

Im Hate-Kommentar wird klar auf ein Opfer Bezug genommen. Während Jugendliche mit geringerer Bildung vor allem das Opfer verteidigen bzw. moralisierend vorgehen, versuchen jene mit höherer Bildung stärker schlagfertig gegenüber Täter*innen aufzutreten und ihnen die Ernsthaftigkeit abzusprechen.

Schwer übergriffiger Kommentar

Beispiel eines geposteten Hate-Kommenars:

10:45
was für ein Wichser du doch bist, ich hoffe du bekommst Hodenkrebs bevor du dich
fortpflanzen kannst

Abbildung 23

Beispielhafte Counter Speech Reaktion der Jugendlichen:

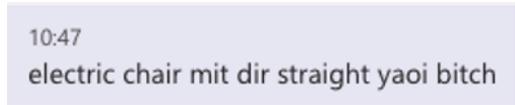
- Sich lustig machende Retourkutsche: „da hätten deine eltern sich wohl doch für das kondom entscheiden sollen...“
- Retourkutsche mit indirekter Drohung: „Schlimm wird es für ich, wenn alles, dass du anderen wünscht - dir passiert.“
- Retourkutsche mit versuchter Abwertung: „Auch lächerlich kann alles nur hoffen“
- Retourkutsche mit Abwertung: „*Ich hab wenigstens irgendwann ne Freundin zum Fortpflanzen, du bist so hässlich das du nie eine bekommen wirst xD*“
- Gegenattacke: „*Du bist selber ein Wichser und kriegst auch ein Hodenkrebs du Vollpfosten*“, „BASTARD“
- Moralische Zurechtweisung: „Das ist extrem respektlos gegenüber Menschen die wirklich mit Krebs zu kämpfen haben.“
- Kurzkomentare mit unterschiedlicher Konnotation
 - ironisch, abwertend: „bisschen cringe brudi“
 - Täter*innen ignorierend: „Juckt?“
 - moralische Bewertung: „shit bro“ „pain“

Die Reaktionen der Jugendlichen fallen hier vergleichsweise breiter und vielfältige aus. Auch hier sind es wieder Jugendliche mit geringerem Bildungshintergrund, die in den Gegenreaktionen eher aggressiv reagieren. Auffallend sind einerseits die verschiedenen Varianten, mittels Retourkutsche zu intervenieren,

indem konkrete Inhalte aus dem Hate-Kommentar aufgegriffen und anders verpack zurückgespielt werden. Das verlangt allerdings eine gewissen Sprachkompetenz, der Versuch eines Jugendlichen mit geringer Bildung („Auch lächerlich kann alles nur hoffen“) fällt dagegen deutlich schwächer aus. Andererseits sind auch die verschiedenen Kurz-Reaktionen auf Englisch auffallend, wie z.B. „shit bro“, „pain“, die von interneterfahrenen Jugendlichen mit höheren Bildung stammen und signalisieren, dass solche Aussagen nicht in Ordnung sind, gleichzeitig aber auch wenig weitere Angriffsfläche offen lassen. Ähnlich zeigt sich das bei den Kurzreaktionen „bisschen cringe brudi“ und „Juckt?“.

Schwer übergreifiger Kommentar mit irritierender Ausdrucksweise

Beispiel eines geposteten Hate-Kommenars:



10:47
electric chair mit dir straight yaoi bitch

Abbildung 24

Beispielhafte Counter Speech Reaktion der Jugendlichen:

- Kurzkommentare mit unterschiedlicher Konnotation:
 - hinterfragend: „aha“
 - ironisch, sich lustig machen: „nice“ „ist das ein date?“
 - abwertend: „Hawara wos“
 - moralische Bewertung: „NANA Nicht gut“
- Gegenattacke: „stu dumb ass bitch“; „solche leute wie du gehören gesteinigt“

Hate-Kommentare solcher Art irritieren und schockieren und bieten wenig Grundlage für fundierte Gegenargumente. Sie sind daher insgesamt schwer zu kontern und lassen als Gegenreaktion oft nur Kurzkommentare zu, die unterschiedlich konnotiert sein können. Insgesamt reagieren aber auch hier wieder Jugendliche mit geringerem Bildungshintergrund aggressiver bzw. auch mit einfachen Ausdrücken einer moralischen Bewertung („NANA Nicht gut“).

Hate Kommentare als Drohung

Beispiel eines geposteten Hate-Kommenars:



10:42
Sniper Pass auf wie du sprichst du Nutte ta qifsha warte ab du kommst nichtmehr nachhause diese Nacht

Abbildung 25

Beispielhafte Counter Speech Reaktion der Jugendlichen:

- Abwertendes Sich lustig machen: „bohyyy ich hab angst chef“; „Plolizai“; „Diese super Deutsch mal wieder. 🤔“
- Gegenattacke: „bitte tu mir den gefallen und lösche dich“
- Verteidigung des Opfers: „Man lass sie doch sie ist keine Nutte sondern eine Frau“

Die Jugendlichen signalisieren in ihren Reaktionen vor allem, die formulierte Drohung nicht ernst zu nehmen. Lediglich ein Jugendlicher mit geringer Bildung versucht mit einer einfachen Aufforderung, das Opfer zu schützen. Solcher Art Drohungen, wo im Internet physische Bedrohungen übermittelt werden, bieten Jugendlichen auch wenig Chancen für Interventionen, da weitere mögliche Schritte in solchen Fällen, wie beispielsweise die Polizei zu informieren, zu viele Kontextinformationen fehlen.

Übergriffe mit fehlendem Kontextbezug

Beispiel eines geposteten Hate-Kommentars:

A screenshot of a social media comment. The background is light purple. The text is white. It shows a timestamp '10:46' and the comment text 'Redet wie ein Azzlack, kein Wunder, dass man ihn nicht ernst nimmt'.

Abbildung 26

Beispielhafte Counter Speech Reaktion der Jugendlichen:

- ironisch, sich lustig machendes Kurzkommentar: „big depression“
- Rückfragen: „Wer sagt das ihn keine ernst nimmt?“
 - abwertend: „Hawara wos“
 - moralische Bewertung: „NANA Nicht gut“

Dass das Fehlen an Kontextinformationen auch die Handlungsoptionen von Jugendlichen stark einschränkt, zeigt auch dieses Beispiel: Insgesamt reagieren die Jugendlichen aufgrund des Fehlens wesentliche Kontextinformationen mit nur sehr wenigen Gegenreaktionen: es ist nicht klar, auf wen sich dieser Übergriff bezieht, vermutlich auf eine Person in einem Video; diese Vermutung scheint offenbar für die Jugendlichen zu wenig Grundlage für Gegenreaktionen zu sein. Möglicherweise wird auch der Begriff „Azzlack“ nicht eindeutig als Übergriff interpretiert. Entsprechend fallen die Reaktionen eher in Form von eher „unverbindlichen“ Kurzkommentaren aus.

Hate-Kommentar mit humorvoller Komponente

Beispiel eines geposteten Hate-Kommenars:

10:44
du bist echt der allerletzte - aber ich hoffe echt dir fehlt n Strohhalm bei der nächsten Capri
Sonne

Abbildung 27

Beispielhafte Counter Speech Reaktion der Jugendlichen:

- Humorvolle, unterstützend: „HAHAHAHAHAHAHAH“
- Humorvoll, ironisch: „das geht zu weit“
- Retourkutsche: „Ich hoffe deine Ärmel rutschen beim händewaschen runter“
- Unverständnis: „Was redet der typ da was meint er mit dir fehlt Strohhalm bei der nächsten Capri Sonne?????“ „lächerlich“

Im Vergleich zu den anderen geposteten Hate-Kommentaren hat dieses auch eine humorvolle Komponente. Entsprechend fallen auch die Gegenreaktionen der Jugendlichen entweder ebenfalls humorvoll aus, oder es wird versucht, ebenso schlagfertig zu reagieren. Gerade Jugendliche mit geringem Bildungshintergrund gelingt eine solche Schlagfertigkeit oft nicht. Sie reagieren sprachlich einsilbiger bzw. mit Unverständnis oder negativen Kommentaren („was redet der typ...“, „Das ist lächerlich“).

Subtilere Formen von Hate Speech: Reproduktion negatives Gedankengut, Diskriminierungsformen

Beispiel 1 eines geposteten Hate-Kommenars:

10:49
Man braucht kein rechter Hetzer zu sein, um zu erkennen, dass der *Islam ein fragwürdiger Kult
ist

Abbildung 28

Beispielhafte Counter Speech Reaktion der Jugendlichen:

- abwertendes, inhaltsleeres Kurzkomentar: „Lächerlich“
- Gegenattacke: „Den einzigen Kult den ich da sehe is dein ekelhafter nazikult du hund“
- Sich lustig machen: „stfu und geh beten ich glaub du brauchst deine hilfe von oben gerade“
- Retourkutsche: „Dein Leben ist auch fragwürdig“
- Sachliche Gegenargumente: „Der Islam ist an sich voll die Friedliche Religion, nur manche Missbrauchen diese Religion um Hass und Gewalt zu verbreiten. Somit wird das Bild des Islam generell schlecht gemacht“

Der Kommentar wird zwar als schwer übergriffig aufgefasst, ist aber subtiler verpackt als das oben genannte Beispiel eines schwer übergriffiger Kommentars („Hodenkrebs“). Aufgrund der Schwere sind aber zeigt sich aber auch eine relativ große Breite an Gegenreaktionen: Auch hier sind es wieder Jugendliche mit geringer Bildung die zu aggressiveren Reaktionen tendieren, jene mit höherer Bildung eher zu „überheblichen“, sich lustig machenden Reaktionen, die auch eine gewissen Schlagfertigkeit verlangen. Interessant ist auch, dass es Jugendlichen mit geringer Bildung vergleichsweise weniger gut gelingt, inhaltlich zu argumentieren („lächerlich“), während andere sachlich fundierte Gegenargumente einbringen können und damit zusätzlich Strategien, wie z.B. Abwertungen von Hater*innen, nicht mehr erforderlich sind.

Beispiel 2 eines geposteten Hate-Kommenars:

10:43

Es gibt so viele Menschen die was gegen homosexuelle haben (ich auch) aber das muss man tolerieren. Manchen ist es egal und manchen eben nicht. Was ist denn daran so schwer zu verstehen. Nicht jeder muss homosexuell tolerieren

Abbildung 29

Beispielhafte Counter Speech Reaktion der Jugendlichen:

- Zustimmung: „Stimmt“
- Plädoyer für Toleranz: „Du solltest zumindest den anstand haben, menschen so zu tollerieren wie sie sind“; „Tolerieren sollte man es, Supporten allerdings muss man es nicht. Lässt die Leute Leben und Lieben wen sie wollen“; „lass die machen was die wollen und misch denen nicht ein“
- Gegenattacke: „der letzte satz beschreitbt deinen Bildungsstand.“
- abwertende Zurechtweisung: „Weißt was sei leies... Nur weil du mit deinem Leben nicht klar kommst.“
- Roturkutsche: „Es gibt so viele Menschen die etwas gegen dich haben (ich auch xdddd) aber das muss man tolerieren. Manchen ist es egal und manchen eben nicht. Was ist daran so schwer zu verstehen. Nicht jeder muss dich tolerieren.“

Das zweite Hate-Speech Kommentar ist ebenfalls subtil verpackt, die Schwere des Übergriffs ist nicht eindeutig zu bewerten. Trotzdem findet sich eine gewisse Vielfalt an Gegenstrategien: Homosexualität ist unter Jugendlichen ein kontrovers diskutiertes Thema, das auch stark im Internet ausgehandelt wird. Insofern haben sich Jugendliche in der Regel bereits eine Meinung gebildet, ob sie Homosexualität tolerieren oder nicht, entsprechend fallen auch die meisten Reaktionen der Jugendlichen als Plädoyer für Toleranz aus. Allerdings kann das sehr unterschiedlich formuliert und konnotiert werden und wirkt dann mehr oder weniger überzeugend. Auch hier zeigen v.a. Jugendliche mit geringer Bildung in der Art der Formulierung Schwächen (z.B. „lass die machen was die wollen und misch denen nicht ein“) und wirken im Auftreten weniger überzeugend. Schlagfertige Gegenreaktionen in Form einer „Retourkutsche“ kommen auch hier wieder ausschließlich von höher gebildeten Jugendlichen. Auffallend ist hier aber auch,

dass in diesem Fall zwar auch mit eher angriffigen Zurechtweisungen reagiert wird, aber nicht mit aggressiven Gegenattacken, da auch der diskriminierende Übergriff deutlich subtiler ausfällt und bzw. eher als Meinungsbeitrag gewertet werden kann – ohne Einsatz von Schimpfwörtern.

Zusammenfassend lässt sich feststellen, dass die Counter Speech Reaktionen der Jugendlichen nicht nur stark von der Art und Weise bzw. vom Typus des Hate-Speech Kommentars abhängen, sondern auch vom Bildungshintergrund der Jugendlichen: Den beteiligten Jugendlichen mit geringem Bildungshintergrund fehlt es deutlich an Basiskompetenzen hinsichtlich Argumentations-, Formulierungs- bzw. Ausdrucksweisen. Jugendliche mit höherem Bildungshintergrund zeigen im Vergleich nicht nur mehr Schlagfertigkeit, sondern verfügen auch über das erforderliche Wissen, um Themen mit sachlichen Gegenargumenten zu kontern. Eindeutige Positionen werden vor allem dann vertreten, wenn es um für Jugendliche zentrale Themen geht, wie z.B. Homosexualität, Religion oder Rassismus, zu denen schon individuelle Einstellungen und Werthaltungen entwickelt wurden. Auch hier ist die Argumentationsweise höher gebildeter Jugendlicher deutlich ausdifferenzierter und versierter als jene der geringer gebildeten. Besonders bei emotionalen, spontanen Reaktionen neigen Jugendliche mit geringem Bildungshintergrund auch häufiger zu aggressiveren Gegenattacken während jene mit höherem Bildungshintergrund versuchen – auch aus einer gewissen Selbstdarstellung heraus –, möglichst schlagfertig zu kontern.

Darüber hinaus orientiert sich der Grad der Aggressivität in den Gegenattacken häufig auch an der wahrgenommenen Aggressivität des Hate-Kommentars selbst: eindeutig übergriffige Hass-Kommentare werden entsprechend auch mit Beschimpfungen gekontert, eher subtilere Form eines Übergriffs werden oft als Aufforderung verstanden, auf argumentative Weise dagegen vorzugehen. Arten solche Übergriffe in konkrete Drohungen aus, so tendieren die beteiligten Jugendliche dazu, Täter*innen zu signalisieren, dass sie nicht ernstgenommen werden. Auffallend ist, dass nur jene beteiligten Jugendlichen mit geringerem Bildungshintergrund auch auf das Opfer Bezug nehmen, um es z.B. zu verteidigen oder zu moralisieren – allerdings auf eine Art und Weise, die sie schwach erscheinen lässt und an die Rolle der Counter Speech Protagonistin im Rollenspiel (siehe Kapitel 3.2.2. in diesem Bericht) erinnert.

Interessant sind auch Reaktionen in Form von Kurzkomentaren, wie „ok“ oder auch z.B. „shit bro“, „pain“, die sehr unterschiedlich konnotiert werden können und gleichzeitig wenig Angriffsfläche offenlassen. Jugendliche mit höherem Bildungshintergrund verwenden solche Ausdrücke versierter, z.B. mit englischsprachigen Ausdrücken gemäß den Gepflogenheiten der Internet-Lingo, Jugendliche mit geringem Bildungshintergrund erscheinen hier weniger versiert (z.B. „NANA Nicht gut“). Solche Kurzkomentare kommen häufig bei Hate-Komentaren vor, die aufgrund fehlender Kontextinformationen schwer verständlich sind bzw. auch keine inhaltlichen Gegenargumente zulassen und insgesamt schwer zu kontern sind. Die beobachteten Defizite in den Counter Speech Kompetenzen Jugendlicher mit geringer Bildung machen es nachvollziehbar, warum es während der Aktionsforschung

einige Jugendliche aus dieser Gruppe gab, die bei dieser Übung kaum oder zum Teil gar nicht kommentierten.

In der nachfolgenden Diskussion mit den Jugendlichen wurde deutlich, dass die beteiligten Jugendlichen die Umsetzung dieser Übung sehr unterschiedlich erlebt haben: Einige eher aktive Internetnutzer*innen, die z.B. auch in der Gaming-Szene aktiv sind, empfanden die geposteten Hate-Kommentare einfach als dumm, peinlich, sinnlos und niveaulos – entsprechend sei es nicht schwer gewesen, darauf zu reagieren. Im normalen Alltag würde es sich aber nicht lohnen, hier zu intervenieren, da man solchen Leuten nur noch mehr Aufmerksamkeit schenken würde und es dem betroffenen Opfer nichts bringen, sondern im Gegenteil nur noch mehr Feuer schüren würde. Der Versuch einer vernünftigen Klärung sei mit solchen Leuten nicht möglich. Jugendliche mit geringem Bildungshintergrund betonten in diesem Zusammenhang auch, dass man solche Leute prinzipiell nicht ernst nehmen sollte: *"Man braucht einfach nicht den Kommentar einfach von irgendwem von der Straße ernst nehmen weil wer überhaupt sowas kommentiert hat eh im Leben was falsch gemacht"* (G4, männlich, min 39:40). Entsprechend seien Gegenreaktionen einfach nur „Zeitverschwendung“: *„Es ist einfach unnötig, dass du dir den Stress antust“* (G4, männlich, min 50:55).

Andere Jugendliche in der Aktionsforschung meinten wiederum, dass ihnen solche Inhalte nicht jeden Tag begegnen und sie insgesamt wahrnehmen, dass auf öffentlichen Plattformen wie z.B. YouTube diese Art von Übergriffen im Vergleich zu früher deutlich weniger geworden sei. Für diese Jugendlichen war es eher schwer zu reagieren, weil sie nicht wussten, wie sie darauf antworten sollen und sie sich auch nicht unbedingt auf das gleiche Niveau begeben wollten – entsprechend fielen die Antworten eher kurz aus, wie z.B. „ok?“. Gleichzeitig gab es auch Jugendliche – insbesondere mit niedrigem Bildungshintergrund – die auf solche übergriffigen Kommentare sehr emotional reagierten und dies in den Kommentaren allerdings auf gleichem Niveau wie die Hate-Kommentare umsetzen: *"Da muss man Zivilcourage machen, also zurückschimpfen"* (G4, männlich, min 40:59). Es gab aber auch Aktionsforschungsteilnehmer*innen, die es einfach nur lustig fanden, zurück zu beleidigen.

Für Jugendliche, die sich prinzipiell dafür entschieden haben, Online-Übergriffe in ihrem Internet-Alltag zu ignorieren – selbst wenn diese gegen sie selbst gerichtet sind – war die Übung ungewohnt. Da man das Opfer in solchen öffentlichen Kontexten in der Regel nicht kennt, wird das Kontern hier eher als Einmischung empfunden, die man üblicherweise vermeidet: *„Diese Person muss dann halt eigene Freunde haben, mit die er reden kann und das melden und die ihm helfen. Also wenn ich die Person wirklich nicht kenne, würde ich mich nicht so einmischen“* (G4, männlich, min 41:19). Die Nutzung der Meldefunktion wird dann eher favorisiert: *"Das einfachste was man machen könnte ist halt einfach das Kommentar melden"* (G4, männlich, min 42:23). Jugendliche, die sich noch nicht eindeutig entschlossen haben, wie sie prinzipiell mit Hate-Kommentaren umgehen, sehen sich zum Teil in einem Dilemma, da sie einerseits solche negativen Inhalte nicht unkommentiert stehen lassen möchten, andererseits die Sinnhaftigkeit und

Wirkung von Gegenreaktionen in Frage stellen: „*Stehen lassen kann man auch einfach nicht aber dagegen was sagen braucht man auch eigentlich nicht.*“ (G4, männlich, min 54:25)

Counter Speech in öffentlichen Kontexten meint nicht nur aktive, zivilcouragierte Gegenrede, sondern auch ein allgemeines Auftreten gegen negative und hasserfüllte Inhalte im Internet. Häufig sind dazu Videobeiträge von Influencer*innen zu finden (oft auch als Anti-Hate-Kampagnen gedacht). Um die Bandbreite an Counter Speech Möglichkeiten zu erweitern, wurden daher im Rahmen der Aktionsforschung **ausgewählte YouTube und TikTok Videos gezeigt, die verschiedene Zugänge im Umgang mit Hate Speech thematisieren**. Ziel war es, diese Videos hinsichtlich ihrer Strategien und Wirksamkeit zur Mobilisierung mitlesender Peers zu besprechen. Dabei wurden verschiedene Arten von Videos gezeigt und ausführlich diskutiert, z.B. mit humorvoller / sich lustig machender Ausrichtung (z.B. YouTuber*innen lesen Hate-Kommentare als Unterhaltung und kontern auf lustige Art und Weise, Videos im Comedy-Format über Hate-Speech; lustige Memes in Bezug auf Hate-Speech); ernsthafte Thematisierung bzw. sensibilisierendes / emotionales Aufmerksam machen (z.B. Videos mit symbolischen Schlägen ins Gesicht durch Hate-Kommentare; Influencer*innen, die sich als Betroffene traurig/enttäuscht/wütend zeigen); kreative Auseinandersetzung (z.B. Gestaltung eines Songtextes auf Basis von Hate-Kommentaren); konstruktives und moralisierendes Eingehen auf abwertende Kritik/Hass durch Gegendarstellungen (z.B. sachliche Erklärvideos/Gegenargumente gegen rassistische Übergriffe); Aufrufe, gegen Hate Speech vorzugehen (z.B. Kampagnenvideo von mehreren Influencer*innen gegen Online-Hass; private Aufrufe von wenig bekannten Personen). Die Ergebnisse werden nachfolgend kurz zusammengefasst:

Humorvoller / sich lustiger machender Umgang mit Hate Speech

Häufig sind solche Videos so gestaltet, dass Influencer*innen erhaltene Hate-Kommentare vorlesen und dazu öffentlich kommentieren, indem sie sich in der Regel über Hater*innen lustig machten. Die beteiligten Jugendlichen empfinden solche Videos als witzig – auch wenn manche meinen, dass z.B. das Vorlesen von Hate-Kommentaren mit der Zeit auch lächerlich werden kann. Die Aufbereitung solcher Videos ist vordergründig unterhaltsam, also kein Aufruf gegen Hate, sondern Hate-Kommentare werden rein als Content für ein gutes Video genutzt. Die Jugendlichen entnehmen solchen Videos keine besondere Botschaft, am ehesten wird für sie vermittelt, dass man sich Hate-Kommentare prinzipiell nicht zu nahe gehen lassen sondern eher mit Humor nehmen soll – Eine „überlegene“ und humorvolle Art und Weise im Umgang mit Hate Kommentaren wird besonders von Jugendlichen mit höherem Bildungshintergrund als ansprechend empfunden. Prinzipiell sind humorvolle und kurzweilige Beiträge z.B. auch in Form von Memes/GIFs für Jugendliche ansprechend, zum Beispiel auch mit Einsatz von Tieren („süßer Hund“) und in Kombination mit spezifischen Botschaften auch als konkrete Counter Speech Interventionen vorstellbar – sofern dies in der Praxis für spezifischen Plattformen überhaupt möglich ist.

Ernsthafter Zugang bzw. Sensibilisierung für das Thema Hate Speech

Jugendliche fühlen sich von ernsthaften Zugängen, wo in Videos z.B. auch die persönliche Betroffenheit aus der Opferperspektive dargestellt wird, nur mäßig angesprochen und würden solche Inhalte z.B. auch nicht an Freund*innen weiterschicken, weil der traurige Inhalt diese nur „runterziehen“ würde. Nach Ansicht der Jugendlichen sind solche Inhalte eher an Hater*innen gerichtet, allerdings wird vermutet, dass sich diese solche Videos ohnehin nicht ansehen. Insgesamt seien solche Videos meist zu ernst und emotional belastend, man würde dadurch auch jetzt nicht stärker gegen Hate Speech aktiv werden.

Kreativer Zugang

Kreative Zugänge im Umgang mit Hate Speech können sehr unterschiedlich gestaltet sein, z.B. als fröhliches Tanz-Video mit Kurzbotschaft auf TikTok, oder die Gestaltung eines Songs dessen Text auf Hate-Kommentaren beruht durch prominente Personen. Jugendliche bewerten solche positiven Zugänge als gut, vor allem, wenn dabei auch humorvolle oder zumindest versöhnliche Komponenten enthalten sind, wo auch aus der Opferperspektive vermittelt wird, dass man sich Hate-Kommentare nicht zu sehr zu Herzen gehen lassen soll. Auf solche Weise wird für Jugendliche in ihrer Rolle als Counter Speaker*innen auch die negative Konnotation von Übergriffen entlastet.

Sachliches, aber auch moralisierendes Eingehen auf Vorurteile und abwertende Kritik

In einem weiteren Videobeispiel stellte ein Mädchen asiatischer Herkunft mit sachlichen Argumenten pointiert dar, warum rassistische Vorurteile nicht zu akzeptieren sind. Die Jugendlichen empfanden dieses Video als sehr gut, informativ und ansprechend. Die guten und sachlich fundierten Argumente würden manche auch dazu anregen, das Video mit anderen zu teilen. Das Video könnte auch potenzielle Counter Speaker*innen unterstützen, da ihnen entsprechende Informationen als Argumentationsgrundlage vermittelt werden: *„Wenn ich ein TikTok Video seh, wo einfach eine Frau sagt, hey stoppt den Hate, aber nicht aufklärt, nicht informiert, nicht irgendwas sagt, welchen Hate überhaupt sondern einfach stoppt Leute zu hate, dann führt das zu nichts finde ich.“* (G3, männlich, min 01:25:41).

Allgemeine Aufrufe gegen Hate Speech vorzugehen

Gezeigt wurde auch ein Kampagnen-Video von mehreren bekannten YouTuber*innen, die dazu aufriefen, dass Hate Speech nicht in Ordnung sei und man dagegen vorgehen solle. Das Video kam vor allem aufgrund der hohen Popularität der teilnehmenden Influencer*innen bei den Jugendlichen gut an. Durch die Teilnahme von mehreren bekannten Personen sei auch die Reichweite größer.

Insgesamt lässt sich aus den Bewertungen der Jugendlichen ableiten, dass humorvolle und positive Zugänge zur Thematik im Vergleich zu ernsthaften und emotional belastenden Beiträgen prinzipiell als angenehmer empfunden werden, allerdings auch wenig hilfreich erscheinen bzw. aus der Sicht Jugendlicher wenig bewegen können. Zudem überlagert der Unterhaltungsfaktor mitunter die Ernsthaftigkeit der Thematik bzw. lässt die Schwere des Themas für manche in einem falschen Licht

erscheinen. Um Jugendliche für dieses Thema zu sensibilisieren, eignen sich daher nach Ansicht der Aktionsforschungs-Teilnehmer*innen eher kurze, emotional berührende Videos von Leuten, die ihre Geschichte erzählen (Storytelling), ähnlich wie das Beispiel von Amanda Todd. Gleichzeitig wird aber auch thematisiert, dass solche Videos insbesondere für jüngere Jugendliche (z.B. für 13/14-Jährige) sehr belastend sein können. Auch kreative Bearbeitungen seine nach Ansicht der Jugendlichen gut geeignet, um zu sensibilisieren, sofern diese professionell ausgearbeitet sind und auch eine versöhnliche und angenehme Komponente zeigen und damit auch unter Freund*innen geteilt werden können.

Besonders gut eignen sich aber auch pointierte, sachliche Aufklärungsvideos mit Fokus auf eine bestimmte Thematik (z.B. Rassismus), die den Jugendlichen die notwendige Argumentationsgrundlage bieten, um als Counter Speaker*innen auch entsprechend argumentieren zu können.

Besonders großen Einfluss auf die Motivation, solche Videos zu dem Thema überhaupt ernsthaft anzusehen, haben aber nach Meinung der Jugendlichen Influencer*innen mit entsprechender Reichweite: Allerdings wird vermutet, dass hier eine eher jüngere Zielgruppe angesprochen wird, z.B. 12-Jährige, die sich aber in diesem Alter noch nicht trauen würden, gegen Hate vorzugehen.

3.2.4 Counter Speech in der Praxis: Vorgehen gegen reale Hate-Kommentare

Um die Ebene der Simulation zu verlassen, versuchten die an der Aktionsforschung teilnehmenden Jugendlichen in einem geschützten Rahmen, als Counter Speaker*innen gegen reale Hate-Kommentare aufzutreten. In einem ersten Versuch wurde dabei mit einem jungen YouTuber kooperiert (zum Zeitpunkt der Durchführung mit geringer Aktivität; das Prozedere der Durchführung wurde vorab ähnlich wie beim größer angelegten Online Feld-Experiment in diesem Bericht genau abgestimmt und eine entsprechende Einverständnis des YouTubers eingeholt – im Zuge der Umsetzung kam es zu keinen Interaktionen mit realen User*innen) und ein **Online Feld-Experiment** in einem kleinem Rahmen umgesetzt: Dabei inszenierte das Forschungsteam (bzw. eine junge, sehr internetaffine Mitarbeiter*in) gezielt Online-Übergriffe am YouTube-Kanal, auf die die vorerst nicht eingeweihten jugendlichen Aktionsforschungsteilnehmer*innen reagieren sollten. Um eine möglichst hohe Authentizität zu gewährleisten, wurden von drei verschiedenen fiktiven Profilen Hasskommentare zeitverteilt geschrieben. Auf diese Weise war es auch möglich, unterschiedliche Arten von Hate Speech zu posten, z.B. abwertende Kritik am Beitrag des YouTubers, direkter Angriff auf die Person des YouTubers, oder auch rassistische Äußerungen, z.B. bezogen auf eine bestimmte Personengruppe, die im Video thematisiert wurde.

Den Jugendlichen selbst wurde ein vom Forschungsteam vorbereiteter Fake-Account zur Verfügung gestellt, um nicht mit den persönlichen Accounts intervenieren zu müssen. Zudem sollte es damit gelingen, eine erste Hemmschwelle zu überwinden, da die wahre Identität nicht preisgegeben wurde. Die

Jugendlichen wurden außerdem vom Forschungsteam entsprechend vorbereitet, indem gezielt mögliche Counter Speech Strategien (z.B. Konfrontation mit Fakten, Rückfragen stellen, klare Benennung von als Hater*in, Nutzung von Hashtags, Humor, Mobilisierung anderer zur Unterstützung, Einsatz visueller Elemente wie Emojis) diskutiert wurden. Die konkrete Umsetzung gestaltete sich so, dass während der gemeinsamen Aktionsforschung ein erster Übergriff inszeniert wurde und noch während der Aktionsforschung die Jugendlichen mit ihren Fake-Accounts reagieren mussten. Sie wurden dabei vom Forschungsteam begleitet und dazu angeregt, zielführende Strategien zu entwickeln, wobei es immer wieder Inputs durch das Forschungsteam gab, aber dennoch der jugendgerechten Umsetzung bzw. individuellen Gestaltung von Counter Speech Strategien ausreichend Raum geboten wurde. Auch die Möglichkeit des gemeinsamen Vorgehens bzw. der gegenseitigen Unterstützung wurde bewusst thematisiert und auf mögliche weitere Reaktionen der Community außerhalb der Aktionsforschungsgruppe vorbereitet.

Bei den Gruppen 3 und 4 konnte die Kooperation mit dem YouTuber nicht mehr umgesetzt werden, da dieser nicht mehr aktiv war. Daher wurde das Prozedere so modifiziert, dass zwar die Fake-Profile genutzt wurden, aber **gegen reale Hate-Kommentare vorgegangen** wurde. Dabei sollten die Jugendlichen selbst Vorschläge für gefundene Hate-Kommentare einbringen (als Aufgabe seit dem letzten Treffen), von denen sie dachten, dass Counter Speech angebracht wäre. Das Forschungsteam hat außerdem zusätzlich selbst nach entsprechenden TikTok oder YouTube-Videos mit Hate-Kommentaren recherchiert für den Fall, dass die Vorschläge der Jugendlichen sich als unpassend erweisen würden (z.B. recherchierten die Jugendlichen häufig mit dem Begriff Hate Speech und stießen damit etwa auf Reaction-Videos der betroffenen Personen – häufig ohne konkrete Hasskommentare). Nach gemeinsamer Entscheidung, gegen welche Hate-Kommentare vorgegangen werden sollte, war das weitere Prozedere bei den Gruppen 3 und 4 aber mit jenem der Gruppen 1 und 2 ident.

Durch die Möglichkeit, die zur Verfügung gestellten Face-Accounts zu nutzen, fühlten sich die meisten Jugendlichen in ihrem Vorgehen gut geschützt. Für manche war es daher die erste Erfahrung im Vorgehen gegen (vermeintlich) reale Hate-Kommentare.

In den folgenden Ausführungen werden die Erfahrungen der Jugendlichen dargestellt, zunächst für jene beiden Gruppen (Gruppen 1 und 2), die gegen die vom Forschungsteam inszenierten Übergriffe vorgehen (Online Feld-Experiment: für die Jugendlichen waren diese jedoch zum Zeitpunkt des Vorgehens vermeintlich reale Kommentare). Dabei werden vor allem die ersten Reaktionen genauer in den Blick genommen, die den weiteren Interaktionsverlauf zwischen Hate- und Counter Speaker*innen wesentlich mitbestimmen. Die folgenden Ausführungen zeigen beispielhaft typische Interaktionsverläufe der einzelnen Gruppen.

Anmerkung: Die auf den nachfolgenden Screenshots anführten Nicknames der Jugendlichen sind fiktiv bzw. jene der eingesetzten Fakeprofile.

Gruppe 1:

Beispiel eines inszenierten Hate-Kommentars:

Bei diesem Hate-Kommentar, dass auch als erster Übergriff inszeniert wurde (es folgten noch weitere auf anderen Videos mit unterschiedlichem, fiktivem Hate-Profil), bezieht sich auf den Inhalt eines Video, das vom YouTuber veröffentlicht wurde und wo von einer Demonstration berichtet wurde.



Abbildung 30

Erste spontane Reaktion der Jugendlichen aus der Gruppe 1 verwiesen zunächst gleich darauf, dass man auf solche Kommentare normalerweise nicht reagiert und begründeten dies damit, dass man solche Kommentare in der Regel nicht ernst nimmt: „Das ist eigentlich so ein Kommentar, auf das ich normalerweise nichts antworten würde. Ich mein es ist zwar nicht ok, dass sie ihn aus dem nichts heraus beschimpft, aber es ist eigentlich nicht viel dahinter.“ (G1, männlich, min 47:30).

Nach der konkreten Aufforderung der Forscher*innen, trotzdem dagegen zu reagieren, musste zunächst geklärt werden, ob dieser Übergriff nicht begründet sei, ob es vielleicht also tatsächlich nicht erlaubt sei, Gesichter auf Demos zu filmen, erst danach folgte eine erste Gegenreaktion aus der Gruppe:

Erstes Counter Speech Kommentar:



Abbildung 31

Es wurde also versucht, einen ersten Schwachpunkt zu finden, der in möglichen Rechtschreibmängeln gefunden wurde. Die vermeintliche Haterin reagierte daraufhin mit folgendem Kommentar:



Abbildung 32

In der Gruppe folgte schließlich eine Diskussion, wie nun eine gelungene Counter Speech Reaktion aussehen sollte: z.B. nochmals zu erklären, dass man sich auf die Rechtschreibung beziehe, oder zu fragen: warum beleidigst du? Schließlich wurde aber mit folgendem Posting reagiert:



Abbildung 33

Mit dieser schlagfertigen Reaktion gepaart mit einer Zurechtweisung wurde eine gewisse Überlegenheit demonstriert, die in den anschließenden Kommentaren auch anderer Jugendlicher aus der gleichen Gruppe auf ähnlich Weise aufgegriffen wurden bzw. war für Folgeaktionen anderer Counter Speaker*innen der „Bann gebrochen“, ebenfalls aktiv zu werden, indem z.B. die vermeintliche Hater*in zurechtgewiesen wurde, dass man auch höflicher formulieren könne, um hinzuweisen, dass auf Demonstrationen keine Gesichter gefilmt werden sollten.

Nach diesen ersten gelungenen Counter Speech Reaktionen kam es allerdings auch zu einer ersten Irritation, da zunächst eine entsprechende Folgeaktion der vermeintlichen Hater*in ausblieb: „*Entweder sie ist gerade nicht am Handy oder sie ignoriert uns.*“ (G1, min 49:50). Es wurde auch überlegt, ob es vielleicht merkwürdig erschienen würde, wenn immer die gleichen Personen gegen die Haterin vorgehen: „*Vielleicht bemerkt sie dann, dass da irgendwas faul ist, wenn immer die gleichen Leute sie anschreiben? ... Ich glaube, wir sollten da zwischendurch warten.*“ (G1, weiblich, min 48:25).

Die weitere Entwicklung (die Jugendlichen kommentierten auch außerhalb der Aktionsforschung über einen Zeitraum von einer Woche) verlief nach weiteren Attacken der vermeintlichen Haterin sehr ähnlich: Die Strategie der Counter Speaker*innen war es, gekonnt schlagfertig zu reagieren und dabei klug, überlegen, humorvoll und zugleich höflich zu wirken. Allerdings verlangt das relativ hohe Kompetenzen im Counter Speech Verhalten, wie folgendes Beispiel als Reaktion auf ein weiteres Hate-Kommentar zeigt, bei dem gefragt wurde, ob diese nicht besseres zu tun habe als zu kommentieren:



Abbildung 34

Neben diesen schlagfertigen Reaktionen waren in dieser Gruppe noch weitere Strategien zu finden, wie Nicht lockerlassen“ („*Du hast meine Frage nicht beantwortet*“), um eine weitere Reaktion bzw.

Rechtfertigung von Hater*innen einzufordern, nicht-aggressive Zurechtweisung mit moralisierendem Grundton („*Warum verwendets du h00rensohn um deine Meinung auszurücken?* Es gibt doch genügend Wörter im Wörterbuch ...“) oder Absprechen verbaler Fähigkeiten („*Wenn du nicht argumentieren kannst, dann las es*“). In allen Strategien wird aber sichtbar, dass der Versuch im Vordergrund steht, die eigenen intellektuellen Fähigkeiten als überlegen darstellen zu wollen. Inhaltlich war der Fokus darauf ausgerichtet, dass Kritik zwar zulässig sei, allerdings ohne Beleidigungen und Schimpfwörter.

Die Idee einer gegenseitigen Unterstützung wurde auch durch die technischen Gegebenheiten erschwert: Es dauerte z.B. unerwartet lange, bis die Kommentare tatsächlich öffentlich zu sehen waren bzw. waren bei der üblichen Einstellung „Top-Kommentare zuerst“ die Kommentare nur schwer zu finden und es musste auf „Neueste Kommentare zuerst“ umgestellt werden. Dieser Umstand machte sowohl das eigene Kommentieren als auch das Unterstützen anderer Kommentare mühsam.

In der Abschlussreflexion eine Woche später hatten die teilnehmenden Jugendlichen Gelegenheit, über ihre Erfahrungen zu berichten: Die selbst gewählte Strategie, die eigene Überlegenheit auf sprachlicher Ebene auszuspielen, wurde zum Teil als anstrengend und belastend erlebt, vor allem aus Angst, durch einen Fehler sofort selbst angegriffen zu werden: „*Ich fand es interessant, irgendwie wollt ich die Person provozieren in einer höflichen Art und Weise, aber man muss auch aufpassen, welche Wörter man verwendet, also man durfte jetzt keinen Satz falsch schreiben, weil dann hätte dich die Person attackieren können.*“ (G1, weiblich, min 08:55). Die Jugendlichen thematisierten aber auch, dass sie zum Teil auch ratlos waren, wie sie reagieren sollten, und dass die Entwicklung passender Counter Speech Kommentare auch manchmal schwierig war, weil man üblicherweise auf solche Inhalte nicht reagieren würde. Besonders herausfordernd war es für sie Hate-Kommentare zu kontern, die diese als „dumm“ wahrnahmen: „*Ich weiß nicht, was kann man auf geh raus aus Österreich antworten? Das ist einfach nur blöd.*“ (G1, weiblich, min 10:52). Üblicherweise werden solche Hate-Kommentare daher auch ignoriert, da eine Reaktion aus Sicht der beteiligten Jugendlichen schlicht unnötig sei.

Auffallend war auch, dass die Jugendlichen trotz ihrer Bemühungen nicht davon überzeugt waren, dass sich ihr Einsatz gelohnt hätte: „*Weil – es hat nicht wirklich einen Effekt. Leute haben einfach, weil sie keine Hobbies haben, keine Ahnung, ich weiß es nicht. Aber sie werden dadurch, dass ich ihnen was sage, glaube ich nicht das lassen ... Das war einfach nur sinnloser Hate, wo es für mich keinen Sinn hätte weil ich weiß, die Leute wollen nichts dazulernen.*“ (G1, männlich, min 09:53). Die gesamte Gruppe hatte nicht das Gefühl, mit ihrem Counter Speech Vorgehen erfolgreich gewesen zu sein, sondern sah lediglich einen gegenseitigen Schlagabtausch: „*Die eine Seite hat die andere Seite attackiert*“ (G1, weiblich, min 12:22). Vermisst wurde ein Zeichen der Einsicht oder des Nachdenkens. Wenn Hater*innen nicht weiter darauf reagieren, ist das für die Jugendlichen nicht unbedingt ein Zeichen des eigenen Erfolgs: „*Naja, manchmal denkt man sich auch, die hat einfach keinen Bock mehr mit mir zu diskutieren und ignoriert mich einfach*“ (G1, weiblich, min 14:25), andere wiederum sahen darin dennoch auch einen persönlichen Sieg. Insgesamt wurde es auch eher als langweilig angesehen, wenn man keine Reaktionen erhält.

Gruppe 2:

Beispiel eines inszenierten Hate-Kommentars:

Bei diesem Hate-Kommentar wurde auf einen Videobeitrag des YouTubers zur beginnenden Corona-Pandemie Bezug genommen und damit eine Personengruppe abgewertet.



Abbildung 35

Im Vergleich zur Gruppe 1 gingen die ersten Versuche des Kommentierens sehr leise vor sich, es gab kaum Austausch, die Jugendlichen überlegten nicht gemeinsam, sondern gingen jeweils eher selbstständig bei den unterschiedlichen inszenierten Hate-Kommentaren vor.

Erstes Counter Speech Kommentar:



Abbildung 36

Auch hier wird versucht inhaltlich klug mit Wissen und Schlagfertigkeit zu punkten, was sich auch im weiteren Interaktionsverlauf bzw. auch bei der Reaktion auf andere Hate-Kommentaren immer wieder zeigt.



Abbildung 37

Es findet sich aber auch eine breite Palette mit weiteren Zugängen, wie Verständnis zeigen und sachlich argumentieren,

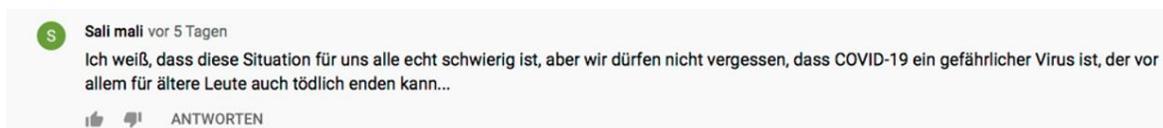


Abbildung 38

Unkenntnis unterstellen,



Abbildung 39

Zurechtweisungen mit moralischem Unterton,



Abbildung 40

Aufrufe zur Deeskalation,



Abbildung 41

oder mit Sarkasmus argumentieren

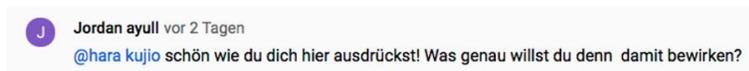


Abbildung 42

Dazwischen sind aber auch angriffige („kümmere dich lieber um deine eigenen Sachen“, „und du bist so perfekt oder was?“, „unnötiges Kindergarten Niveau ...“) und auch Gegenangriffe zu finden, insbesondere dann, wenn Hate-Kommentare zunehmend übergriff werden,



Abbildung 43

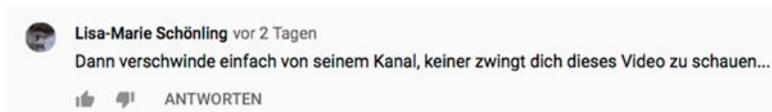


Abbildung 44

beziehungsweise werden Kommentare auch zunehmend emotionaler.



Abbildung 45

Die Gruppe zeigt also prinzipiell ein breiteres Spektrum möglicher Counter Speech Reaktionen als Gruppe 1, was wahrscheinlich auch am individuell ausgerichteten Vorgehen liegt. Die Gruppe war während des vereinbarten Zeitraums von einer Woche zudem besonders aktiv. Auffallend war, dass auch hier zunächst vor allem auf Strategien wie Schlagfertigkeit oder Demonstration der intellektuellen Überlegenheit gesetzt wurde, die Tonalität aber mit der Zeit zunehmend angriffiger wurde – was vermutlich auch mit der Langatmigkeit der Interaktionen zu tun hatte, die scheinbar zu keinem wirklichen Ende führten und die Counter Speaker*innen zunehmend frustrierten. Zusätzlich orientierte sich die Tonalität aber auch am Hate-Kommentar selbst.

Auch in der nachfolgenden Diskussion bei Abschluss des Feldversuchs vermittelten die Jugendlichen, dass es nicht einfach war, motiviert zu bleiben bzw. wirkte die Gruppe insgesamt resignierend: *„Die waren schon ziemlich hartnäckig und die hat man irgendwie nicht klein bekommen“* (G2, weiblich, min 02:01). Auch in dieser Gruppe herrschte das Gefühl, mit Counter Speech im Grunde nichts erreichen zu können: *„Es hat sich nicht so angefühlt, als hätte er irgendwas akzeptiert, was wir geschrieben haben“* (G2, männlich, min 32:08). Entsprechend gab es nach Ansicht der Jugendlichen auch keine Strategie, die tatsächlich Wirkung gezeigt hätte – außer das Potenzial, gemeinsam aufzutreten: *„Es ist schon so, wenn man zu fünf auf einen losgeht dann muss der sich schon was dabei denken“* (G2, männlich, min 37:05).

Um irgendein Gefühl von Erfolg zu haben, hätten sich die Jugendlichen zumindest einen Schlusspunkt gewünscht, vor allem bei einer bereits länger andauernden Konversation. Das wäre auch in einer ganz einfachen Form vorstellbar, wie z.B. ja egal, chillt mal. Eine Entschuldigung würde man sich ohnehin nicht erwarten. Ansonsten wäre ein Erfolg vielleicht daran erkennbar, wenn Hater*innen in ihrer Kommunikation zunehmend in die Ecke getrieben werden: *„Wenn man merkt, dass er jetzt einfach nichts mehr sagen kann, weil er einfach merkt, dass er es selber verkackt hat und deshalb immer auf so Kleinigkeiten rumreitet die eigentlich schon egal sind.“* (G2, weiblich, min 41:50).

Auffallend war in dieser Gruppe auch, dass viele Ausdrücke, die in den Hate-Kommentaren vorkamen, von den Jugendlichen gegoogelt werden mussten und ihnen als durchschnittliche Internet-Nutzer*innen nicht geläufig waren. Sind Begrifflichkeiten nicht bekannt oder werden diese in falschen Kontexten verwendet, läuft man Gefahr, bloßgestellt zu werden. So wurde z.B. der Begriff „simp“ in einem Hate-Kommentar verwendet, ein Begriff, der vor allem von Männern mit misogynem Gedankengut verwendet wird, um Männer, die sich freundlich gegenüber Frauen verhalten, als charakterlos, unterwürfig und schwach zu bezeichnen. In der entsprechenden Gegenreaktion einer jugendlichen Counter Speakerin wird dieser Begriff aufgegriffen, aber unpassend eingesetzt, wie folgendes Beispiel zeigt:



Abbildung 46



Abbildung 47

Gruppe 3:

In dieser Gruppe – ebenso für Gruppe 4 – wurde nicht gegen inszenierte, sondern gegen reale Hate-Kommentare vorgegangen. Die Jugendlichen sollten dafür Beispiele suchen, gegen die dann mit Hilfe der Fake-Profilen vorgegangen werden konnte. Den Jugendlichen fiel diese Aufgabe schwerer als erwartet, da sie im Alltag eher zufällig auf Hate-Kommentare stoßen und diese in der Regel ignorieren. Bei der gezielten Suche fanden sich zudem einige Beispiele, die z.B. als Reaktionsvideos von Influencer*innen selbst gestaltet waren. Bei anderen wiederum war der ursprüngliche Hate-Kommentar nicht mehr sichtbar, da er schon gelöscht wurde (z.B. von Plattformbetreiber*innen) und ließ sich lediglich aufgrund von Reaktionen anderer erschließen.

Für das konkrete Vorgehen wurden schließlich zwei Videos ausgewählt:

- 1) Tiktok-Video 1: Hate-Kommentare bei einer TikTokerin, die „*Sachen gepostet hat, die nicht so gut von der Gesellschaft aufgenommen werden*“ (G3, weiblich, 09:50). Ihr wird in entsprechenden Kommentaren indirekt der Tod gewünscht.
- 2) Tiktok-Video 2: Hate-Kommentare bei einer beeinträchtigt wirkenden Person in einem TikTok-Video, die Cosplay macht (= das Sichverkleiden als eine Figur z.B. aus einem Manga, häufig als Wettbewerb um die beste Darstellung): „*Ich hab halt eben wieder in die Kommentare geschaut und die haben ihn halt alle so ziemlich verarscht. Von ‚auf Wish* bestellt‘ bis zu ‚auf Sonderschulkiosk‘ und da hab ich mir auch gedacht das ist nicht ganz so okay und deswegen hab ich’s dann in die Gruppe geschickt. Aber ich wüsste jetzt persönlich echt nicht, wie ich dagegen vorgehen kann. Also was ich jetzt da drauf schreiben könnte ... weil ich mir nicht sicher bin, ob ich auf die Hate-Kommentare eingehen soll oder ob ich ihm quasi was Positives sagen soll*“ (G3, männlich, 13:50).

* Wish: Online-Shopping-App mit besonders billigen Preisen

Auch in dieser Gruppe wurden die ersten Versuche, mit den Fake-Profilen dagegen vorzugehen, eher individuell gestartet.

Die folgenden Ausführungen zeigen wieder, welche Strategien die Jugendlichen hier eingesetzt haben:

Zu Video 1

Benennung des Übergriffs als unmoralisches klares „No Go“

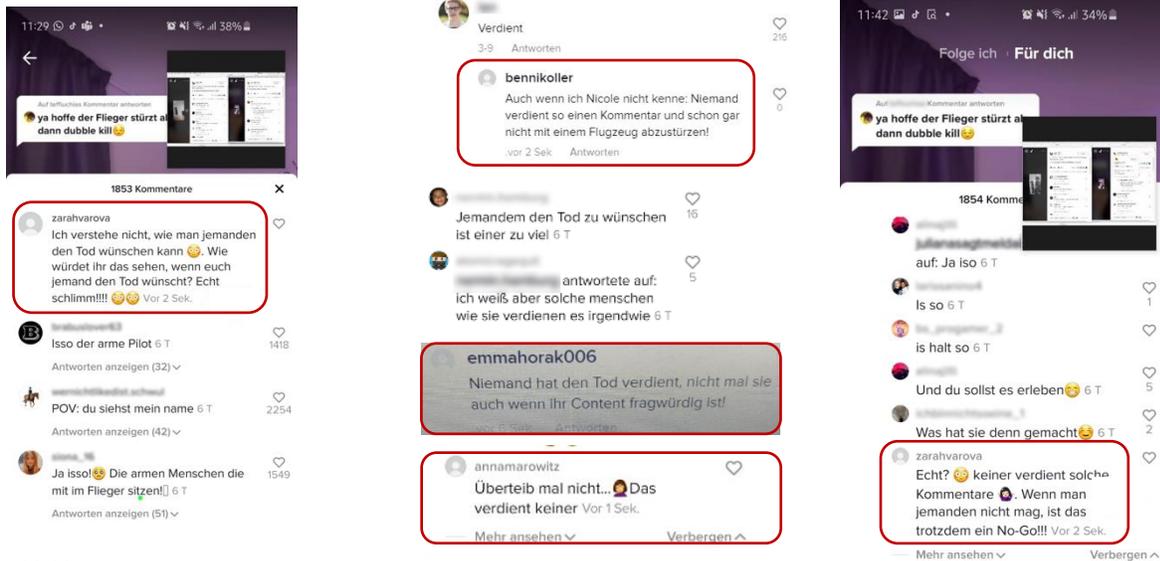


Abbildung 48

Retourkutsche – mit Bezug auf „No-Go“



Abbildung 49

Die klare Benennung als No-Go Hate Kommentar wird zum Teil noch mit anderen Strategien verknüpft, wie z.B. mit der Aufforderung, einfach zu entfolgen („... aber dann folgt ihr doch einfach nicht? Jemanden den Tod zu wünschen geht zu weit“), oder es wird ein schlechtes Gewissen gemacht („Würdest du es feiern wenn dir jemand den Tod wünscht?“).

Jemandem den Tod zu wünschen, ist für die beteiligten Jugendlichen offenbar moralisch nicht vertretbar und ein absolutes „No-Go“. Entsprechend wird auch in allen Counter Speech Strategien darauf Bezug genommen. Aus den Diskussionen wird klar, dass eine Verteidigung des Opfers nur möglich war, da der Opferstatus ohne diesem No-Go-Hate Hate Kommentar sonst nicht klar genug gewesen wäre, da die

angefeindete Person offenbar auch selbst Fehlritte gemacht hatte, die nun entsprechend von der Community sanktioniert wurden.

Zu Video 2

Die Jugendlichen reagieren mit einem offensiven Angriff in Form einer Verteidigung des Opfers, wobei viel auf Englisch reagiert wird.

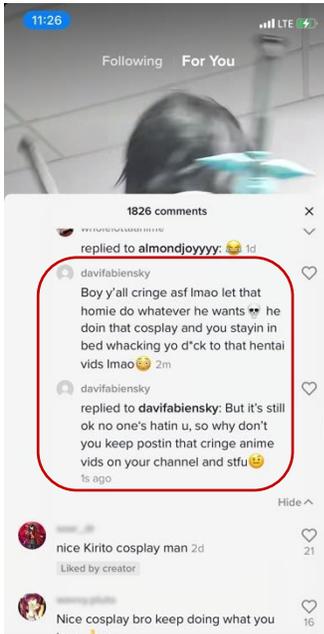


Abbildung 50

Bei vielen Videos, die von Jugendlichen gesehen werden, werden die Kommentare sowohl auf Deutsch als auch auf Englisch bzw. auch ausschließlich auf Englisch gemacht. Counter Speech erfordert hier zusätzliche Kompetenzen, insbesondere auch, um die oft in diesem Zusammenhang vorkommenden Internet-Slang Abkürzungen (z.B. stfu = shut the fuck up; Übersetzung ähnlich wie „halt dein Maul“; gtfo = get the fuck out; Übersetzung ähnliche wie „Verpiss dich!“) richtig einsetzen zu können, die z.B. auch häufig in der Gaming-Szene zu finden sind. Entsprechend sind es eher sehr interneterfahrene Jugendliche mit höherer Bildung, die auf diese Weise kontern. Auf die Bedeutung von Englisch bzw. Fremdsprachen für die Formulierung von Gegenreaktionen wurde bereits an anderer Stelle hingewiesen: die emotionale und kognitive Distanz erleichtert hier, sofern die erforderlichen Sprachkenntnisse gegeben sind, oft das Vorgehen.

Inhaltlich fordern die Kommentare im Grunde zu Toleranz auf bzw. betonen die Selbstbestimmtheit des Opfers mit der klaren Aufforderung, die Übergriff zu stoppen bzw. werden Hater*innen in die Schranken gewiesen.



Abbildung 51

Neben dieser vorrangig klaren Verteidigung des Opfers bei gleichzeitigem Gegenangriff von Hater*innen sind vereinzelt auch moralische Zugänge zu finden, die aber ebenfalls Täter*innen abwerten (z.B. „*Musst du unbedingt andere Leute runtermachen? Du bist nicht lustig mit sowas sondern einfach nur peinlich*“).

Insgesamt nehmen die Jugendlichen hier stark die Position ein, das Opfer zu verteidigen – was entgegen den bisherigen Ausführungen der Jugendlichen sonst eher vermieden wird. Erklärbar wird das dadurch, dass die im TikTok-Video angegriffene Person durch ihre Beeinträchtigung auch klar als hilfsbedürftiges Opfer wahrgenommen wird, das unverschuldet gehated wird. Damit sind einerseits ausreichend Kontextinformationen vorhanden, um die Situation richtig einschätzen zu können, andererseits kann das Opfer nicht mit durchschnittlichen User*innen verglichen werden, von denen in der Regel erwartet wird, dass sie selbst intervenieren.

In der anschließenden Diskussion eine Woche später schilderten die beteiligten Jugendlichen, dass das Vorgehen gegen Hate-Kommentare eher eine ungewohnte und zum Teil „komische“ Erfahrung war, die meisten von ihnen üblicherweise darauf nicht reagieren würden und sie anfangs nicht wussten, was sie schreiben sollen, sie es am Ende aber ganz lustig fanden. Vereinzelt wurde auch versucht, die Kommentare untereinander zu unterstützen (mit Likes), wobei hier vor allem das Forschungsteam dazu angeregt hat. Die Jugendlichen waren allerdings nicht überzeugt davon, ob eine strategische gegenseitige Unterstützung etwas bewirken kann bzw. wurde das Ausmaß an Unterstützung in der Regel als zu gering erachtet: „*Naja man wird erst was erreichen, wenn es so 100 Likes oder ein paar so Recommends hat oder so, weil es bringt halt nichts wenn es einer liked und ich glaub, keine Ahnung wenn's vier nochmal kommentieren, bringt halt nichts. Es liest halt auch keiner.*“ (G3, männlich, 01:17:57). Einige haben dennoch ihre eigenen Kommentare geliked, damit es etwas „cooler“ aussieht.

Die Möglichkeit, gezielt nach Unterstützung aufzurufen, wurde bisher von den beteiligten Jugendlichen noch nicht angedacht. Ein Teilnehmer der Aktionsforschung meinte, er könnte sich vorstellen, das auf diese Weise zu formulieren – hat aber selbst keine konkreten Erfahrungen dazu:



Abbildung 52

In der Gruppe wurde auch erwähnt, dass bezogen auf die betroffenen Personen auch nach Hintergrundinformationen auf den betreffenden Profilen gesucht wurde, sowohl bei Hater*innen – hier um nach möglichen Angriffspunkten zu suchen (z.B. wurde dann im Konter-Kommentar aufgenommen), als auch beim Opfer selbst: „*Ich muss auch nochmal sagen, bei dem Kiriko Cosplay ich glaube man muss*

sich auch ein bisschen mehr mit der Person die das Video gemacht hat beschäftigen, damit man sie gut verteidigen kann. Ich hab dann irgendwo mal gelesen, dass das ein Krebs-Patient ist zum Beispiel und wenn man halt solche Details weiß, kann man die Person viel besser verteidigen.“ (G3, männlich, 01:22:14). Kontextinformationen sind also wichtig, um eigene Interventionen als Counter Speaker*in zu begründen und für sich selbst zu rechtfertigen.

Aber auch diese Gruppe würde ihren Counter Speech Einsatz nicht unbedingt als erfolgreich bewerten: Erfolgreich wäre es für sie gewesen, wenn zumindest auch fremde Personen Counter Speech Kommentare geliked oder darauf reagiert hätten. Da das aber nicht passiert sei, wurde das Vorgehen insgesamt als eher demotivierend empfunden. Kleinere Erfolgsmomente stellten sich nur dann ein, wenn man mit der eigenen Antwort zufrieden war. Grundsätzlich wären die Jugendlichen ohne den expliziten Auftrag durch die Teilnahme an der Aktionsforschung nicht eigenständig gegen die Hate-Kommentare vorgegangen, da sie das Gefühl hatten, dass es ohnehin nichts bringen würde. Daher würde man das eher ignorieren – außer vielleicht bei Personen, die einem besonders wichtig oder nahe seien.

Gruppe 4

Auch in dieser Gruppe wurden die Jugendlichen gebeten, nach Hate-Kommentaren zu suchen, um anschließend dagegen vorzugehen. Den Jugendlichen fiel es aber offenbar schwer, passende Beispiele zu finden, da sie auf ihrer ForYou-Page nur Inhalte vorgeschlagen bekommen würden, wo es wenig Hate gibt – man müsse dann gezielt z.B. TikToks mit bestimmten Inhalten (z.B. zum Thema Homophobie) anschauen. Andere Jugendliche argumentierten damit, dass sie selbst wenig in die Kommentare schauen, sondern nur das Video ansehen würden. Insgesamt wird aber vermutet, dass aktuell die meisten Hasskommentare auf TikTok zu finden wären, wenn diese auch seit der Einführung von Blacklists weniger geworden zu sein scheinen. Wichtig war auch, die Wahrnehmung des Opfers selbst, das im Auftreten auch einen klaren Opferstatus erkennen lassen muss, um auf entsprechende Verteidigung zu hoffen: *„Es ist halt so ne Sache was man ins Internet reinstellt ... zum Beispiel beim Mädchen vorher, da hätte ich nichts drauf geschrieben, weil ich finde sowas einfach nur cringe ... das Benehmen, Füße rauf, das Lied. Würde da keine Hasskommentare schreiben aber sie auch nicht verteidigen“* (G3, weiblich, 57:47).

Ausgewählt wurde schließlich ein TikTok-Video eines jungen Mannes, der für sein Aussehen (z.B. Frisur) gehated wurde. Die betroffene Person war auch in früheren Videos aufgrund seines Auftretens (z.B. öfter mit nacktem Oberkörper) angefeindet worden.

Im konkreten Counter Speech Vorgehen wurden von den Jugendlichen folgende Strategien umgesetzt:

Die Jugendlichen reagieren hier damit, den Angriff zu entkräften Gleichgültigkeit zu signalisieren, dass Verhalten zu hinterfragen und zum Aufhören aufzufordern, wobei alle Kommentare relativ kurz ausfallen.

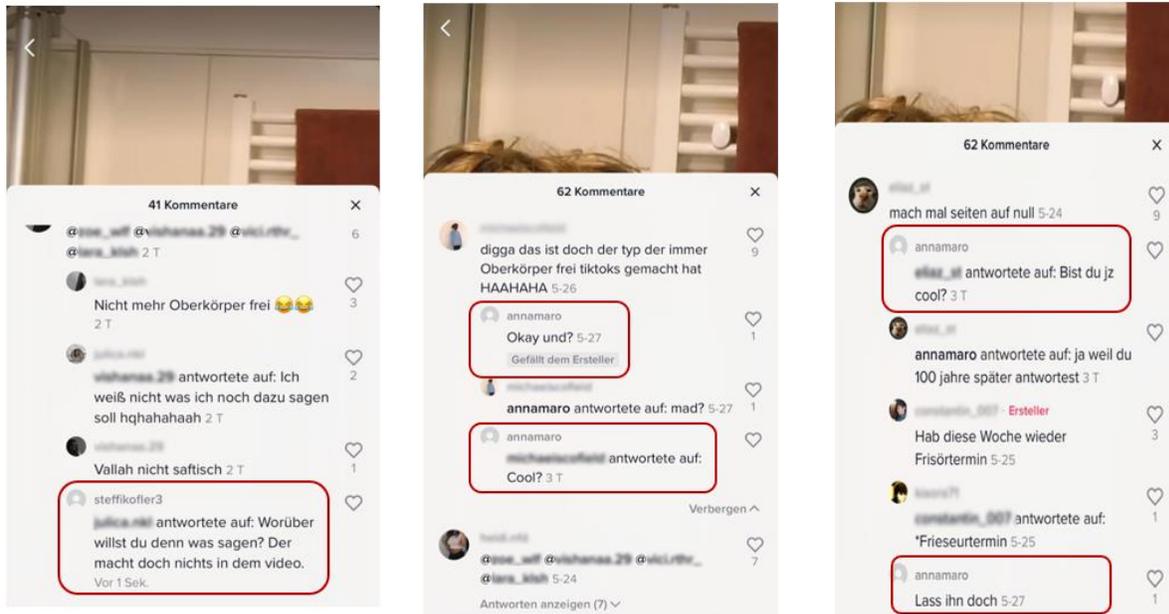


Abbildung 53

In dieser Gruppe war die Aktivität der Jugendlichen in ihrer Rolle als Counter Speaker*innen deutlich geringer ausgeprägt als bei den anderen Gruppen. Es wurde nicht nur weniger häufig, sondern auch deutlich kürzer interveniert, und die Begründungen waren kaum ausformuliert. Insgesamt zeigt sich, dass in dieser Gruppe die vorhandenen Counter Speech Kompetenzen am wenigsten ausgeprägt waren bzw. auch die Sensibilisierung in der Wahrnehmung von Online-Übergriffen geringer zu sein schien.

In der nachfolgenden Diskussion am Ende der Aktion meinten die einige Teilnehmer*innen, sie hätten entweder nichts gefunden, wogegen sie reagieren hätten können, oder sie hätten nicht gewusst, was sie kommentieren sollen bzw. hatten keine Lust, „herumzudiskutieren“. Hätten sie reagiert, dann am ehesten mit z.B. *"Lass ihn, er is so"* weil: *„er ist ein normaler Mensch, er kann machen was er will“* (G3, männlich, 01:11:15). Argumentiert wird auch, dass so etwas einfach auch keine „gescheite Diskussion“ sei, bei der man reagieren könnte. Eine Teilnehmerin hat außerdem angemerkt, dass sie auf TikTok bei ihren Antworten vom System nachgefragt wurde, ob sie das tatsächlich senden will – vermutlich aufgrund der Verwendung von Schimpfwörtern.

Erfolg im Counter Speech Vorgehen wird auch in dieser Gruppe nicht konzediert, angemerkt wird aber, dass es mehr Sinn machen würde, wenn mehrere dagegen vorgehen – auch wenn das in der Gruppe nicht praktiziert wurde: *„Ich glaub es wär auch gescheiter wenn mehrere [dagegen kommentieren würden] weil Aussage gegen Aussage bringt sich ja wieder nichts.“* (G3, weiblich, 01:29:08).

Während der Diskussion wurde aber auch folgendes Beispiel von einer Teilnehmerin gezeigt, ein Reaktionsvideo eines bekannten TikTokers, das sich auf ein beeinträchtigtes Mädchen bezieht, das sehr viele Hate-Kommentare erhält. Das offensive Anprangern durch eine prominente Person dieses

Fehlverhalten („Das ist nicht euer Ernst! Schämt euch in Grund und Boden!“) wurde von der Community offenbar sofort aufgegriffen und unterstützt.

Reaktionsvideo auf TikTok

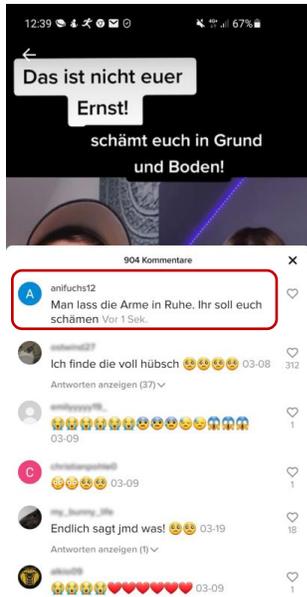


Abbildung 54: Video, <https://vm.tiktok.com/ZMdd3nMoT/> (Letztzugriff: 23.09.2021)

TikTok Video des Opfers



Abbildung 55: Video, <https://vm.tiktok.com/ZMddTRqeV/> (Letztzugriff: 23.09.2021)

Für die beteiligten Jugendlichen hat ein solches ein Vorgehen grundsätzlich Potenzial, andere zu mobilisieren, zukünftig aktiver gegen solche negativen Inhalte aufzutreten – das sei aber nicht vorhersagbar. Einige meinten auch, dass es mehr bringen würde, am Profil von Hater*innen nach Schwachstellen oder besonderen Informationen zu suchen und dann darauf bezogen Gegenangriffe zu machen: Den*die Hater*in zu haten sei legitim - dem wurde allerdings auch von anderen widersprochen, da es Ziel sein müsse, gar nicht mehr zu haten.

Insgesamt zeigt sich, dass die vorhanden Counter Speech Möglichkeiten aus der Sicht der beteiligten Jugendlichen in der Praxis deutlich eingeschränkter umgesetzt werden können und der Fokus auf einer gelungenen, vorrangig textbasierten Vorgehensweise liegt, die auch entsprechendes Argumentationsvermögen und Schlagfertigkeit verlangt.

3.2.5 Wahrnehmung und Bewertung des Erfolgs von Counter Speech

Im Zuge der gesamten Aktionsforschung wurde mehrmals deutlich, dass Jugendliche ihr Counter Speech Vorgehen als nicht oder nur wenig erfolgreich einschätzten. Um dennoch mögliche Erfolgs-Determinanten gemeinsam mit Jugendlichen zu erarbeiten, wurde eine Übung in Form eines Quiz umgesetzt: Als Ausgangspunkt wurde das Saferinternet-Quiz, das gezielt zum Thema Online-Zivilcourage entwickelt wurde (<https://app.involve.me/siat/zivilcourage-quiz>), herangezogen. Die Jugendlichen sollten zunächst die Quizfragen selbständig beantworten, anschließend wurden diese gemeinsam reflektiert. Als konkrete Übung mussten die Jugendlichen Antwortoptionen für folgende neue Quizfrage entwickeln:

„Du hast dir vorgenommen, gegen Hate-Kommentare vorzugehen. Nach einiger Zeit denkst du: „Ich hab keine Ahnung, ob das überhaupt was bringt. Woran merke ich eigentlich, dass ich erfolgreich bin? 🤔“.

Aufgabe war es, jeweils zwei falsche und zwei richtige Antworten zu überlegen, die man anderen Jugendlichen stellen könnte und diese dann im MS-Teams Chat zu posten. Die verschiedenen Vorschläge wurden anschließend wieder diskutiert.

Von den Jugendlichen kamen zusammenfassend folgende Antwortvorschläge, die diese als „richtige“ Antwort titulierte und aus Sicht von Jugendlichen als Zeichen eines Erfolgs von Counter Speech interpretierbar sind:

- Einsicht von Hater*innen: wobei konkrete Entschuldigungen nicht zu erwarten wären
- keine weiteren Folgeaktionen von Hater*innen: dem wird in der Diskussion zum Teil widersprochen
- allgemein positive Reaktionen und Unterstützung durch andere Peers, z.B. auch Likes oder auch Danke!
- prinzipiell Reaktionen durch andere, unabhängig ob positiv oder negativ
- Erfolgsbewertung anhand von Likes und Unterstützungskommentaren über einen längeren Zeitraum; es wird damit argumentiert, dass es manchmal sein kann, dass Unterstützung kommt, ein anderes Mal nicht, daher müsse eine längere Zeitperiode in Betracht gezogen werden
- Entwicklung eines eigenen gutes Gefühls / Gewissens
- Mithilfe zum Erfolg anderer: wenn man selber andere Counter Speaker*innen unterstützt

Insgesamt wird deutlich, dass sich Jugendliche vor allem grundsätzliche Reaktionen auf ihr Counter Speech Verhalten wünschen – allerdings ist das schwer steuerbar. Daher bleibt in der Regel übrig, durch den eigenen Einsatz ein gutes Gewissen zu erzeugen oder zumindest selbst Vorbildwirkung für andere zu sein, indem aktiv auch andere Counter Speaker*innen unterstützt werden.

Die aus der Sicht Jugendlicher „falschen“ Antwortoptionen lassen sich wie folgt zusammenfassen und verweisen auf den Misserfolg von Counter Speech aus der Perspektive Jugendlicher:

- wenn man als Counter Speaker*in selbst zum Opfer wird

- wenn die eigenen Interventionen noch mehr Hate hervorrufen
- wenn man sich ausschließlich an der Zahl oder Aktualität der Likes für den eigenen Counter Speech Kommentar orientiert, da dies nicht kontrolliert werden kann
- wenn man den Erfolg anhand der Kommentare beurteilt, die man zurückbekommt, da diese (unberechtigterweise) auch negativ ausfallen können
- wenn man Hater*innen zum Verschwinden gebracht hat (z.B. durch Gegenattacke), aber Konflikte nicht gelöst wurden
- wenn man die Situation zwar nicht verschlimmert, aber ignoriert

Damit werden klar die Risiken zum Ausdruck gebracht, wonach Counter Speech Hate Speech weiter eskalieren kann. Gleichzeitig wird eine ambivalente Haltung gegenüber möglichen Unterstützungleistungen anderer sichtbar: zwar wünschen sich Jugendliche Reaktionen als Zeichen des Counter Speech Erfolgs, gleichzeitig können solche Reaktionen häufig nicht als Bewertungsgrundlage für den tatsächlichen Erfolg herangezogen werden. Gegenattacken sind zumindest für einen Teil der Jugendlichen keine zielführenden Counter Speech Optionen. Dass auch Ignorieren falsch ist, wird nur von einem Teil der Jugendlichen vertreten und stammt wohl auch aus dem Kontext der Aktionsforschung, wo das Forschungsteam Ignorieren als Strategie immer wieder in Frage stellte. Ein jugendlicher Teilnehmer fasst dies so zusammen: „*Wenn du dich nicht einmischst wird halt nichts passieren, aber es wird sich auch nichts ins Positive ändern*“ (G3, männlich, 1:49:20)

Ob und wann Counter Speech erfolgreich ist oder sein kann, wurde auch in den Diskussionen im Zuge der anderen Übungen (insbesondere Vorgehen gegen reale Hate-Kommentare) besprochen. Dabei wird deutlich, dass diese Frage für die Jugendlichen schwer zu beurteilen ist. Jedenfalls ist aus Sicht der Jugendlichen die Wirkung von persönlichen Konfrontationen deutlich höher, da sich Hater*innen im Internet auch sehr leicht einer Auseinandersetzung entziehen können bzw. dann einfach nicht erreichbar sind: „*[Der Erfolg ist] sicher nicht so, als würde man das persönlich sagen. Kommt auch auf die Person an, ob sich die das durchliest und Gedanken darüber macht oder einfach sagt, das ist nicht meine Meinung.*“ (G3, männlich, min 30:44).

Wenn Counter Speech (bei fremden Personen) praktiziert wird, dann sind das oft auch spontane Reaktionen, die nicht weiterverfolgt werden. Entsprechend bleibt dann aber auch unklar, ob der eigene Einsatz letztlich erfolgreich war oder nicht: „*Ich hab eigentlich nicht mehr nachgeschaut ob das jetzt erfolgreich war oder nicht, ich hab einfach das Kommentar hinterlassen*“ (G4, weiblich, min 44:00). Erschwert wird die Einschätzung des Erfolgs auch dadurch, dass eindeutige Zeichen, wie etwa eine Entschuldigung in der Regel ausbleiben: „*Also bei mir hab ich das noch nie gesehen. Noch niemand hat sich jemals entschuldigt*“ (G4, männlich, min 40:03).

Prinzipiell herrscht die Auffassung, dass Hater*innen im Internet ihre Meinung nicht ändern und mit Counter Speech eher die Gefahr besteht, dass sich vorgefasste Meinungen nur noch mehr verfestigen und die Situation insgesamt stärker eskaliert. Die Nutzung der Meldefunktion wird in solchen Fällen manchmal sogar als effektiver wahrgenommen – wenn auch manche der Meinung sind, dass dies meistens auch nicht sehr viel nützen würde, da das in der Regel nur wenige machen würden. Es gäbe zudem auch negative Inhalte im Internet, wo Counter Speech prinzipiell nicht umsetzbar zu sein scheint, z.B. wenn auf Seiten bestimmte negative Inhalte propagiert werden und damit einen schlechten Einfluss ausüben. Genannt werden z.B. Seiten, wo Leute zeigen, wie sich ritzen, oder Seiten über Magersucht.

In den Diskussionen wurde auch von eigenen Opfer-Erfahrungen berichtet: so schildert etwa ein männlicher Teilnehmer der Aktionsforschung, dass er mit 10/11 Jahren zunächst die Hate-Kommentare gelöscht hätte, später habe er versucht, sachlicher dagegen vorzugehen – wobei er meint, dass dies nur bei einem geschätzten Anteil von 5% gelungen sei – und dann hätte er begonnen, diese einfach zu ignorieren. Auch solche Erfahrungen bestärken die Jugendlichen darin, dass Ignorieren letztlich die bessere Strategie sei als Counter Speech, da dies Hater*innen eher dazu anzuregen oder sogar daran erinnern würde, weiter zu haken.

Als Motivation für das eigene Auftreten als Counter Speaker*in bleibt dann letztlich das gute/befreiende Gefühl bzw. Gewissen, wenn auf Situationen reagiert wird, die dem eigenen Norm- bzw. Wertesystem widersprechen: *„Ich hab das ein paar Mal gemacht auf TikTok bei Rassismus-Postings oder wo auch Leute mit menschlicher Behinderung diskriminiert wurden ... das war dann so irgendwie befreiend, dass ich denen gesagt hab, ja, warum tust du sowas, das ist einfach nur unter aller Würde“* (G3, weiblich, min 36:55). Um Jüngeren zu vermitteln, dass Counter Speech Sinn macht, müsste man nach Ansicht mancher Jugendlicher vermitteln, dass man sich im Nachhinein gut fühlt, weil man etwas gemacht hat und dass es genau so sinnvoll sei, im Internet zu reagieren wie offline. Die Motivation, selbst spontan zu intervenieren, könne aber auch dann erhöht werden, wenn man sich an der Vorbildwirkung anderer orientiert: *„[Ich habe auf Hate reagiert weil] ich hab auch gesehen, dass auch mehrere andere diese Position verteidigt haben.“* (G4, weiblich, min 44:00)

3.3 Zusammenfassung und Schlussfolgerungen

Im Rahmen einer partizipativen Online Aktionsforschung (umgesetzt von Februar 2020 bis Juni 2021) mit 31 Jugendlichen im Alter von 14 bis 22 Jahren (17 Mädchen, 14 Jungen) wurde in Zusammenarbeit mit dem Projektteam die Anwendbarkeit verschiedener Gestaltungsmöglichkeiten von Online Zivilcourage im Umgang mit Hate Speech und Cyberbullying erforscht. Durchgeführt wurde die Aktionsforschung mit vier Gruppen à 5-11 Teilnehmer*innen, die sich jeweils nach Geschlechtszusammensetzung und Bildungshintergrund unterschieden. Jede Gruppe traf sich je vier Mal (= 16 Treffen insgesamt). Zwischen

den einzelnen Gruppentreffen erhielten die Jugendlichen kurze Arbeitsaufträge (z.B. Sammeln von Screenshots mit Online-Übergriffen und Konterreaktionen). Folgende Aktions- und Forschungsphasen wurden umgesetzt: gemeinsame Analyse von Optimierungspotenzialen jugendlicher Interventionsstrategien auf Basis von Online-Rollenspiel-Auszügen auf WhatsApp; Durchführung einer Simulationsübung in Form von „Speed Counter Speech“ im Vorgehen gegen Hate-Kommentare auf größeren öffentlichen Plattformen wie z.B. YouTube; Bewertungen von unterschiedlich ausgerichteten Videobeiträgen von Influencer*innen, um für ein Vorgehen gegen Hate Speech zu sensibilisieren; reales Vorgehen gegen echte Hate-Kommentare mit Hilfe von Fake-Profilen; sowie die Klärung der Frage nach dem Erfolg von Counter Speech durch die Entwicklung von Quizfragen.

Die Ergebnisse zeigen zunächst, dass das, was die beteiligten Jugendlichen mit Counter Speech verbinden, nur bedingt der Definition von Counter Speech als zivilcouragierte Handlung entspricht: Genannt wurden etwa Reaktionsvideos von Influencer*innen, die z.B. auf den provozierenden oder negativen Content anderer Influencer*innen aufmerksam machen und dadurch aggressive Gegenattacken von unterstützenden User*innen auslösen, die dann allgemein auf Verständnis stoßen. Ebenso nannten die beteiligten Jugendlichen die Praktizierung der „Cancel Culture“, d.h. der systematische Ausschluss bzw. Angriff von Personen, die sich nach Ansicht der Community falsch verhalten haben und dadurch faktisch zensuriert werden. Weiters nannten die teilnehmenden Jugendlichen als Beispiele teils faktenbasierte, teils unreflektierte Reaktionen bei Auseinandersetzungen in Bezug auf kontroverse Themen (z.B. Veganismus, Homosexualität); Reaktionen von Fan-Gruppen, die Celebrities unterstützen; Gegenreaktionen von Opfern selbst, die in der Regel mehr Anerkennung erfahren als jene durch dritte Personen. Interventionen durch Dritte fielen den Teilnehmer*innen eher dadurch auf, dass diese entweder zum Ausdruck bringen, dass sie selbst genervt sind, oder dass sie für andere einfach unterhaltsam zu lesen sind. Die Vorstellungen der Jugendlichen decken sich am ehesten dann mit der Definition von Counter Speech im Sinne von Zivilcourage, wenn es um den Einsatz für Freund*innen geht.

Die an der Aktionsforschung teilnehmenden Jugendlichen hatten insgesamt wenig eigene Erfahrung im Praktizieren von Counter Speech und maßen dem Thema auch keine vordergründige Bedeutung bei, obwohl es ihren Moralvorstellungen klar widersprach, wenn Personen im Internet Online-Übergriffen ausgesetzt sind. Die teilnehmenden Jugendlichen vermittelten den Eindruck, Begegnungen mit den vielfältigen Formen von digitaler Gewalt seien inzwischen zur Normalität geworden und das Ignorieren im Umgang damit eine gängige und akzeptierte Strategie. Wenn es auch als wichtig betrachtet wurde, darüber zu informieren, dass Hass keine Meinung ist, so wurde dennoch Ignorieren im Vergleich zu Counter Speech als die effizientere Lösung gesehen. Counter Speech wurde prinzipiell mit mühsamen und langwierigen Interaktionen assoziiert, oft seien zugehörige Kontexte unklar oder überfordernd in der Wahrnehmung und wurden letztlich als Zeitverschwendung oder auch unnötiges Einmischen interpretiert.

Um dieser wenig motivierenden Grundhaltung der Jugendlichen bezüglich der Sinnhaftigkeit von Counter Speech entgegen zu treten, wurde im Rahmen der Aktionsforschung gemeinsam nach konstruktiven und praxistauglichen Counter Speech Vorschlägen gesucht und dafür unterschiedliche Phasen der Forschung und Aktion umgesetzt. Die Aktionsforschung basierte auf der Frage, wie aus Sicht der Jugendlichen Counter Speech gestaltet bzw. optimiert werden kann. Die zentralen Ergebnisse werden nachfolgend kurz zusammengefasst.

Zunächst ist festzuhalten, dass jugendliche Counter Speaker*innen in der Regel **vor einem großen Kreis an Zuschauer*innen agieren**. Entsprechend ist es für Jugendliche wichtig genau zu überlegen, was gepostet wird bzw. wie man sich dabei selbst präsentiert und welche möglichen Gegenreaktionen zu erwarten sind. Wichtig ist es in diesem Zusammenhang zu unterscheiden, ob man als Counter Speaker*in in einem privateren Kontext, z.B. in einer WhatsApp-Gruppe auftritt, wo man in der Regel die meisten Personen kennt, oder in größeren, öffentlichen Kontexten, wie z.B. auf YouTube, mit vielen unbekanntem oder anonymen Akteur*innen.

Bei den Vorschlägen zur **Gestaltung des Auftretens als Counter Speaker*in in privaten Kontexten** verweisen die Ergebnisse auf deutliche bildungsspezifische Differenzen: Jugendliche mit höherer Bildung verfolgten das Ziel, in ihrem Auftreten als Counter Speaker*in ernst genommen zu werden, indem sie ihren intellektuelle Vorteil ausspielten und z.B. die Umsetzung strafrechtlicher Konsequenzen androhten, verbunden mit dem Posten konkreter Gesetzestexte. Jugendliche mit niedrigerer Bildung hingegen versuchten eher, durch Gegenattacken ernst genommen zu werden. Die jeweiligen Erfolgsaussichten wurden durchaus skeptisch eingeschätzt: Jugendliche mit geringer Bildung vermuteten, dass das Posten von Gesetzestexten von Hater*innen entweder gar nicht gelesen oder nicht ernst genommen würde, da man aus Erfahrung wisse, dass ohnehin nie eine Anzeige folgen würde. Jugendliche mit höherer Bildung wiederum empfanden Gegenattacken als wenig zielführend, da dies die Situation nur noch mehr aufheizen würde. Es wurden also jeweils die bevorzugten Strategien der jeweils anderen Gruppe als wenig wirksam interpretiert. Darüber hinaus gelang es Jugendlichen mit höherer Bildung in ihren Counter Speech Vorschlägen eher, ihre Konter-Kommentare mit bildgestützten Gegenreaktionen wie z.B. Memes zu kombinieren und damit ihren Botschaften mehr Aufmerksamkeit zu verleihen bzw. diese unterschiedlich zu konnotieren. Dadurch sollte z.B. gezielt versucht werden, die Situation aufzulockern (z.B. mit humorvollen GIFs), die eigenen intellektuellen Kapazitäten noch stärker zu betonen (z.B. Mitschicken einer Reverse-Karte), oder auf subtilem Wege die eigene Einschätzung der Situation mitzuteilen (z.B. Bild, wo sich jemand an die Stirn greift). Jugendliche mit geringer Bildung meinten hingegen, dass ihnen dazu nicht viel einfallen würde bzw. sie sich nicht ausreichend kreativ fühlten. Der Einsatz von Memes / GIFs wurde aber generell eher positiv konnotiert und mit guter Laune verbunden. Insofern bestünde nach Ansicht der Jugendlichen die Gefahr, die missverständliche Botschaft zu übermitteln, dass die Situation nicht so schlimm sei, daher seien bildgestützte Gegenreaktionen oft unpassend. Jugendliche mit niedrigerem Bildungshintergrund schlugen im Vergleich zu jenen mit höherer Bildung auch stärker moralisierende Zugänge vor mit dem Ziel, Hater*innen ihr Fehlverhalten verständlich bzw. ihnen ein schlechtes Gewissen

zu machen; oft wurde dabei aus einer Opferperspektive heraus argumentiert. Insbesondere aus Sicht der Jugendlichen mit höherer Bildung wurden sie aber genau dadurch als eher schwach erlebt.

Bei den Vorschlägen zur **Gestaltung des Auftretens als Counter Speaker*in in größeren öffentlichen Kontexten** zeigten sich ebenfalls bildungsspezifische Unterschiede: Jugendliche mit geringerer Bildung zeigten eher aggressiv-emotionale Reaktionen, jene mit höherer Bildung eher schlagfertige, überhebliche, sich lustig machende Counter Speech Strategien. In der Ausformulierung sachlich fundierter Gegenargumente waren die beteiligten Jugendlichen mit höherer Bildung gegenüber jenen mit niedriger Bildung sprachlich im Vorteil, ebenso wie in der argumentativen Darlegung der eigenen Positionierung bei kontrovers diskutierten Themen (z.B. persönliche Haltung zum Umgang mit Homosexualität). Insgesamt wirkten die beteiligten Jugendlichen mit geringer Bildung im Auftreten weniger überzeugend und sprachlich weniger versiert (z.B. „lass die machen was die wollen und misch denen nicht ein“). Schlagfertige Gegenreaktionen, die eine Unterstützung durch mitlesende Peers wahrscheinlicher machen, z.B. in Form einer Retourkutsche, kamen in dieser Studie ausschließlich von höher gebildeten Jugendlichen. Auffallend ist auch hier, dass nur Jugendliche mit geringer Bildung in ihren Gegenreaktionen auch Bezug zum Opfer nahmen (verteidigend, moralisierend).

Die Counter Speech Reaktionen der Jugendlichen waren aber auch stark von der Art und Weise bzw. vom Typus des Hate-Speech Kommentars selbst abhängig: auf klar aggressive und übergriffige Hass-Kommentare wurden auch entsprechend mit Beschimpfungen reagiert, gegen subtilere Formen eines Übergriffs wurde eher auf argumentative Weise vorgegangen. Als besonders herausfordernd erlebten die teilnehmenden Jugendlichen Reaktionen auf als „dumm“ wahrgenommene Hate-Kommentare, d.h. Kommentare, die keine Grundlage für Gegenargumente bieten oder aufgrund fehlender Kontextinformationen schwer verständlich sind. Hier wurden unterschiedlich konnotierte Kurzkommentare vorgeschlagen, wie „ok“, „shit bro“, „pain“, die signalisieren, dass solche Aussagen nicht in Ordnung sind, gleichzeitig aber auch wenig weitere Angriffsfläche offen lassen. Jugendliche mit höherem Bildungshintergrund verwendeten solche Ausdrücke insgesamt versierter (z.B. auch englischsprachige Ausdrücke), Jugendliche mit geringem Bildungshintergrund formulierten sprachlich einfacher (z.B. „NANA Nicht gut“). Interessant ist auch, dass es den beteiligten Jugendlichen vielfach schlicht an Übung fehlte, da sie in der Regel auf Hate-Kommentare nicht reagieren.

Counter Speech meint nicht nur aktive, zivilcouragierte Gegenrede, sondern auch ein **allgemeines Auftreten gegen negative und hasserfüllte Inhalte im Internet**. Häufig sind dazu Videobeiträge von Influencer*innen zu finden (oft auch als Anti-Hate-Kampagnen gedacht), deren Sinnhaftigkeit und Wirksamkeit ebenfalls mit den beteiligten Jugendlichen diskutiert wurde. Sie betrachteten humorvolle und positive Darstellungen (z.B. Hate Kommentare vorlesen und sich lustig machen) prinzipiell angenehmer als ernsthafte und emotional belastende Beiträge (z.B. Darstellung, dass einen Hate betroffen macht); diese dienen nach Ansicht der Jugendlichen aber ausschließlich der Unterhaltung und der Selbstdarstellung von Influencer*innen. Humorvolle Beiträge würde man auch viel eher mit anderen teilen als ernsthafte, da

man sonst womöglich selbst Rückfragen erhalten würde, ob alles in Ordnung sei. Gut aufgenommen wurden pointierte, sachliche Aufklärungsvideos mit Fokus auf eine bestimmte Thematik (z.B. Rassismus) mit dem Vorteil, dass den Jugendlichen hier die notwendige Argumentationsgrundlage für das eigene Kontern vermittelt wird. Auch kreative Bearbeitungen (z.B. Aufbereitung von Hate-Kommentaren zu einem Song) kamen bei den teilnehmenden Jugendlichen gut an, sofern diese professionell ausgearbeitet und angenehm zu konsumieren sind, vorzugsweise mit einer positiven Komponente. Kampagnenvideos gegen Online Hass, z.B. in Form eines Zusammenschlusses von mehreren YouTuber*innen, um Jugendliche für ein Vorgehen gegen Hate zu mobilisieren, fanden ebenfalls Anklang, vor allem aufgrund der hohen Popularität der Influencer*innen und deren Reichweite – entsprechend könnten nach Ansicht der Jugendlichen z.B. auch persönliche Aufrufe von Einzelpersonen kaum etwas erreichen. Insgesamt wurde betont, dass bei allen gezeigten Videos die persönliche Favorisierung eines Influencers entscheidend sei, Videos zu diesem Thema überhaupt anzusehen. Trotz positiver Bewertungen einiger Videobeispiele waren die teilnehmenden Jugendlichen aber hinsichtlich ihrer Sinnhaftigkeit und Wirkung skeptisch: sie vermuteten, dass Hater*innen mit solchen Videos ohnehin nicht erreichbar wären und Online-Hass dadurch nicht reduziert werden könne. Am ehesten wären junge Internetnutzer*innen erreichbar, welche die teilnehmenden prominenten Influencer*innen als Idole sehen – allerdings würde man sich in jungen Jahren oft einfach noch nicht trauen, gegen Hate-Kommentare aktiv aufzutreten.

Die von den Jugendlichen genannten Gestaltungsvorschläge wurden schließlich hinsichtlich ihrer **Praktikabilität im Vorgehen gegen reale Hate-Kommentare auf YouTube und TikTok getestet**. Dabei zeigte sich, dass viele strategische Zugänge, die eher für privatere Kontexte überlegt wurden, nicht praktikierbar waren, beispielsweise aufgrund der Anonymität (z.B. Konfrontation mit Fakten zu möglichen strafrechtlichen Folgen wird nicht ernst genommen) oder technischer Gegebenheiten (z.B. Posten von Memes auf YouTube und TikTok nicht möglich).

Jugendliche mit höherer Bildung verwendeten als vorrangige Strategie die Demonstration ihrer intellektuellen Vorteile, indem sie versuchten, auf textbasierter Ebene schlagfertig zu reagieren bzw. mit Wissen zu punkten und dabei klug, überlegen, humorvoll und zugleich höflich zu wirken. Gleichzeitig wurde versucht, als Gruppe nicht locker zu lassen und immer wieder durch Rückfragen Hater*innen darauf hinzuweisen, dass Kritik zwar zulässig sei, aber ohne Beleidigungen und Schimpfwörter. Die Jugendlichen berichteten allerdings auch, dass ein solches Vorgehen als anstrengend und belastend empfunden wurde und sie nicht immer schlagfertige Reaktionen parat hatten. Dies äußerte sich auch dadurch, dass die Tonalität bei einigen Jugendlichen mit der Zeit auch zunehmend angriffiger wurde. Zudem hatten sie Angst, Fehler zu machen (z.B. Rechtschreibfehler oder Unkenntnis gebräuchlicher Begriffe im Netzjargon wie „simp“) und damit sofort Angriffspunkte zu eröffnen. Ihre Erfahrungen endeten schließlich mit dem resignierenden Gefühl, letztlich nicht erfolgreich gewesen zu sein. Potenzial wurde lediglich in der Idee gesehen, gemeinsam aufzutreten, was aber aufgrund der technischen Gegebenheiten z.B. auf YouTube erschwert wurde: So brauchten z.B. gepostete Kommentare unverhältnismäßig lange, bis diese tatsächlich

zu sehen waren, zusätzlich waren diese schwer zu finden, da sie bei der üblichen Einstellung „Top Kommentare zuerst“ diese meist weit unten gelistet waren.

Für Jugendliche mit geringerem Bildungshintergrund erschienen die hohen Anforderungen an die kognitiven Kompetenzen im textbasierten Counter Speech Verhalten als große Herausforderung: es wurde nicht nur weniger häufig, sondern auch mit deutlich kürzeren Formulierungen (z.B. „lass ihn doch“) und kaum ausformulierten Begründungen interveniert. Argumentiert wurde dies von den teilnehmenden Jugendlichen damit, dass sie entweder keine Lust hätten, „herumzudiskutieren“, oder ihnen einfach nichts einfallen würde. Leichter fiel diesen Jugendlichen die Reaktion dann, wenn gegen klar erkennbare No-Go-Hate-Kommentare vorgegangen werden sollte, z.B. jemandem den Tod zu wünschen: hier konnte relativ einfach darauf hingewiesen werden, dass solche Übergriffe zu weit gehen; zudem ist es in solchen Fällen auch wahrscheinlicher, Unterstützung von mitlesenden Peers zu erhalten. Auch wenn das Opfer als eine klar hilfsbedürftige Person wahrgenommen wurde (z.B. Übergriffe an offensichtlich beeinträchtigten Personen), nahmen geringer gebildete Jugendliche klar die Position ein, das Opfer öffentlich zu verteidigen. Insgesamt erwies sich die Wahrnehmung des Opfers (z.B. auch als sympathisch oder unsympathisch bzw. er/sie hat Hate „verdient“ oder nicht) als für das eigene Eintreten besonders bedeutsam. Aber auch hier beendeten die Jugendlichen ihren Einsatz als Counter Speaker*innen mit dem Gefühl, erfolglos gewesen zu sein. Sich als Counter Speaker*innen gegenseitig zu unterstützen, wurde auch in dieser Gruppe zwar prinzipiell als gute Idee eingestuft, allerdings müsste das Ausmaß deutlich höher sein als das innerhalb der Gruppe möglich war. Das offensive Anprangern von Hater*innen durch Influencer*innen (z.B. bekannter TikToker weist auf ein beeinträchtigtes Mädchen hin, das gehatet wird mit dem klaren Hinweis: „Schämt euch in Grund und Boden!“) zeigte für die Mobilisierung von Jugendlichen zu aktiver Gegenrede Wirkung – allerdings häufig in Form von wenig wünschenswerten Gegenattacken auf die Hater*innen.

Schließlich zeigen die Ergebnisse, dass die erarbeiteten **Counter Speech Strategien letztlich daran scheiterten, dass ihre Wirkung und ihr Erfolg für Jugendliche nicht erkennbar war**. Wenn auch konkrete Entschuldigungen in der Regel nicht zu erwarten sind, so wünschten sich die beteiligten Jugendlichen dennoch grundsätzlich positive und unterstützende Reaktionen auf ihr Counter Speech Verhalten, z.B. in Form von Likes oder Kommentaren.

Insgesamt lässt sich schlussfolgernd zusammenfassen, dass für eine Förderung von Counter Speech durch Jugendliche zunächst Aufklärung braucht, was Counter Speech ist, wovon sie sich abgrenzt und was sie bewirkt. Festzuhalten ist außerdem, dass Counter Speech im Sinne von Online Zivilcourage durch Jugendliche nur im Maßstab vorhandener Basiskompetenzen gestaltbar ist, da es insbesondere Jugendlichen mit geringer Bildung häufig an Argumentations- und Formulierungskompetenzen mangelt bzw. deren Schlagfertigkeit geringer ausgeprägt ist. Darüber hinaus mangelt es an Übung, da Online-Übergriffe in der Regel ignoriert werden. Die Gestaltung von Counter Speech ist zudem auch stark an die

technologischen Gegebenheiten gebunden, da alternative, nicht textbasierte Zugänge wie die Verwendung von Memes oft nicht einsetzbar sind.

Die Nutzung verfügbarer digitaler Strukturen zur Mobilisierung und Vernetzung mit anderen Peers, um nicht allein als Counter Speaker*in auftreten zu müssen (z.B. Instagram-Story mit dem Hate-Posting machen und dann Leute zum Kommentieren auffordern; einen Link zum Hate-Kommentar kopieren und in Communities posten, die das unterstützen könnten), wird weder von Jugendlichen mit höherer Bildung noch von jenen mit geringer Bildung mitgedacht – was vermutlich daran liegt, dass die Sinnhaftigkeit von Counter Speech aufgrund des nicht erkennbaren Erfolgs grundsätzlich in Frage gestellt wird. Neben einer notwendigen Information, Schulung und Training sind aber auch zukünftige technologische Entwicklungen gefordert, Counter Speaker*innen in ihrem Handeln stärker zu unterstützen, indem z.B. Jugendlichen mit geringeren Kompetenzen automatisierte Counter Speech Vorschläge zur Verfügung gestellt werden könnten, und zusätzlich auch der Erfolg von Counter Speech durch gezielte Unterstützung z.B. durch Likes und positive Kommentare sichtbar gestaltet werden könnte.

4. Online Feld-Experiment zur Prüfung von Counter Speech Wirkungspotenzialen

Um den jugendgerechten Einsatz von Counter Speech als Antwort auf Hate Speech in einem möglichst realen Setting zu prüfen, wurde ein Online Feld-Experiment in Kooperation mit einem YouTuber umgesetzt. Im Rahmen dieses Online Feld-Experiments sollte auf einem realen YouTube-Kanal eines*einer kooperierenden YouTuber*in, Online-Übergriffe kontrolliert inszeniert und damit zu Counter Speech angeregt werden. Ziel war es zu überprüfen, was Counter Speech durch Jugendliche in der konkreten alltäglichen Verwendung bewirken kann bzw. ob und unter welchen Bedingungen sich Counter Speech als erfolgreich erweist. Dazu sollten vor allem solche jugendgerechten Counter Speech Strategien verwendet werden, die sich auf Basis der bisherigen Erfahrungen (Online-Rollenspiele, Aktionsforschung, Inhaltsanalyse von YouTube-Daten) als besonders vielversprechend erwiesen haben (z.B. sachliche Gegenargumentationen, Fragen stellen, Humor) und nicht als aggressive Gegenattacken formuliert wurden.

4.1 Methodisches Design

Für die **methodische Umsetzung** des Online Feld-Experiments wurde nach YouTuber*innen gesucht, die (a) vorrangig Jugendliche als Zielgruppe ansprechen, (b) in der inhaltlichen Ausrichtung ihrer Kanäle sowohl weibliche als auch männliche Jugendliche ansprechen und (c) eine möglichst hohe Reichweite haben. Die Suche nach geeigneten Kooperationspartner*innen gestaltete sich deutlich schwieriger als erwartet: Es gab große Bedenken der in Frage kommenden YouTuber*innen, da diese im Falle einer stark konflikthaften Dynamik aufgrund der durch das Forschungsteam inszenierten Hate Speech negative Auswirkungen auf die Akzeptanz und Reichweite des Kanals befürchteten bzw. annahmen, dass ihre Community bzw. Teile der Community sich von ihnen distanzieren könnte(n). Schließlich ist es gelungen, einen YouTuber als **Kooperationspartner** zu gewinnen, der allen Anforderungen entspricht und auch unter Jugendlichen eine relativ hohe Bekanntheit hat:

Joseph DeChangeman: [Joseph DeChangeman - YouTube](#)



Abbildung 1

Der Kanal hat aktuell 202.000 Abonnent*innen und spricht laut seinen Analytics hauptsächlich 16- bis 26-Jährige an. Aus Art und Formulierung der Kommentare kann geschlossen werden, dass diese eher einer höheren Bildungsschicht zuzuordnen sind. Joseph kooperiert auch immer wieder mit anderen bekannten Influencer*innen wie z.B. Taddl, Zeo, Unge, die bei Jugendlichen ebenfalls eine hohe Reichweite haben.

Sein Account zeichnet sich durch regelmäßige Selbstexperimente aus, z.B. eine Woche nicht sprechen, ein Monat ohne Spiegel oder eine Woche ohne Uhr. Auf seinem Kanal finden sich auch Dokumentationen, beispielsweise über den YouTuber Rezo. Durch die Art seiner Videos wird seine Community immer wieder animiert, eigene Erfahrungen mit dem behandelten Thema in den Kommentaren zu erörtern. Dieser Anregung kommen viele User*innen auch häufig nach. Die Community erscheint insgesamt als sehr aufgeschlossen, diskussionsfreudig und anspruchsvoll hinsichtlich des Contents. In den Kommentaren finden sich viele Komplimente von seinen Abonnent*innen zur hohen Qualität der Videos bzw. zu deren Inhalten.

Konzeption und Aufbau des Online Feld-Experiments wurden über folgende 6 Phasen geplant und orientierten sich an der jeweiligen Neuerscheinung eines Videos auf Josephs Kanal:

- **Phase 1:** Möglichst unmittelbar nach Veröffentlichung des neuesten Videos auf dem Kanal wurden vom Forschungsteam zwei Online-Übergriffe inszeniert: Diese richteten sich einmal gegen eine*n vermeintliche*n User*in (= eingeweihte Person des Forschungsteams) und einmal gegen den kooperierenden YouTuber (Kanalbetreiber). Dazu wurden unterschiedliche fiktive „Hater*innen-Profil“, sowie ein „Fan-Profil“ für die*den vermeintliche*n User*in vorbereitet. Diese Profile wurden möglichst „jugendgerecht“, ohne besondere weiterführende Kontextinformationen gestaltet, d.h. diese wurden nicht z.B. rechtsradikal dargestellt. Um im konkreten Einsatz eine gewisse Flexibilität und Variabilität zu gewährleisten, wurden im Laufe des Experiments mehrere nutzbare fiktive Profile eingerichtet, die alle dem gleichen Schema entsprachen. Sobald auf die Kommentare der fiktiven Hater*innen-Profil reagiert wurde (Counter Speech), wurden diese auch genutzt, um die inszenierten negativen Kommentare zu verstärken. Umgesetzt wurden diese Übergriffe durch eine internetaffine und YouTube-erfahrene junge Person im Projektteam – die realitätsnahe und authentische Umsetzung eines solchen Übergriffs wurde bereits im Rahmen der Online-Aktionsforschung erfolgreich erprobt. Alle Interaktionen, inklusive spontane Reaktionen der jeweiligen (realen) Online-Community, wurden mittels Screenshots dokumentiert.

Analyseziel dieser ersten Phase war es zu erfassen, wie die reale Online-Community auf Hate Speech (HS) reagiert bzw. ob Counter Speech (CS) durch die Community vorkommt und wie diese gestaltet wird. Dies sollte auch als Vergleichsbasis für Phase 2 dienen, wo neben Hate Speech gezielt auch Counter Speech inszeniert wurde (siehe nachfolgender Punkt).

- **Phase 2:** Im nächsten (=zweiten) neu erschienenen Video des YouTube Kanals wurde das Setting von Phase 1 wiederholt (Übergriff gegen vermeintliche*n User*in und YouTuber*in), aber zusätzlich zu Hate Speech *gezielt mit Counter Speech reagiert* (vor allem solche CS-Strategien, die

sich aus bisherigen Erfahrungen im Projektverlauf bewährt haben, keine aggressiven Gegenattacken). Um Counter Speech durch Jugendliche möglichst realitätsnah umzusetzen, wurde dies von ausgewählten Jugendlichen aus den Aktionsforschungsgruppen in Begleitung der Forscher*innen durchgeführt. Die freiwilligen Teilnehmer*innen erhielten Online-Einkaufsgutscheine als Dankeschön. Es wurde eine Counter Speech Gruppe von zehn 14-19-jährigen Jugendlichen (geschlechtergemischt) eingerichtet. Diese erhielten genaue Informationen zum Prozedere und wurden hinsichtlich unterschiedlicher Counter Speech Strategien zusätzlich in einem gemeinsamen Online-Workshop vom Forschungsteam gecoacht. Die Umsetzung erfolgte mittels dafür vom Forschungsteam vorbereiteter Fake-Accounts, sodass die Jugendlichen nicht mit ihrem tatsächlichen, privaten Profil aufschienen. Um die Interaktionen zwischen Hater*innen und Counter Speaker*innen möglichst realitätsnah zu gestalten, wurden die fiktiven Hater*innen Profile ebenfalls wieder benutzt, um negative Inhalte als Reaktion auf Counter Speech weiter zu verstärken. Sämtliche Interaktionen wurden wieder mittels Screenshots dokumentiert.

Analyseziel dieser zweiten Phase war es zu prüfen, inwieweit sich in den Reaktionen der Community Unterschiede zeigen, wenn bei einem Video nur Hate Speech und bei dem anderen zusätzlich gezielt Counter Speech inszeniert wird.

- **Phasen 3-6:** Bei den nachfolgenden jeweils neue erschienenen Videos wurde das Prozedere von Phase 1 bzw. Phase 2 abwechselnd wiederholt: D.h. beim dritten und fünften neuen Video wurde das Setting von Phase 1 wiederholt (Hate Speech ohne inszeniertem Counter Speech), beim vierten und sechsten das Setting von Phase 2 (Hate Speech mit inszeniertem Counter Speech). Sämtliche Interaktionen inklusive der spontanen Reaktionen der realen Online Community wurden mittels Screenshots dokumentiert.

Analyseziel dieser Forschungsphasen war es zu prüfen, inwieweit es Unterschiede im Verhalten der realen Community gibt, wenn man Counter Speech gezielt einführt und dann wieder absetzt.

Das Online Feld-Experiment umfasst also insgesamt drei Interventionen, bei denen ausschließlich Hate Speech inszeniert wurde und weitere drei Interventionen, bei denen zusätzlich zu Hate Speech auch gezielt mit Counter Speech interveniert wurde, diese Interventionen wurden abwechselnd umgesetzt und lassen sich wie folgt zusammenfassen: **HS (Phase 1) / HS+CS (Phase 2) / HS (Phase 3) / HS+CS (Phase 4) / HS (Phase 5) / HS+CS (Phase 6)**. Das Untersuchungsdesign ermöglichte dadurch folgendes **Kontrollgruppen-Setting**: sowohl die Community als auch die Person des YouTubers wurde konstant gehalten, während der Einsatz von Counter Speech gezielt variiert wurde. Damit konnte analysiert werden, ob die reale Community bei Videos, zu denen ausschließlich Hate Speech inszeniert wurde, im Vergleich zu jenen, wo zusätzlich auch Counter Speech eingesetzt wurde, unterschiedlich reagiert. Gleichzeitig sollte damit geprüft werden, ob es gelingt, durch gezielte Counter Speech auch andere User*innen zu mehr Counter Speech zu motivieren.

Die **Zusammenarbeit mit dem kooperierenden YouTuber** wurde genau geregelt. Dieser wurde über die Ziele, Konzeption und Herangehensweisen des Feld-Experiments detailliert informiert, die inszenierten Übergriffe wurden vorab mit ihm abgesprochen. Vor dem ersten Durchgang wurde ihm

außerdem ein Dokument mit einer Liste von „Muster“-Hasskommentaren geschickt und um sein Einverständnis gebeten. Da der YouTuber auch selbst Filter gegen bestimmte Hasskommentare eingerichtet hatte, wurden die entsprechenden Worte dem Forschungsteam mitgeteilt. Dadurch konnten die inszenierten Hate-Kommentare so formuliert werden, dass diese nicht automatisch von diesen Filtern gelöscht wurden. Zusätzlich wurde der kooperierende YouTuber vor jedem Durchlauf des Feld-Experiments kontaktiert: Er wurde informiert, dass unter dem jeweils aktuellsten Video Hate- bzw. Counter Speech) durch das Team inszeniert wird, und sein Einverständnis dafür wurde eingeholt. Der YouTuber wurde außerdem gebeten, über den gesamten Untersuchungszeitraum hinweg keine negativen Inhalte zu löschen oder zu kommentieren. Weiters wurde vereinbart, dass je nach Verlauf des Online Feld-Experiments am Ende besprochen wird, ob, wie und von wem die Aufklärung der Online Community durchgeführt wird. Angedacht wurde die Gestaltung eines eigenen Videobeitrags durch den YouTuber, inklusive Interview mit dem Forschungsteam.

In der **Umsetzung des Untersuchungsdesigns** waren folgende Herausforderungen zu bewältigen:

(1) Die geplanten Interventionen (inszenierte Hate als auch Counter Speech) sollten möglichst zeitnah unmittelbar nach Veröffentlichung eines Videos erfolgen, um die Chance zu erhöhen, von der Community entsprechend wahrgenommen zu werden. Das Forschungsteam wurde durch den YouTuber jeweils über den Zeitpunkt einer bevorstehenden Veröffentlichung eines Videos informiert (in der Regel sonntags), sodass die geplanten Interventionen zeitnah umgesetzt werden konnten.

(2) Durch den YouTube-Algorithmus war die Wahrscheinlichkeit aber dennoch hoch, dass einfache Hate-Kommentare nicht in den Top-Kommentaren gelistet werden und damit die inszenierten Übergriffe von der Online-Community nicht gesehen bzw. nicht wahrgenommen werden. Damit wäre das Risiko hoch gewesen, dass keinerlei Interaktionen mit realen User*innen entstehen würden. Als Gegenstrategie wurde daher entschieden, Likes zu kaufen und damit die inszenierten Übergriffe weiter nach oben, unter die Top-Kommentare, zu reihen. Um die Anzahl an notwendigen Likes zu eruieren, wurden zunächst Testläufe mit „unauffälligen“ Kommentaren ohne negative Inhalte durchgeführt. Es wurden im Laufe des Experiments auch verschiedene Plattformen zum Erwerb von Likes ausprobiert, da diese nicht immer zuverlässig bzw. manchmal zeitverzögert geliefert wurden. Da jener Hasskommentar, der sich gegen eine*n (inszenierte*n) User*in richtete, eine Reaktion auf einen positiven Ursprungskommentar dieser fiktiven Person war, war es hier notwendig, die Likes nicht für den Hasskommentar, sondern für den Ursprungskommentar zu kaufen, um den gesamten Kommentarblock nach oben zu pushen. Es zeigte sich, dass idealerweise eine Anzahl zwischen 100 und 250 Likes geeignet schienen, wobei für den direkten Hate Kommentar meist etwas weniger Likes gekauft wurden als für den positiven Kommentar, gegen den dann ein Hate-Kommentar geschrieben wurde.

4.2 Durchführung des Online Feld-Experiments

In den folgenden Ausführungen wird der gesamte Prozess des Online Feld-Experiments inklusive Screenshots dargestellt. Zur Inszenierung von Hasskommentaren sowie eines Fan-Kommentars (= vermeintliche*r User*in), unter dem ein Hasskommentar inszeniert wurde, wurden im Laufe des Experiments insgesamt sechs Accounts erstellt. Zur leichteren Orientierung auf den nachfolgenden Screenshots werden hier die verwendeten **Nicknames** in den eingesetzten Fake-Profilen aufgelistet:

verwendete Nicknames für vermeintliche*n User*in:

hara kujio, Der Eagle

verwendete Nicknames für vermeintliche*n Hater*in:

schni mk, M, Count, Grim Grin, hara kujio

verwendete Nicknames für Counter-Speaker*innen (fiktive Namen):

Leah Fuchs, Emma Horak, Zarah Varova, Sam Pollak, Davi Fabiensky, Anna Maro, Marvin Leeg, Jarik Zeuner, Benni Koller

In vereinzelt Fällen haben Counter Speaker*innen des Forschungsteams mit privaten YouTube-Accounts kommentiert. In diesen Fällen wurde der Name geschwärzt.

Im gesamten Bericht wurden bei den Screenshots der Interventionen außerdem die Nicknames aller echten User*innen ab dem dritten Buchstaben geschwärzt, um so die Anonymität zu gewährleisten, während gleichzeitig der Kommunikationsverlauf verständlich bleibt.

4.2.1 Erste Intervention Hate Speech OHNE Counter Speech (Experiment-Phase 1)

Intervention 1a: Übergriff gegen User*in



Abbildung 2

Kontextinformationen zur Umsetzung

- rote Markierung: echter User*innen-Kommentar, nicht inszeniert
- „hara kujio“: inszenierter User*innen-Kommentar: am 18.04.21 um 12.17 Uhr gepostet
- „schni mk“: inszenierter Hate-Kommentar: am 18.04.21 um 12.28 Uhr gepostet
- Anzahl Likes gekauft: 175 (eingetroffen am 21.04.21)
- Anzahl Likes final: 177

Bei diesem ersten Übergriff gegen eine*n vermeintliche*n User*in gab es zumindest eine Person aus der realen Community, die mit Counter Speech (CS) reagierte (rote Markierung). Dabei orientierte sich diese offenbar stark an der Tonalität der inhaltlichen Ausrichtung des Hate-Kommentars, indem ebenfalls eher aggressiv zurückgeschimpft wurde und auch der im Hate-Kommentar verwendete Opfer-Begriff wieder aufgegriffen wurde. Der*Die Counter Speaker*in reagierte am Folgetag, nachdem der Hate-Kommentar gepostet wurde, es war also keine unmittelbare Reaktion. Allerdings blieben hier weitere Interaktionen durch andere User*innen gänzlich aus, sodass die Kommunikation schließlich mit einem nochmaligen Hate-Kommentar endete und damit Counter Speech hier als eher wenig erfolgreich zu bewerten ist. Im gesamten Kommentarblock fanden sich auch keine weiteren Hinweise auf mögliche indirekte Formen von Counter Speech. Die Anzahl der gekauften Likes für den

inszenierten Hass-Kommentar hat sich im Untersuchungszeitraum ganz leicht erhöht, sodass hier sogar eher von einer Verstärkung von Hate Speech (HS) gesprochen werden kann. Der*die Counter Speaker*in bekam keine Unterstützung durch andere User*innen (Bystander*innen).

Intervention 1b: Übergriff gegen YouTuber



Abbildung 3

Kontextinformationen zur Umsetzung:

- „M“: inszenierter Hate-Kommentar: am 18.04.21 um 12.38 Uhr gepostet
- Anzahl Likes gekauft: 250 (eingetroffen am 22.04.21)
- Anzahl Likes final: 265

Der zweite inszenierte Übergriff in Phase 1 war gegen den YouTuber selbst gerichtet und löste keinerlei Reaktionen durch Kommentare aus. Allerdings hat sich die Anzahl an Likes um 15 erhöht, das heißt, hier wurde Hate Speech durch andere (echte) User*innen verstärkt, Counter Speech kam nicht vor. Neben den dokumentierten Reaktionen der Community wurden sämtliche Kommentare unter dem Video auch auf andere reale Hate Speech Kommentare untersucht: Dabei fanden sich insgesamt folgende zwei Kommentare, die sich bezugnehmend auf einen im Video verwendeten geschlechtssensibel formulierten Begriff („Musiker*innen“) beziehen:

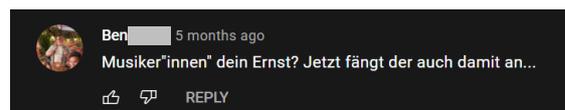


Abbildung 4

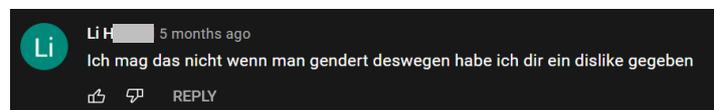


Abbildung 5

Auch hier waren keine Counter Speech Kommentare durch andere User*innen auffindbar.

Damit lässt sich für diese erste Phase zusammenfassen, dass zwar insgesamt in der Community des teilnehmenden YouTubers allgemein wenig Hate Speech zu finden ist (es überwiegt positives Feedback zu den Videos), dass aber beim Vorkommen einzelner (inszenierter) Hate Speech Kommentare Counter Speech eher wenig wahrscheinlich ist. Ganz im Gegenteil wird durch Likes zumindest indirekt eher Hate Speech verstärkt – wenn auch in einem verhältnismäßig geringen Ausmaß.

4.2.2 Erste Intervention Hate Speech MIT Counter Speech (Experiment-Phase 2)

Intervention 2a: Übergriff gegen User*in

The image shows a screenshot of a social media comment thread. At the top, a comment by 'Der Eagle' is marked as 'Markierter Kommentar'. Below it, several other comments are visible. Two comments are highlighted with red rounded rectangles:

- th0** (vor 2 Wochen):
@schni mk Es gibt Menschen, die überdenken, bevor sie etwas machen. Menschen, die den Sinn dahinter (/Inspiration) suchen, bevor sie etwas beginnen, oder um etwas loszulassen, was sie in den letzten Monaten/Jahren zu ihrer Gewohnheit gemacht haben.
- th0** (vor 2 Wochen):
@schni mk Nur die Frage ist, ob man die Andersdenkenden toleriert, oder ob man nach unlösbaren Problemen dahinter sucht.
Glückwunsch, du hast meine Message zu 0% verstanden und suchst lieber aus dem Nichts, ohne Argumente Stress 🤔🤔🤔

Other visible comments include:

- schni mk** (vor 2 Wochen): alter wie absolut peinlich bist du denn mach doch einfach selbst du opfer warum brauchst du den dazu junge
- Lea Fuchs** (vor 2 Wochen): @schni mk Was laberst du überhaupt? Du bist peinlich!
- Anna Maro** (vor 2 Wochen): @schni mk ich finde es traurig, dass du nichts besseres zu tun hast als hier herum zu haten.
- schni mk** (vor 2 Wochen): @Lea Fuchs sag mal warum peinliche
- Emma Horak** (vor 2 Wochen): Bin voll deiner Meinung, auch in der corona zeit sind wir zu viel am handy. Finde das Video voll interessant
- schni mk** (vor 2 Wochen): @th0 wow menschen denken du musst ja ein richtig riesiges gehirn haben was erklärst du mir als nächstes, wasser ist nass? hahahahahah junge du opfer
- schni mk** (vor 1 Woche): @th0 junge nerv wen anders mit deiner pseudophilosophie
- Zarah Varova** (vor 1 Woche): @schni mk Warum sollte das peinlich sein? Anscheinend findet er solche Videos interessant und wenn es dir nicht gefällt scroll einfach weiter.
- Sam Pollak** (vor 1 Woche): @schni mk ich glaube dir würde ein screen detox auch mal gut tun
- Grim Grin** (vor 5 Tagen): @Sam Pollak halt doch selber mal die fresse
- Davi Fablensky** (vor 4 Tagen): @schni mk hast du nichts besseres zu tun, als leute runterzuziehen?
- Anna Maro** (vor 4 Tagen): @Zarah Varova komplett deiner Meinung!
- Anna Maro** (vor 4 Tagen): @Grim Grin finde es lustig wie du nur hier bist um andere schlecht zu machen und hate zu verbreiten...
- Grim Grin** (vor 3 Tagen): @Anna Maro dann lach woanders th0t

Abbildung 6

Kontextinformationen zur Umsetzung:

- rote Markierungen: echter User*innen-Kommentare, nicht inszeniert
- „Der Eagle“: inszenierter User*innen-Kommentar: am 16.05.21 um 12:05 Uhr gepostet
- schni mk: inszenierter Hater*innen Kommentar: wurde ursprünglich am 16.05.21 um 12:35 Uhr gepostet, wurde aber gefiltert, daher erneut um 13:05 gepostet.
- „Der Eagle“: Anzahl Likes gekauft: 342 (eingetroffen am 25.05.21)
- „Der Eagle“: Anzahl Likes final: 361

In dieser zweiten Untersuchungsphase wurde gezielt und intensiv mit Counter Speech durch die eingeweihten Jugendlichen im Forschungsteam reagiert und insgesamt fünf Hate-Profile (hara kujio, schni mk, Grim Grin, Count, M) eingesetzt. Es wurde dabei darauf geachtet, Counter Speech nicht als Gegenattacken zu formulieren, wenn auch in einzelnen Fällen aus einer Emotionalität heraus der aggressive Inhalt des Hate-Kommentars auch für leichtere Gegenangriffe genutzt wurde. Die Jugendlichen wurden dazu motiviert, eine möglichst große Bandbreite an Counter Speech Strategien einzusetzen, z.B. Online Hass klar benennen, das Opfer unterstützen, humorvoll/ironisch reagieren, Rückfragen stellen. In den Interaktionen gab es eine Person aus der realen Community, die ebenfalls mit Counter Speech reagierte, dabei aber nicht auf die anderen Counter Speaker*innen aus dem Forschungsteam einging oder darauf Bezug nahm. Counter Speech wurde durch diese reale Person sehr allgemein, tendenziell indirekt und nicht konfrontativ und inhaltlich eher intellektuell formuliert. Auch in der zweiten Reaktion bezugnehmend auf den entsprechenden Hate-Kommentar reagierte diese*r reale User*in, indem er*sie dem*der vermeintlichen Hater*in die intellektuellen Fähigkeiten abzusprechen.

Die Anzahl der gekauften Likes erhöhte sich beim positiv inszenierten Ursprungskommentar von 342 auf 361. Allerdings bleibt unklar, ob mit diesen Likes durch reale User*innen das Opfer gegen den Übergriff, oder ob lediglich das positive Feedback für den YouTuber unterstützt werden sollte. Außerhalb der Inszenierung gab es keine weiteren Hate Speech Kommentare oder indirekte Formen von Counter Speech in den Kommentarspalten.

Intervention 2b: Übergriff gegen YouTuber

Markierter Kommentar

hara kujio vor 2 Wochen
alter wenn ich mit dir wochenlang im urwald festsitzen würde dann würd ich auch nie wieder etwas von dir hören wollen absoluter cringelord
👍 170 🗨️ ANTWORTEN

🔼 23 Antworten ausblenden

Lea Fuchs vor 2 Wochen
Was genau soll daran cringe sein?
👍 🗨️ ANTWORTEN

Anna Maro vor 2 Wochen
Deine Aussage is ja mal totaler Blödsinn 😂
👍 🗨️ ANTWORTEN

hara kujio vor 2 Wochen
@Lea Fuchs ich würd mich schämen zuzugeben wenn mir jemand so lange aus dem weg geht alter sau cringe jetzt hier irgendwie kontakt aufnehmen zu wollen
👍 🗨️ ANTWORTEN

Emma Horak vor 2 Wochen
Hoffe du fühlst dich jetzt besser nach deinem unnötigen Kommentar
👍 1 🗨️ ANTWORTEN

Emma Horak vor 2 Wochen
@hara kujio wenn dir das Video nicht passt solltest du es dir einfach nicht anschauen und nicht unnötigen hass verbreiten!!
👍 1 🗨️ ANTWORTEN

hara kujio vor 2 Wochen
@Emma Horak danke mir gehts großartig und jetzt halts maul
👍 🗨️ ANTWORTEN

th0 vor 2 Wochen
@hara kujio schni mk :) 🤔🤔🤔
👍 🗨️ ANTWORTEN

Emma Horak vor 1 Woche
@hara kujio echt Respektlos.. bin echt enttäuscht wie viel Hass es im Internet gibt vielleicht wirst du das irgendwann mal verstehen man kann es dir nicht übel nehmen von deinem Profilbild und Namen würde ich vermuten das du echt jung bist nh🤔
👍 🗨️ ANTWORTEN

Zarah Varova vor 1 Woche
@Emma Horak Was genau soll jetzt daran cringe sein?
👍 🗨️ ANTWORTEN

Count vor 1 Woche
@th0 Du versuchst zu argumentieren aber hast offensichtlich einen Schlaganfall ich empfehle dir ein Krankenhaus und keine Youtube Comment Section in der du sinnlose Kommentare schreibst.
👍 🗨️ ANTWORTEN

Count vor 1 Woche
@Emma Horak Und was du redest darauf kann man nicht mal eingehen nur ein kompletter Versager findet nichts anderes an jemandem auszusetzen als Namen und Profilbild du redest von Hass und bist selber absolut keinen Deut besser.
👍 🗨️ ANTWORTEN

Sam Pollak vor 1 Woche
@hara kujio wieso so wütend? Man muss nicht immer gleich herablassend werden.
👍 🗨️ ANTWORTEN

hara kujio vor 1 Woche
@Sam Pollak du tust als hätte ich ihm gesagt er soll sich umbringen übertreib mal nicht
👍 🗨️ ANTWORTEN

Emma Horak vor 1 Woche
@Count danke du hast recht aber ich finde, dass Hass im Netz nichts zu suchen hat
👍 🗨️ ANTWORTEN

Count vor 5 Tagen
@Emma Horak
Was im Netz Platz hat und was nicht dürfen zum Glück keine dummen Gören wie du entscheiden, die Menschen ihre Meinung verbieten wollen.
👍 🗨️ ANTWORTEN

Emma Horak vor 5 Tagen
@Count schon mal was von Konstruktiver Kritik gehört??? Das kann man mit Hass-Kommentaren garnicht erst vergleichen. Jeder darf seine Meinung sagen aber nicht mit starken Ausdrücken die, die andere Person verletzen könnten!
„Dumme Göre“ genau sowas kann manche Mädchen echt runter machen mich zum Glück nicht aber darüber solltest du dir Gedanken machen🙏
👍 🗨️ ANTWORTEN

hara kujio vor 5 Tagen
@Emma Horak hast du echt nix besseres zu tun als eine woche lang unter einem kommentar rumzudiskutieren scheiß sjw halt mal dein maul ich brauch keine notifications mehr von dir
👍 🗨️ ANTWORTEN

Fortsetzung nächste Seite



Abbildung 7

Kontextinformationen zur Umsetzung:

- „hara kujio“: erster inszenierter Hate-Kommentar: am 16.05 um 12.40 Uhr
- Anzahl Likes gekauft: Nach Schwierigkeiten mit den Likes wurden schließlich 175 Likes gekauft (eingetroffen am 25.05.21)
- Anzahl Likes final: 166

Der erste Übergriff direkt gegen den YouTuber wurde in Phase 2 am 16.05.21 um 12.40 Uhr vom Fake-Account Hara Kujio gepostet. Nachdem die bestellten Likes bis zum 20.05 nicht sichtbar waren, wurde an diesem Tag ein weiteres Hassposting vom Fake-Account M gepostet (siehe nächste Seite) und erneut Likes gekauft. Durch einen notwendigen Anbieterwechsel wurde diese am 25.05.21 geliefert.

Bei diesem ersten Übergriff, der sich direkt an den YouTuber richtete, fand sich eine Counter Speech Reaktion jener Person aus der realen Community wieder, die bereits beim Übergriff gegen eine*n User*in kommentiert hatte. Der*die reale Counter Speaker*in reagierte durch Markierung des Hater*innen Profils des vorhergegangenen Experiments (Phase 2(1)) und verknüpft dies mit einem Zwinker-Smiley und Applaus-Emojis. Damit sollte offenbar mitgeteilt werden, dass durchschaut wurde, dass die Accounts hara kujio und schni mk von derselben Person betrieben wurden. In diesem Kommentarblock kam zur Verstärkung der negativen Kommentare noch ein weiteres inszeniertes Profil zum Einsatz (Count). Auch dieser Account reagierte direkt auf das Community-Mitglied, von diesem kam jedoch keine weitere Reaktion. Zwischen den inszenierten Hater*innen und den geschulten Counter Speaker*innen herrschte eine intensive Diskussion, jedoch ohne weitere Reaktion der Online-Community.

Es ist nicht restlos geklärt, aus welchen Gründen die Anzahl der Likes mit der Zeit abgenommen hat und was daraus geschlossen werden kann. Vorstellbar ist, dass YouTube diese, durch Bots generierten Fake-Likes, filtert und dementsprechend die Anzahl nach und nach etwas kleiner wird. Da im Vergleich zu anderen Plattformen wie Facebook auf YouTube nicht sichtbar ist, von welchen Accounts diese Likes

stammen, ist es nicht möglich, komplett transparent darzustellen, wie hoch die Anzahl der Fake-Likes und der echten Likes grundsätzlich ist, da wie beschrieben auch die Fake Likes keinen stabilen Referenzpunkt darstellen.

Intervention 2c: Weiterer Übergriff gegen YouTuber

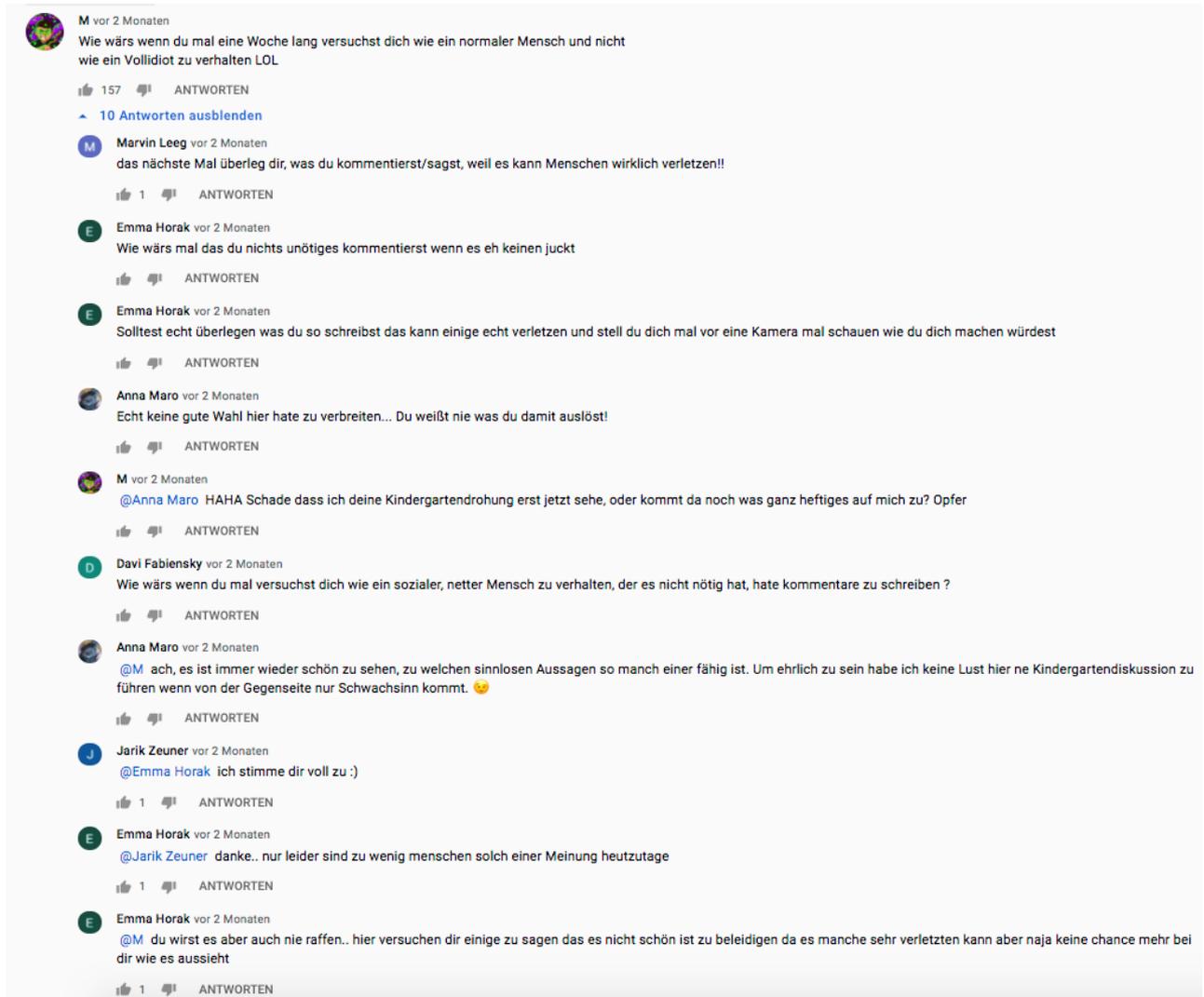


Abbildung 8

Kontextinformationen zur Umsetzung:

- „M“: zweiter inszenierter Hate-Kommentar: am 20.05.21 um 15.20 Uhr gepostet
- Anzahl Likes gekauft: 167 (eingetroffen am 25.05.21)
- Anzahl Likes final: 157 (160, dazwischen waren es mehr)

Im zweiten inszenierten Übergriff gegen den YouTuber ist es trotz Bemühungen mit intensiver Counter Speech inklusive beispielhafter gegenseitiger Unterstützung („ich stimme dir voll zu“) nicht gelungen, andere User*innen zum Kommentieren zu motivieren, d.h. es gab außerhalb des Forschungsteams keinerlei Reaktionen dazu (auch nicht indirekt; auch kein weiterer Hate Speech Kommentar außerhalb

des Forschungsteams). Die verringerte Anzahl an Likes ist auch hier nicht vollständig erklärbar, steht aber vermutlich in Zusammenhang mit dem Filtern von Fake-Likes durch YouTube.

Insgesamt ist für diese zweite Phase festzuhalten, dass die intensiven Counter Speech Bemühungen durch das Forschungsteam kaum bzw. praktisch nicht dazu beigetragen haben, Counter Speech durch andere, reale User*innen zu motivieren.

4.2.3 Zweite Intervention Hate Speech OHNE Counter Speech (Experiment-Phase 3)

Intervention 3a: Übergriff gegen User*in

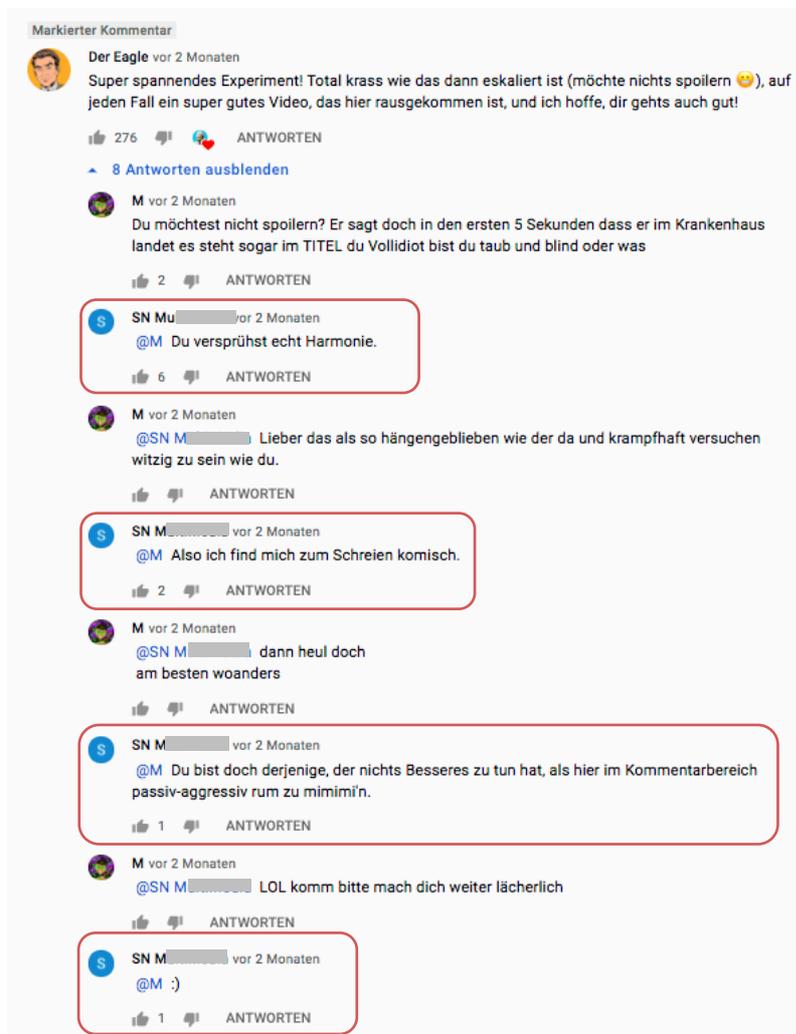


Abbildung 9

Kontextinformationen zur Umsetzung:

- rote Markierung: echter User*innen-Kommentar, nicht inszeniert
- „Der Eagle“: inszenierter User*innen-Kommentar: am 13.06.21 um 13.30 Uhr gepostet
- „M“: inszenierter Hate-Kommentar: am 13.06.21 um 13:30 Uhr gepostet

- „Der Eagle“: Anzahl Likes gekauft: 253 (eingetroffen am 15.06.21)
- Anzahl Likes final: 276

In dieser Phase 3 wurde nun ein weiteres Mal Hate Speech ohne gezielte Counter Speech durch das Forschungsteam inszeniert. Dabei gab es eine Person der realen Community, die als Counter Speaker*in auftrat: die Strategien waren hier eher humorvoll bzw. ironisch ausgerichtet, es gab insgesamt vier Interaktionsfolgen zwischen der*dem vermeintlichen Hater*in und der*dem realen Counter Speaker*in, wobei die letzte Intervention mit einem einfachen Smiley abgeschlossen wurde, womit signalisiert wurde, die Übergriffe nicht wirklich ernst zu nehmen. Zudem bot sich hier keine unmittelbar weitere Angriffsfläche mehr für einen weiteren Hate Speech Kommentar an.

Auffällig ist hier, dass zwar keine direkte Counter Speech in Form von Kommentaren aufscheint, sehr wohl aber einzelne Likes an den*die Counter Speaker*in vergeben wurden und damit diese*r zumindest indirekt unterstützt wurde. Der inszenierte Hate Speech Kommentar erhielt zwar ebenfalls leichte indirekte Unterstützung durch zwei Likes, wurde aber im weiteren Interaktionsverlauf nicht mehr unterstützt. Insgesamt scheint hier der Erfolg eher zugunsten von Counter Speech ausgefallen zu sein, da diese letztlich mehr Unterstützung erhielt und auch in der Interaktionskette als abschließender Kommentar aufscheint.

Intervention 3b: Übergriff gegen YouTuber



Abbildung 10

Kontextinformationen zur Umsetzung:

- „schni mk: inszenierter Hate-Kommentar: am 13.06.21 um 13:30 Uhr gepostet
- Anzahl Likes gekauft: 167 (eingetroffen am 15.06.21)
- Anzahl Likes final: 206

Beim inszenierten Übergriff gegen den YouTuber kam es allerdings zu keinerlei weiteren Interaktionen, auch nicht zu Counter Speech. Allerdings hat sich hier die Anzahl an Likes für Hate Speech deutlich erhöht, sodass hier wieder indirekte Unterstützung für Hate Speech vorkommt.

Um mögliche Interaktionen durch die Community anzuregen, wurde folgender zweiter Hate Speech Kommentar zwei Tage später gepostet:



Abbildung 11

Kontextinformationen zur Umsetzung:

- rote Markierung: echter User*innen-Kommentare, nicht inszeniert
- „Count“: inszenierter Hate-Kommentar: am 15.06.21 um 21:50 Uhr gepostet
- Anzahl Likes gekauft: 167 (eingetroffen am 17.06.21)
- Anzahl Likes final: 173 (zwischenzeitlich 177)

Dieses Mal traten zwei Personen der realen Community als Counter Speaker*innen auf, die in ihrem strategischen Vorgehen zunächst eher humorvoll / ironisch auftreten, abschließend wird der Übergriff auch klar als Beleidigung deklariert. Die beiden realen Counter Speaker*innen wurden auch durch Likes zumindest indirekt unterstützt, gleichzeitig erhielt aber auch der Ursprungs-Hate Kommentar Unterstützung durch Likes, nicht aber im Interaktionsverlauf mit den Counter Speaker*innen.

Bei der Analyse der Kommentarspalten konnten weder für den inszenierten Übergriff gegen eine*n vermeintliche*n User*in noch gegen den YouTuber weitere reale Hate Speech Kommentare außerhalb der Inszenierung durch das Forschungsteam festgestellt werden. Auch indirekte Formen von Counter Speech wurden nicht verwendet.

Insgesamt lässt sich festhalten, dass diese Phase 3 in Bezug auf Counter Speech bisher am erfolgreichsten war, da am meisten reale Counter Speech aus der Community vorzufinden waren. Möglicherweise haben auch die intensiven Counter Speech Bemühungen aus Phase 2 die etwas

stärkere Mobilisierung von realen User*innen in ihrer Rolle als Counter Speaker*innen beeinflusst. Nachdem die Mobilisierung insgesamt aber im bisherigen Verlauf deutlich geringer ausfiel als erwartet, sind solche Vermutung nur sehr vorsichtig zu interpretieren.

4.2.4 Zweite Intervention Hate Speech MIT Counter Speech (Experiment-Phase 4)

Intervention 4a: Übergriff gegen User*in

Markierter Kommentar:

Der Eagle vor 3 Wochen
Super spannend! Ich wünschte, ich hätte das Geld um sowas auch auszuprobieren... Danke, dass du dieses Video gemacht hast, dann fühle ich mich wenigstens ein bisschen als hätt ichs auch gemacht!
👍 280 🗨️ ANTWORTEN
11 Antworten ausblenden

Emma Horak vor 3 Wochen
Bin voll deiner Meinung das Video ist echt super 😊
👍 1 🗨️ ANTWORTEN

Emma Horak vor 3 Wochen
Wenn man so ein Video sieht hat man das Gefühl auch einen teil gemacht zu haben
👍 1 🗨️ ANTWORTEN

Anna Maro vor 3 Wochen
Ja total!!
👍 1 🗨️ ANTWORTEN

Count vor 3 Wochen
geh doch arbeiten und heul nicht rum
👍 🗨️ ANTWORTEN

Emma Horak vor 2 Wochen
@Count was meinst du?
👍 1 🗨️ ANTWORTEN

Emma Horak vor 2 Wochen
@Count ich finde, der Kommentar musste nicht sein ich meine @Der Eagle wollte nur positive Rückmeldung geben verstehe nicht was es daran auszusetzen gibt
👍 2 🗨️ ANTWORTEN

Marvin Leeg vor 2 Wochen
@Count wieso gleich so unsympathisch?
👍 2 🗨️ ANTWORTEN

Count vor 2 Wochen
@Emma Horak steht doch da was gibts da dumm nachzufragen lern lesen
👍 🗨️ ANTWORTEN

Cin... vor 2 Wochen
@Count Es gibt auch Menschen, die können nicht arbeiten und es gibt Menschen die gehen arbeiten und verdienen trotzdem nen Dreck. Oder Menschen, die ne Familie versorgen müssen... Willst du noch mehr Beispiele?
👍 4 🗨️ ANTWORTEN

Count vor 2 Wochen
@Cin... ja bitte verschwende noch mehr von meiner zeit
👍 🗨️ ANTWORTEN

Benni Koller vor 2 Wochen
@Count dude, dir ist hoffentlich klar, dass es vielleicht so manchen in deinem umfeld aehnlich geht, die art wie du so redest kann auch schaedlich fuer die in deiner unmittelbaren naehe scheiße sein
👍 2 🗨️ ANTWORTEN

Abbildung 12

Kontextinformationen zur Umsetzung:

- rote Markierung: echter User*innen-Kommentar, nicht inszeniert
- „Der Eagle“: inszenierter User*innen-Kommentar: am 11.07.21 um 14:05 Uhr gepostet
- „Count“: inszenierter Hate-Kommentar: am 11.07.21 um 14:05 Uhr gepostet
- „Der Eagle“: Anzahl Kommentare gekauft: 228 (eingetroffen am 12.07.21)

- „Der Eagle“: Anzahl Kommentare final: 280

In dieser Phase wurde zum zweiten Mal neben Hate Speech auch Counter Speech inszeniert. Es gab eine reale Person, die ebenfalls als Counter Speaker*in auftrat und von der inhaltlichen Ausrichtung her versucht hat, durch sachliches Argumentieren (aber auf eher angriffslustige Weise), dem Hate Kommentar entgegen zu treten. Es blieb allerdings bei einem einzigen Kommentar, eine weitere Reaktion der vermeintlichen Hater*in wurde nicht mehr beantwortet.

Im Vergleich zur Phase 2 (ebenfalls mit Counter Speech) wurden die inszenierten Interaktionen insgesamt etwas kürzer gehalten, mit der Zielsetzung, dadurch mehr Interaktionen durch die Community anzuregen, was allerdings nicht gelang.

Intervention 4b: Übergriff gegen YouTuber

The screenshot shows a YouTube comment thread. At the top, a comment by Grim Grin (3 weeks ago) is marked as 'Markierter Kommentar'. It contains a direct insult: 'Trendkleidung???? juuuunge du würdest RICHTIGEN style nichmal sehen wenn er dir in die fresse schlagen würde bitte lern erst was über style bevor du hier was dazu sagst peinlich'. This comment has 273 likes and 10 replies. The replies include:

- Emma Horak (3 weeks ago): 'Vielleicht denkst du bisschen besser nach für jeden ist "Trendkleidung" was anderes es gibt keine richtige antwort' (2 likes).
- Emma Horak (3 weeks ago): 'Finde diesen Kommentar echt unhöflich und unnötig von dir !!!' (4 likes).
- Grim Grin (3 weeks ago) replies to Emma Horak: '@Emma Horak junge absolut peinlich als wären Trends etwas das leute sich selbst ausdenken können was trend ist und was nicht ist nunmal FAKT egal was du dir einreden willst wenn man keine ahnung hat einfach die fresse halten'.
- Emma Horak (2 weeks ago) replies to Grim Grin: '@Grim Grin erstens solltest du mal nachdenken wie du mit Menschen redest. Ob ich meine "Fresse" halte oder nicht kann immer noch ich selbst entscheiden!! Vielleicht ist dir das nicht bewusst, aber ein Trend wird durch 1 Menschen gesetzt einer macht es vor und die anderen nach weil die selber keine anderen Ideen haben!'.
- Sam Pollak (2 weeks ago) replies to Grim Grin: '@Grim Grin Meinungen sind Keine Fakten, nur weil dir individualität nicht in dein Weltbild passt musst du dich hier nicht über soetwas aufregen'.
- Davi Fabiinsky (2 weeks ago, edited): 'vielleicht solltest du erstmal lernen, Leute nicht niederzumachen, nur weil sie sich anders kleiden oder einen anderen Geschmack haben?! Der Kommentar ist einfach absolut respektlos und unnötig!'.

Below this thread, a red-bordered box highlights a comment by Ika (2 weeks ago, edited): 'Diese Trendkleidung hat ihm eine Designerin ausgesucht. Ich frage mich allerdings, was in deinen Augen denn "RICHTIGER" Style ist, wovon du hier redest. Offenbar scheinst du es ja zu wissen...Oo'. This comment has 1 like and 1 reply.

The reply from Grim Grin (2 weeks ago) to Ika is: '@Ika Mir egal ob sie Designerin ist es sieht trotzdem aus wie New yorker 2016 und hat absolut 0 mit streetstyle, high fashion oder sonst was zu tun einfach peinlich'. This reply also has 1 like and 1 reply.

Below this, another red-bordered box highlights a comment by Ika (2 weeks ago) to Grim Grin: '@Grim Grin Also, mir persönlich gefällt es auch nicht, aber genauso wenig finde ich diesen ganzen high fashion Kram etc. nicht sehr stylish. Dementsprechend frage ich mich, was macht dich zu Experten? Kommst du aus der Modebranche, dass du weißt, was Style ist und was nicht?'. This comment has 1 like and 1 reply.

The final reply from Grim Grin (2 weeks ago) to Ika is: '@Ika man muss nicht "aus der Modebranche kommen" um zu verstehen dass das hier Bullshit ist'. This reply has 1 like and 1 reply.

Abbildung 13

Kontextinformationen zur Umsetzung:

- rote Markierung: echter User*innen-Kommentare, nicht inszeniert
- „Grim Grin“: inszenierter Hate-Kommentar: am 11.07.21 um 14:05 Uhr gepostet
- Anzahl Kommentare gekauft: 168 (eingetroffen am 12.07.21)
- Anzahl Kommentare final: 273

Beim inszenierten Übergriff gegen den YouTuber gab es ebenfalls eine reale Person, die als Counter Speaker*in mit insgesamt zwei Kommentaren auftrat. Diese versuchte ebenfalls, eher sachlich inhaltlich dagegen zu argumentieren und das Verhalten der*des vermeintlichen Hater*in in Frage zu stellen, was in der Herangehensweise der Argumentationsweise der inszenierten Counter Speaker*innen sehr ähnlich war. Die*der reale Counter Speaker*in klinkte sich im Auftreten sozusagen in eine Art Diskussion ein, fand dabei aber weder konkrete Unterstützung durch andere Counter Speaker*innen noch durch Likes.

Im Gegensatz dazu hat sich die Anzahl an Likes für den Hate Kommentar erhöht. Außerdem fand sich in der Analyse der weiteren Kommentare unter dem Video folgender, realer Hate Speech Kommentar, dass allerdings von niemanden unterstützt oder beantwortet wurde – auch nicht durch Counter Speech.

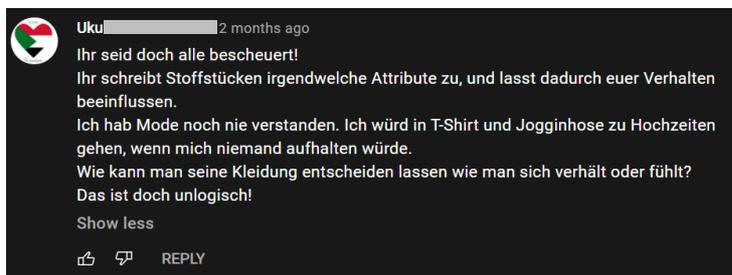


Abbildung 14

Insgesamt scheint die Erfolgs-Bewertung hier wieder eher zugunsten von Hate Speech auszufallen. Es ist trotz Counter Speech Bemühungen zwecks Mobilisierung anderer zur Gegenrede wieder nur in Einzelfällen gelungen, auch reale User*innen zu Counter Speech zu mobilisieren.

4.2.5 Dritte Intervention Hate Speech OHNE Counter Speech (Experiment-Phase 5)

Intervention 5a: Übergriff gegen User*in

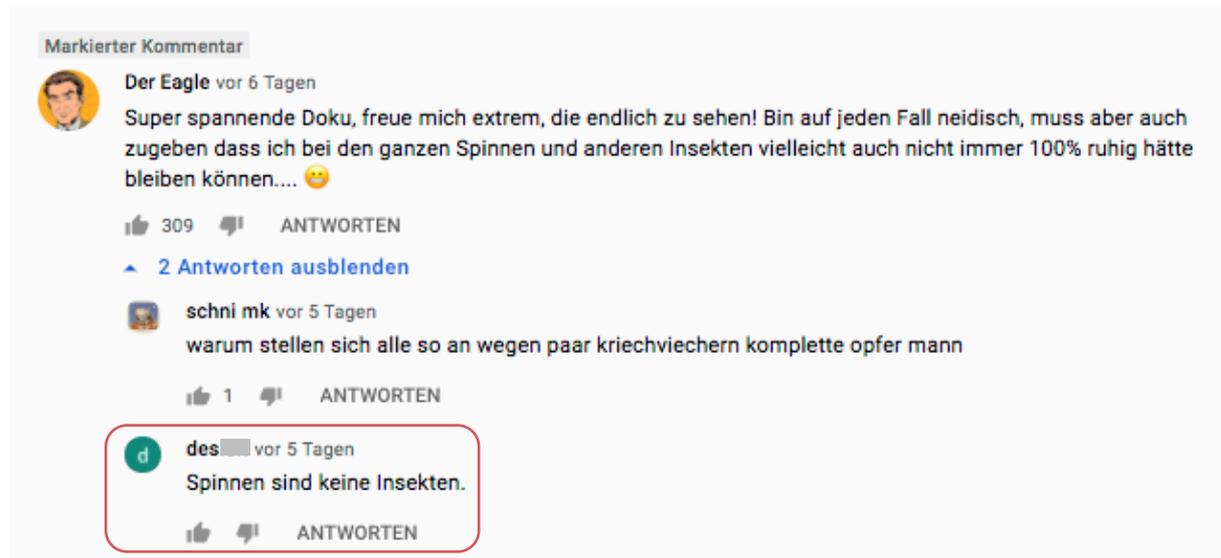


Abbildung 15

Kontextinformationen zur Umsetzung:

- rote Markierung: echter User*innen-Kommentar, nicht inszeniert
- „Der Eagle“: inszenierter User*innen-Kommentar: am 16.08.21 um 21:25 Uhr gepostet
- „schni mk“: inszenierter Hate-Kommentar: am 16.08.21 um 21:25 Uhr gepostet
- „Der Eagle“: Like gekauft: 269 (eingetroffen am 18.08.21)
- „Der Eagle“: Anzahl Likes final: 285

In Phase 5 wurde wieder ausschließlich Hate Speech inszeniert. Es gab zwar eine reale Person, die kommentiert hat, diese hat jedoch kein Counter Speech betrieben, sondern den Übergriff sogar zusätzlich eher verstärkt – wenn auch keine besonders abfälligen Äußerungen getätigt wurden.

Intervention 5b: Übergriff gegen YouTuber

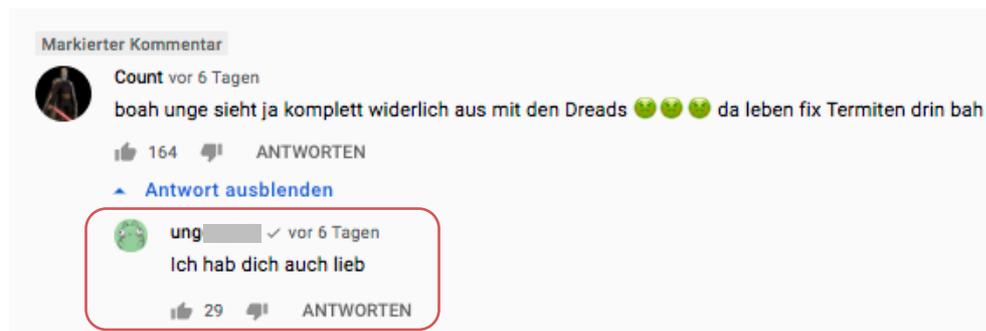


Abbildung 16

Kontextinformationen zur Umsetzung:

- rote Markierung: echter User*innen-Kommentar, nicht inszeniert: stammt hier vom Opfer selbst = bekannter YouTuber „unge“, der im Video vorkommt
- „Count“: inszenierter Hate-Kommentar: am 16.08.21 um 21:25 Uhr gepostet; ursprünglicher Hass-Kommentar wurde gefiltert, unmittelbar danach wurde ein neues gepostet
- Anzahl Kommentare gekauft: 168 (eingetroffen am 18.08.21)
- Anzahl Kommentare final: 164

Beim inszenierten Übergriff gegen den YouTuber gab es keine Reaktion von realen User*innen, aber vom Opfer selbst – und zwar ein bekannter anderer YouTuber, der eher humorvoll/ironisch reagierte. Dieser Kommentar erhielt auch einige Likes, konkrete Kommentare blieben aber überraschenderweise aus.

Bei der Analyse aller weiteren Kommentare unter diesem Video fanden sich weder weitere Hate-Kommentare noch indirekte Counter Speech. Die Verringerung um vier Likes lässt sich hier vermutlich wieder auf das Herausfiltern von Fake-Likes durch YouTube zurückführen.

Insgesamt ist es damit in Phase 5 nicht gelungen, Personen zu Counter Speech zu mobilisieren, sondern im Gegenteil wurde sogar eher Hate Speech verstärkt.

4.2.6 Dritte Intervention Hate Speech MIT Counter Speech (Experiment-Phase 6)

Intervention 6a: Übergriff gegen User*in

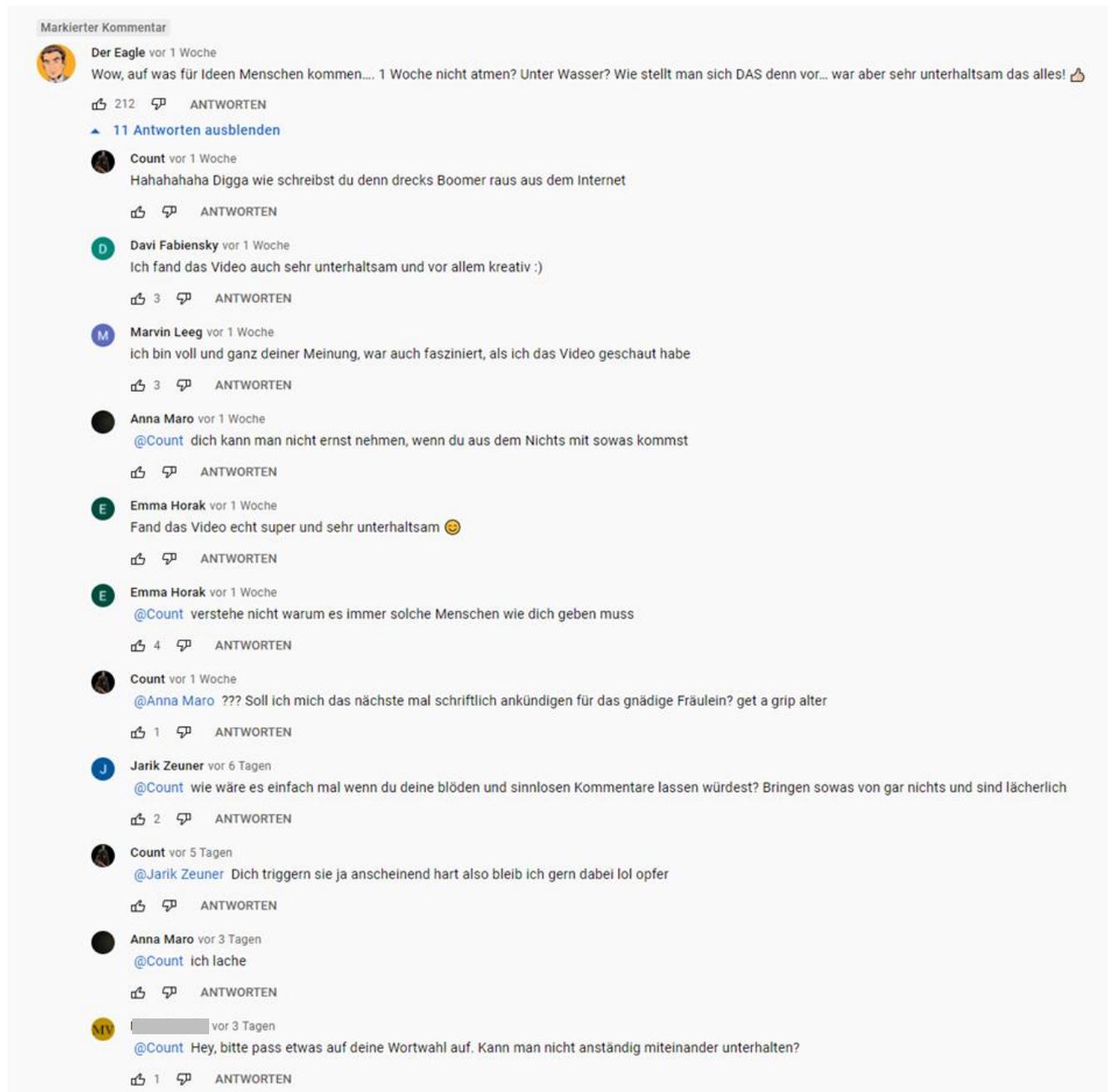


Abbildung 17

Kontextinformationen zur Umsetzung:

- „Der Eagle“: inszenierter User*innen-Kommentar: am 05.09.21 um 15:00 Uhr gepostet
- „Count“: inszenierter Hate-Kommentar: am 05.09.21 um 15:00 Uhr gepostet
- „Der Eagle“: Anzahl Likes gekauft: 200 (eingetroffen am 06.09.21)
- „Der Eagle“: Anzahl Likes final: 212

Auch in der 6. und damit letzten Phase des Online Feld-Experiments gelang es trotz Counter Speech Bemühungen durch das Forschungsteam scheinbar nicht mehr, reale User*innen zur Gegenrede gegen Hate Speech zu mobilisieren. Es fanden insgesamt keine direkten Interaktionen mit anderen Personen außerhalb des Feld-Experiments statt. Allerdings zeigte die Analyse der übrigen Kommentare, dass zumindest indirekte Formen von Counter Speech gepostet wurden (siehe dazu weiter unten).

Intervention 6b: Übergriff gegen YouTuber

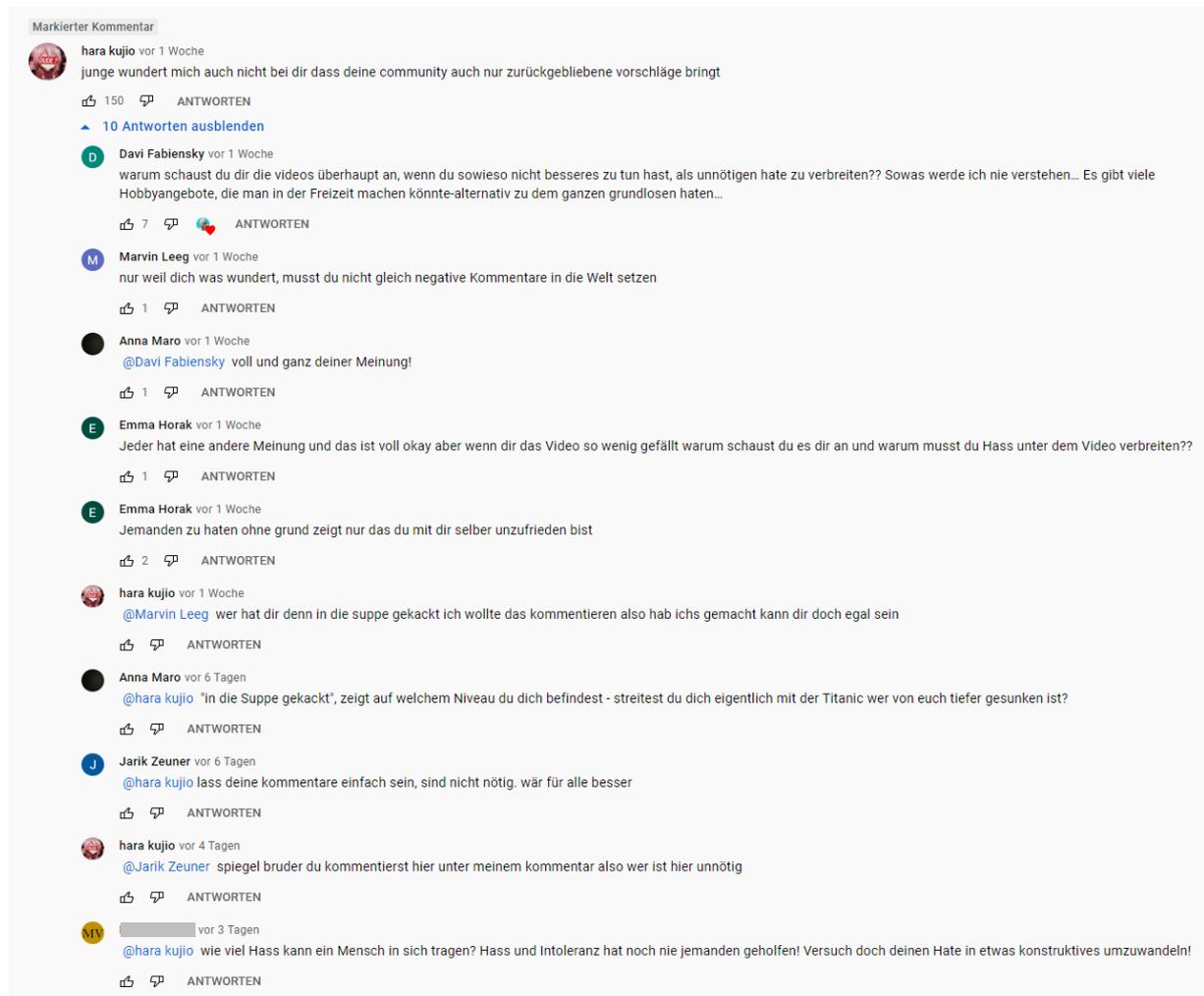


Abbildung 18

Kontextinformationen zur Umsetzung:

- „hara kujio“: inszenierter Hate-Kommentar: am 05.09.21 um 15:00 Uhr gepostet
- Anzahl Likes gekauft: 150 (eingetroffen am 06.09.21)
- Anzahl Likes final: 150

Auch beim inszenierten Übergriff den YouTuber gab es scheinbar keine Reaktionen mehr durch andere, reale User*innen.

Nach kurzer Zeit allerdings bemerkte das Forschungsteam, dass die inszenierten Hasskommentare teilweise nicht mehr sichtbar waren. Der Grund dafür waren vermutlich durch YouTube veranlasste Shadowbans: Das bedeutet, dass bestimmte Beiträge von spezifischen Accounts weniger bzw. nicht mehr sichtbar sind, ohne den entsprechenden User*innen Account zu sperren, zu benachrichtigen oder vorzuwarnen. Shadowbans sind also Einschränkungen, die von User*innen möglichst nicht bemerkt werden sollen. Aufgrund der mangelnden Sichtbarkeit wurden folgende neuen Hate Speech Kommentare mit neuem Account gepostet, wobei die inhaltliche Formulierung der Hate Speech Kommentare abgeschwächt werden musste, um die Sichtbarkeit zu gewährleisten.



Abbildung 19

Kontextinformationen zur Umsetzung:

- „Grim Grin“: inszenierter Hate-Kommentar: am 09.09.21 um 15:00 Uhr gepostet
- Anzahl Likes gekauft: 25 (eingetroffen am 09.09.21)
- Anzahl Likes final: 26



Abbildung 20

Kontextinformationen zur Umsetzung:

- „schni mk“: inszenierter Hate-Kommentar: am 09.09.21 um 19:00 Uhr gepostet
- Anzahl Kommentare gekauft: 0
- Anzahl Kommentare final: 1

Bei allen inszenierten Übergriffen inklusive Counter Speech durch das Forschungsteam gelang es offenbar wieder nicht, andere User* zu motivieren, ebenfalls zu kommentieren.

Bei der Analyse aller Kommentarspalten unter diesem Video fanden sich allerdings folgende beiden indirekten Counter Speech Kommentare:



Abbildung 21

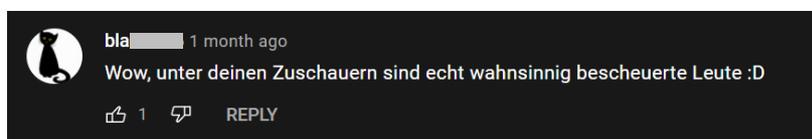


Abbildung 22

Nachdem unter diesem Video ausschließlich vom Forschungsteam inszenierte Hate Speech vorkommt, ist es sehr wahrscheinlich, dass sich diese Kommentare als Reaktion darauf beziehen. Counter Speech hat sich demnach von direkten Interaktionen auf diese indirekte Ebene verlagert. Diese Vermutung wird auch durch die offensichtlichen Shadowbans unterstützt: Es ist anzunehmen, dass reale User*innen die Hate Kommentare (bzw. die Profile, von denen Hate Speech ausging) gemeldet haben, anstatt konkrete Counter Speech Kommentare zu posten und daher YouTube mit diesen Restriktionen reagiert hat.

4.3 Zusammenfassung und Schlussfolgerungen

Der folgende Überblick bietet eine kompakte Zusammenfassung der Ergebnisse des Online Feld-Experiments:

- Phase 1: Inszenierte Hate Speech ohne Counter Speech
Bei allen inszenierten Übergriffen reagiert zwar zumindest ein*e reale Counter Speaker*in, insgesamt überwiegt aber der Anteil an Hate Speech bzw. wird diese sogar eher verstärkt, wenn auch in einem verhältnismäßig geringen Ausmaß.
- Phase 2: Inszenierte Hate Speech mit inszenierter Counter Speech
Insgesamt gab es bei allen inszenierten Angriffen nur eine Person, die mit Counter Speech reagiert hat – wenn auch zumindest drei Mal. Trotz der intensiven Counter Speech Bemühungen durch das Forschungsteam ist es damit aber nicht gelungen, auch andere, reale User*innen zu aktiver Gegenrede zu mobilisieren.
- Phase 3: Inszenierte Hate Speech ohne Counter Speech
Im Vergleich zu allen übrigen Forschungsphasen war diese in Bezug auf Counter Speech am erfolgreichsten, d.h. es konnte die meisten Reaktionen verzeichnet werden: Es reagierten insgesamt drei reale Counter Speaker*innen auf die inszenierten Übergriffe, zwei davon auch mehrmals (d.h. eine Person 4x, die zweite Person 2x, die dritte 1x). Möglicherweise zeigten hier die Counter Speech Bemühungen aus Phase 2 Wirkung, allerdings ist dies aufgrund der geringen Anzahl nur mit Vorsicht zu interpretieren.
- Phase 4: Inszenierte Hate Speech mit inszenierter Counter Speech
Bei allen inszenierten Übergriffen reagierten zwei Personen mit Counter Speech, eine davon nur einmal, die zweite zwei Mal. Allerdings gab es auch außerhalb der durch das Forschungsteam inszenierten Hasskommentare ein realer Hate Speech Kommentar durch eine*n andere*n User*in, das aber unkommentiert blieb, sodass insgesamt trotz inszenierter Counter Speech wieder mehr Hate Speech als reale Counter Speech zu finden war.
- Phase 5: Inszenierte Hate Speech ohne Counter Speech
Es gab bei allen Übergriffen zwei reale Personen, die einmalig darauf reagiert haben: Allerdings hat die erste Person Hate Speech sogar eher verstärkt, die zweite Person war das Opfer des Übergriffs, ein im Video genannter bekannter YouTuber. Dieser erhielt zwar Likes als Unterstützung, es kamen aber keine weiteren Kommentare dazu.
- Phase 6: Inszenierte Hate Speech mit inszenierter Counter Speech
In dieser letzten Phase gab es keine einzige reale Person, die auf die inszenierten Übergriffe bzw. inszenierte Counter Speech reagiert hat. Allerdings fanden sich zwei indirekte Counter Speech Kommentare, in denen das unangebrachte Verhalten thematisiert wurde. Das Forschungsteam hatte zudem mit Shadowbans zu kämpfen, sodass anzunehmen ist, dass die benutzten Hate-Accounts gemeldet wurden. Counter Speech scheint sich auf eine eher indirekte Ebene (Nutzung der Meldefunktion) ohne direkte Konfrontation verlagert zu haben.

Aus diesen Ergebnissen lassen sich insgesamt nur vorsichtige Schlüsse ziehen: Sehr deutlich zeigt sich, dass die Mobilisierung von User*innen zu realer Counter Speech durch inszenierte Gegenrede sehr schwer umzusetzen ist. Effekte waren erst in der Folgephase zu sehen und sind als eher kurzfristig zu werten. Letztlich hat sich Counter Speech in der letzten Runde dieses Experiments auf eine indirekte Ebene ohne direkte Konfrontation verlagert, das heißt, die Nutzung der Meldefunktion und indirektes Verfassen von Kommentaren, die nicht als direkte Antwort formuliert wurden.

Insgesamt war in der untersuchten Community wenig Hate Speech auffindbar und damit auch wenig Counter Speech möglich. Damit bestätigen sich die bisherigen Erkenntnisse zu Counter Speech durch Jugendliche, dass nämlich direkte Gegenreaktionen und damit für alle sichtbare Counter Speech weitgehend vermieden wird.

Festzuhalten ist außerdem, dass die Counter Speech Reaktionen der realen Community insgesamt sehr gering ausfielen. Neben der scheinbar geringen Motivation zu Counter Speech könnten folgende weitere Gründe für die geringe Anzahl an Reaktionen vermutet werden:

- 1) In jenen Runden, wo neben Hate Speech auch Counter Speech durch das Forschungsteam inszeniert wurde, waren die Interaktionsketten teilweise sehr lang. Es ist möglich, dass die eigentliche Community sogar davon abgehalten wurde zu reagieren, weil die Counter Speech Interaktion ohnehin schon von anderen übernommen wurden. Insgesamt könnte so der Anreiz für andere potenzielle Counter Speaker*innen zu gering gewesen sein, insbesondere dann, wenn der letzte Kommentar in der Interaktionskette einen Counter Speech Kommentar darstellte.
- 2) Die intensiven Counter Speech Bemühungen durch das Forschungsteam wurden von manchen User*innen vielleicht als ungewöhnlich empfunden, wodurch eventuell auch Skepsis im Umgang damit entstanden ist, was diese ebenfalls von weiteren Interventionen abhalten hätte können.
- 3) Die Einschränkungen bei der konkreten Ausformulierung der Hate Speech Kommentare, die durch die kanalspezifischen (Joseph DeChangemans YouTube Kanal) und plattformspezifischen (YouTube) Filter gegeben waren, führten möglicherweise dazu, dass diese als „nicht schlimm“ genug interpretiert wurden: Es kam mehrmals vor, dass bestimmte Schimpfwörter automatisch von den YouTube Filtern gelöscht wurden. Um dies zu umgehen, wurden vermeintliche Tippfehler in die Kommentare eingebaut oder ein Teil des Wortes zensiert, beispielsweise „du hruensohn“ oder „du hu*****“ statt „du hurensohn“. So musste beim Verfassen neuer Hasskommentare stets darauf geachtet werden, die richtige Balance zwischen „schwerwiegend genug, um eine Reaktion zu provozieren“ und „nicht zu schwerwiegend, um keinen Shadowban oder ein Löschen durch YouTubes Filter zu provozieren“ zu finden.
- 4) Es ist anzunehmen, dass einige Community-Mitglieder mit der Nutzung der Melde-Funktion reagiert haben, anstatt direkt einen Counter Speech Kommentar zu posten: Dies ist vor allem auch deshalb naheliegend, da einige der vom Forschungsteam benutzen Accounts zur Inszenierung von Hate Speech am Ende des Feld-Experiments augenscheinlich einem Shadowban ausgesetzt waren. Es ist also anzunehmen, dass User*innen anstelle von Counter Speech als Gegenmaßnahme möglicherweise die Kommentare oder Profile von denen Hatespeech ausging, gemeldet haben und YouTube folglich vermutlich mit Shadowbans reagiert hat.

- 5) Bei der Analyse der Kommentarspalten der einzelnen Videos wurden teilweise auch indirekte Formen von Counter Speech gefunden. Dadurch ist es naheliegend, dass die inszenierten Hate Speech Kommentare zwar gesehen wurden, aber Counter Speech als direkte Reaktion offenbar nicht unbedingt die Regel ist.
- 6) Die Erhöhung der Sichtbarkeit von Hate und Counter Speech durch zugekaufte Likes erhöht nicht unbedingt die Bereitschaft zu Counter Speech. In der Regel wurden die Likes direkt nach der Inszenierung der Kommentare gekauft, da die meisten Personen ein Video innerhalb der ersten 48 Stunden anschauen. Zwar trafen bei den ersten Videos die zugekauften Likes oft erst sehr spät ein, was mitunter ein Grund für das geringe Ausmaß an Counter Speech gewesen sein könnte, allerdings gab es auch bei solchen Durchläufen, wo die Likes relativ früh eintrafen (vor allem in späteren Runden) insgesamt sehr wenig Counter Speech.

Die Frage, ob eine erhöhte Sichtbarkeit von Counter Speech, wie das beispielsweise auch in automatisierter Form inszeniert werden könnte, tatsächlich dazu beitragen kann, User*innen zu realer Counter Speech zu motivieren, kann auf Grundlage des durchgeführten Online Feld-Experiments nicht eindeutig beantwortet werden. Es scheint aber zumindest Potenzial für kurzzeitige (Folge-)Effekte zu geben bzw. ist es möglich, User*innen verstärkt zu indirekten Reaktionen (z.B. Nutzung der Meldefunktion) zu animieren. Eine mögliche Idee, automatisierte Counter Speech Reaktion auf identifizierte Hate Speech zu forcieren, müsste zudem sehr behutsam umgesetzt werden, um keine Skepsis hervorzurufen bzw. könnte vermehrte Counter Speech auch dazu führen, dass es so wirkt, als ob sich ohnehin schon andere darum kümmern und der eigene Einsatz damit nicht mehr erforderlich sei (Verantwortungsdiffusion). Eine direkte Aufforderung von User*innen zu Counter Speech könnte hier vermutlich wirksamer sein.

Unterstützungsmaßnahmen der Praxispartner*innen zur Förderung von Counter Speech

Die Projektpartner*innen ÖIAT/Saferinternet.at, Mauthausen Komitee Österreich, das Büro 1.6 Kriminalprävention und Opferhilfe des Bundeskriminalamts/BM.I und das Bundeskanzleramt / Sektion VI: Familie und Jugend /REFERAT VI/5/a: Kompetenzzentrum Jugend wurden während des gesamten Forschungsprozesses aktiv in die Konzeption und Umsetzung eingebunden, Teilergebnisse wurden in gemeinsamen Treffen regelmäßig diskutiert und ausgetauscht. Auf diese Weise wurde sichergestellt, dass die Praxispartner*innen die Studie in jeder Phase mit ihrem Wissen unterstützen und bereichern, dass sie ausreichend in das Projekt involviert sind und über entsprechendes Hintergrundwissen zur Studie verfügen.

Ziel war schließlich, dass die Projektpartner*innen die Studienergebnisse konkret für die Adaptierung bestehender und/oder für die Entwicklung neuer Trainings- bzw. Schulungskonzepte und Informationsmaßnahmen einsetzen können. Da die beteiligten Projektpartner*innen zentrale Positionen für Zivilcourage bzw. für den sicheren Umgang von digitalen Medien durch Jugendliche in einem österreichweiten und internationalen Netzwerk einnehmen, ist auch ein hohes und breites Nutzungs- und Einsatzpotenzial vorhanden:

- Das **Österreichische Institut für Angewandte Telekommunikation / Saferinternet.at (ÖIAT)** ist Mitglied im ACR, dem Verband der kooperativen Forschungseinrichtungen Österreichs sowie in zahlreichen nationalen und internationalen Gremien vertreten (Beirat Informationsgesellschaft des Bundeskanzleramts, Beirat Stoptline, Steuerungsgruppe des INSAFE-Netzwerks der EU-Kommission etc.). Das ÖIAT bietet kostenlose Broschüren, Folder, Videos und Unterrichtsmaterialien zu verschiedenen Themen sowie zielgruppengerechte und individuell gestaltbare Workshops und Vorträge in ganz Österreich an. Bisher wurden über 300 Informationsmaterialien erstellt und über 5 Mio.-fach digital und analog verbreitet, ca. 19.000 Workshops und Vorträge abgehalten und damit rund 450.000 Personen erreicht. Zusätzlich wurden zahlreiche Aktionen und Events, wie etwa der jährliche internationale Safer Internet Day umgesetzt, an dem sich in Österreich jedes Jahr mehr als 200 Schulen, Unternehmen & Institutionen mit eigenen Projekten beteiligen. Über die in sozialen Netzwerken und auf der Website regelmäßig zur Verfügung gestellten Informationen werden monatlich rund 325.000 Personen erreicht.
- Das **Mauthausen Komitee Österreich** führt seit April 2010 mit Jugendlichen Zivilcourage-Trainings durch, die sich an Jugendliche ab der 10. Schulstufe richten und in Schulen, Lehrwerkstätten, Jugendzentren, etc. in ganz Österreich durchgeführt werden. Seit drei Jahren wird im Rahmen der

Trainings auch das Schwerpunktthema "Online Zivilcourage" angeboten, das aufgrund der großen Nachfrage stetig weiter ausgebaut wird. Das MKÖ verfügt über ein österreichweites Netzwerk zur Vermarktung und Vermittlung der Trainings. Die zentralen Ziele des Zivilcourage-Trainings sind die Erhöhung der Sensibilität für zivilcouragiertes Handeln, die Erweiterung des eigenen Verhaltensrepertoires auf unterschiedlichen Ebenen, sowie die Vermittlung von Zivilcourage und gesellschaftlicher Verantwortung. Ein wichtiger Bestandteil der Zivilcourage-Trainings des Mauthausen Komitees Österreich ist die Erarbeitung von Handlungskompetenzen, sowie die Stärkung der Artikulations- und Argumentationsfähigkeit, der produktiven Konfliktfähigkeit, Selbstsicherheit, Selbstwirksamkeit, Reflexionsfähigkeit und die Ermutigung zum freiwilligen Engagement. Die Zivilcourage-Trainings bieten den Teilnehmer*innen einen geschützten Raum, in dem sie die Möglichkeit haben, neue Verhaltensweisen auszuprobieren und zu reflektieren mit dem Ziel, Sicherheit im Umgang mit schwierigen Situationen zu gewinnen. Seit Beginn des Projekts wird die Nachfrage nach Zivilcourage-Trainings zunehmend größer, bis jetzt konnte mit mehr als 80.000 Jugendlichen Zivilcourage trainiert werden.

- Das **Büro 1.6 Kriminalprävention und Opferhilfe des Bundeskriminalamts** führte bis zum Beginn der Corona-Pandemie laufend – und seither entsprechend eingeschränkt - Programme zur Kriminalprävention mit der Zielgruppe Jugendliche - UNDER18 - durch. Durch die österreichweit standardisierte Ausbildung und Umsetzung für die Zielgruppe 13- bis 17-Jährige kann eine nachhaltige Präventionsarbeit geleistet werden. Diese Aufgabe leisten österreichweit derzeit insgesamt mehr als 430 Präventionsbedienstete. In Zusammenhang mit dem vorliegenden Forschungsprojekt ist insbesondere das Projekt Click & Check von Interesse, das von der Polizei österreichweit für Schulen und Schüler*innen angeboten wird: Ziel ist die Förderung eines verantwortungsvollen Umgangs mit digitalen Medien und die Erweiterung bzw. Verinnerlichung von Handlungsstrategien in sozialen Netzwerken. Dabei werden Themen wie Chats und Soziale Netzwerke, Cybermobbing, Hate Speech, ermutigen zu Digitalcourage oder auch Themen wie Grooming und Sexting gemeinsam mit den Jugendlichen erarbeitet. Neben der Stärkung des Rechtsbewusstseins der Jugendlichen, soll auch deren Empathiefähigkeit gefördert werden. Neben Kindern und Jugendlichen werden auch das Lehrpersonal, die Erziehungsberechtigten in die Umsetzung der Kriminalprävention mit der Zielgruppe Jugendliche eingebunden, um das Qualitätskriterium des Mehrebenenansatzes zu erreichen. Noch 2019 wurden mit Click & Check österreichweit rund 3.000 Schulungsmaßnahmen umgesetzt und mehr als 70.000 Schüler*innen, Erziehungsberechtigte und Lehrer*innen mit dem Präventionsprogramm „Click & Check“ erreicht.
- Das **Bundeskanzleramt/BKA / Sektion VI: Familie und Jugend /REFERAT VI/5/a: Kompetenzzentrum Jugend** ist für die Umsetzung und Weiterentwicklung der Österreichischen Jugendstrategie, Prävention und Gesundheitsförderung im Bereich Jugend verantwortlich. Die Projektergebnisse sind dabei insbesondere für die Koordination mit der Beratungsstelle Extremismus und des Komitees No

Hate Speech in Österreich von Bedeutung. Die Ergebnisse können zudem in die Agenden der Jugendforschung eingebracht werden. Die bestehende enge Zusammenarbeit mit der Bundes-Jugendvertretung, dem bundesweiten Netzwerk Offene Jugendarbeit sowie dem Bundesnetzwerk Österreichische Jugendinfos kann zudem für die Dissemination der Ergebnisse eingesetzt werden. Das BKA ist Fördergeber und Vernetzungspartner von Jugendarbeitsstrukturen in Österreich und dient somit als Schnittstelle.

Entwickelte Maßnahmen

Aufbauend auf den Studienergebnissen haben die Praxispartner*innen konkrete Maßnahmen ausgearbeitet, um die Ergebnisse in die Praxis zu transferieren. Diese werden im Folgenden vorgestellt:

ÖIAT/Saferinternet.at

1) Konzeption und Umsetzung eines Saferinternet-Quiz für 8- bis 12-Jährige:

Das ÖIAT war aktiv im Forschungsprojekt involviert und hat insbesondere auch die Aktionsforschungsgruppen mitkonzipiert und mitbegleitet. In diesem Rahmen wurden auch viele Diskussionen mit Jugendlichen geführt, wodurch ein bereichernder Austausch zwischen Jugendlichen als Expert*innen ihrer (Online-)Lebenswelt und erwachsenen Fachexpert*innen möglich wurde. Im Zuge dieser intensiven Auseinandersetzung wurde ein Aspekt herausgegriffen, der bislang zu wenig Beachtung fand: Wie nämlich durch das Vorleben / die „Vorbildwirkung“ Erwachsener nun auch ganz junge Internetnutzende sich Hasspostings und Hassrede zu eigen machen. Die Jugendlichen machten in ihren Diskussionen deutlich, dass es vor allem auch schon jüngeren Kindern an ganz konkreten Ideen fehlt, wie sie gegen Hass im Netz vorgehen könnten und welche Optionen es für sie eigentlich gäbe. Dies wurde zum Anlass genommen, um aus der sehr beliebten Reihe von Saferinternet-Quizzes für die Altersgruppe der 8-bis 12-Jährigen ein entsprechendes Quiz zu entwickeln, das als kostenloses Angebot bereits vielfach genutzt wird: <https://siat.involve.me/gut-gekontert>. Dabei wurde darauf geachtet, sich sowohl auch sprachlich, als auch inhaltlich dieser Altersgruppe anzupassen. Die folgenden Abbildungen zeigen beispielhafte Ausschnitte aus diesem Quiz:



Abbildung 1

2) Entwicklung eines Unterrichtsmaterials für Lehrende für die Zielgruppe 9- bis 12-Jährige:

Um Lehrenden eine Hilfestellung zu bieten, wie solche Fragen im Unterricht aufgegriffen werden können, wurde ein umfassendes Unterrichtsmaterial für Lehrende entwickelt. Ziel war es, Übungen zu ausarbeiten, damit Kinder ein Gefühl entwickeln können, was im Internet ok ist und was nicht. Dazu gehört es auch, gegen Gemeinheiten einschreiten zu können. Das Unterrichtsmaterial „Hass im Netz kompetent entgegnen - Wir werden zu Cyber-Heroes: Wir unternehmen etwas, anstatt zuzuschauen und mitzumachen!“ besteht aus zahlreichen Übungen, die thematisch/inhaltlich aufbereitet werden, sowie mehreren Arbeitsblättern und Tipps für Kinder/Jugendliche und Pädagog*innen. Es knüpft inhaltlich an den Lehrplan zur Digitalen Grundbildung an, wo Medienkompetenz, Digitale Kommunikation und Umgang mit Social Media vermittelt werden sollen. Im entwickelten Unterrichtsmaterial werden Themen behandelt wie etwa Sensibilisierung für passende und unpassende Postings, Erarbeitung eines Handlungsrepertoires, um bei problematische Situationen im Internet nicht zu ignorieren, sondern zu handeln, Stärkung der Selbstwirksamkeit im Internet, Erklärungen und Entwicklung eines Verständnisses, wie Hate Speech entsteht, Entwicklung von Counter Speech Kompetenzen, Vermittlung der Wichtigkeit zur Mobilisierung anderer zur Unterstützung und zur Nutzung der Meldefunktion. Auf den folgenden Seiten finden sich beispielhaft entwickelte Arbeitsblätter aus diesem Unterrichtsmaterial:

Beispiel 1: Arbeitsblatt zu Übung 1 „Was ist okay?“

Stell dir vor, du findest solche Chats und Kommentare in WhatsApp, Instagram oder TikTok. Was sagst du, findest du es okay, so mit jemandem zu schreiben? Kreuze an!

Ist okay
 Ist nicht okay

Boah Maxi, du nervst!

Maxi soll sich mal endlich waschen, seine Haut ist schon ganz dunkel. Ekelhaft!

Ist okay
 Ist nicht okay

Maxi, du bist sch****!
Raus aus der Gruppe mit dir!

Ist okay
 Ist nicht okay

Maxi zeichnet wie ein Kindergartenkind, voll peinlich!

Ist okay
 Ist nicht okay

Maxi kann nicht Computerspielen, weil sie ist ein Mädchen. Und Mädchen können das nicht.

Ist okay
 Ist nicht okay

Abbildung 2

Beispiel 2: Arbeitsblatt zu Übung 2 „Ich kann was tun!“

Hier findest du Ideen, was man gegen Online- Beschimpfungen tun könnte. Natürlich müssen nicht alle Vorschläge für dich passen. Überleg dir auch selbst, was du tun könntest. Verbinde die geschilderten Situationen mit den Lösungsansätzen, die dir passend erscheinen.

<p>Das Opfer in Schutz nehmen. „Genug jetzt. Sie hat nix getan.“</p>	<p>Gegen Vorurteile schreiben. „Das stimmt doch nicht. Wer glaubt so einen Blödsinn?“</p>	<p>Warum die Täterinnen und Täter das tun „Der glaubt vielleicht, das ist cool.“</p>
<p>Das Opfer trösten. „Mach dir keine Sorgen, die sind ja nur neidisch.“</p>	<p>Unterstützung anbieten. „Ich bin für dich da.“</p>	<p>Die unterstützen, die dagegen sind. „Ich sehe das auch so. Wir sollten das melden.“</p>
<p>Ablenken. „Gemma Kino!“ „STOPSTOPSTOPST“</p>	<p>Stimmung auflockern. lustige Memes posten; Gruppennamen in „<u>peace</u>“ ändern.</p>	<p>Nachfragen. „Was ist eigentlich los?“</p>

Abbildung 3

Stell dir vor, du siehst folgende Dinge im Internet. Welche der oben genannten Ideen würdest du anwenden?

<p>Deine beste Freundin wird gehatet und die ganze Klasse schaut zu oder macht mit.</p>	<p>Dein Klassensprecher wird beschimpft, weil er nichts gegen die viele HÜ gesagt hat.</p>	<p>Alle hacken schon seit Tagen auf einer Person herum.</p>
<p>Ein Nacktbild macht die Runde. Du kennst die Person nicht.</p>	<p>Jemand aus deiner Klasse wurde aus der Klassengruppe ausgeschlossen.</p>	<p>Ein echt schönes Bild wird als hässlich bezeichnet.</p>
<p>Eine gute Computerspielerin wird beschimpft. Die sind wohl neidisch, weil sie besser ist.</p>	<p>Zwei Kinder in der Klasse beschimpfen sich wegen ihrer unterschiedlichen Religion.</p>	<p>Jemand aus deiner Klasse schreibt einer Person: „Geh sterben.“</p>

Abbildung 4

Beispiel 3: Arbeitsblatt zu Übung 5 „Mach doch mit!“

Finde Argumente, warum es gemeinsam viel leichter ist, gegen Online-Beschimpfungen vorzugehen. Wie kannst du andere Menschen dazu motivieren, sich dir anzuschließen? Was sagst oder schreibst du ihnen? Notiere deine Argumente, orientiere dich dafür an den unten genannten Ideen, überlege aber auch nach eigenen Antworten!

Zu reagieren, bringt eh nix.

Das geht mich nix an.

Aber das stimmt doch.

Die ist selbst schuld, wenn sie sowas macht.

Ich weiß doch nicht, was passiert ist.

Das nervt.

Wenn ich was mach, wird's doch bloß schlimmer.

Ein paar Ideen:

Wir alle machen Fehler.

Es geht uns alle was an.

Wir machen da nicht mit.

Wir können gemeinsam nachfragen.

Kritik ist okay, aber Beschimpfen nicht.

Gemeinsam können wir das stoppen.

Lasst uns doch gemeinsam Spaß haben, nicht Stunk.

Man muss nicht alle Menschen mögen, aber andere zu beschimpfen ist trotzdem nicht okay.

Abbildung 5

3) Adaptierung und Erweiterung bestehender Flyer für Jugendliche zu Online Zivilcourage:

Der bereits im Zuge des KIRAS Projekt „Zivilcourage 2.0“ entwickelte Flyer für Jugendliche wurde um konkrete Ideen, die von den Jugendlichen im Rahmen des Forschungsprojektes entwickelt und benannt wurden, ergänzt (<https://www.saferinternet.at/zielgruppen/jugendliche/?file=4530>). Dabei wurde insbesondere auch auf konkrete Counter Speech Ideen hingewiesen (siehe nachfolgende Abbildung).

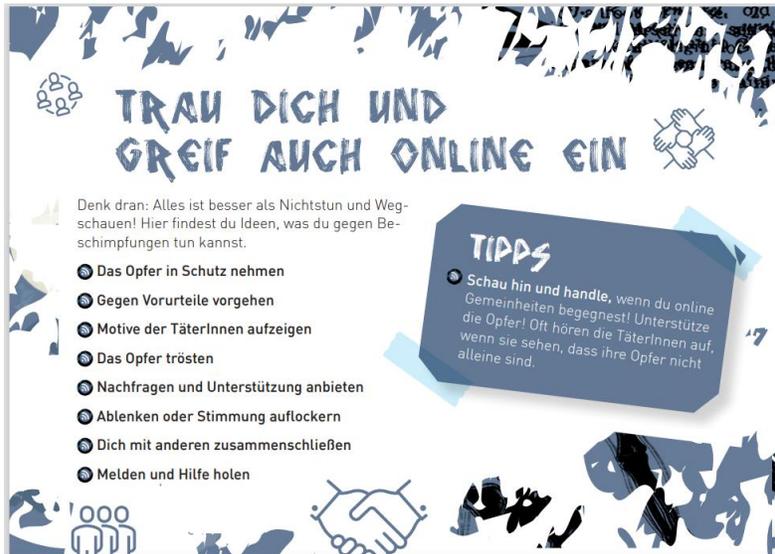


Abbildung 6

4) Integration des Themas in Saferinternet.at-Workshops:

Die Forschungsergebnisse werden auch im Rahmen der Weiterbildung der österreichweit tätigen insgesamt rund 35 Saferinternet.at-Trainer*innen weitergetragen und in bestehende Workshop-Formate integriert. Dabei werden die Ergebnisse der Studie auch für die Erweiterung und Ergänzung der Saferinternet.at Workshops- und Trainingsunterlagen verwendet. Ziel ist es, Jugendliche zu aktiver Counter Speech zu ermutigen, sie dafür zu sensibilisieren und ihnen Anregungen und Ideen für konstruktive Counter Speech mitzugeben – im Rahmen ihrer verfügbaren Basiskompetenzen.

Mauthausen Komitee Österreich / MKÖ

1) Entwicklung von Online Informations-, Übungs- und Trainingsmaterial für Jugendliche:

Auf Basis der empirischen Erkenntnisse und gemeinsam mit Expert*innen entwickelte das MKÖ im Rahmen des Projektes „ZIVIL.COURAGE.ONLINE“ Informations-, Übungs- und Trainingsmaterial für Jugendliche speziell auf der ZIVIL.COURAGE.ONLINE App und der Website. Für die beiden Plattformen wurden Materialien und interaktive Übungen zu folgenden Themenerarbeitet: Hass im Netz, Argumentation, Beispiele von Menschen die Online Zivilcourage gezeigt haben, Umgang mit Fake

News, Verschwörungstheorien, Gesetze & Rechtliches, Meldestellen; Literaturhinweise & Links, Cybermobbing, Media Literacy, Social Media; das Angebot wird durch Übungen für mehr Zivilcourage auf Online Plattformen zusätzlich bereichert. Dabei werden insbesondere auch Detailergebnisse aus den Projektergebnissen genutzt, um z.B. geplante Chat-Übung zu Counter Speech zu adaptieren. Konkret wird etwa seit 2022 eine **Chat-Übung im Rahmen der App bei den Trainings zum trainieren von Counter Speech** verwendet.

Das Projekt „ZIVIL.COURAGE.ONLINE“ hat zum Ziel, Zivilcourage auf Onlineplattformen unter Jugendlichen gezielt zu fördern. Gemeinsam mit ExpertInnen hat das MKÖ ZIVIL.COURAGE.ONLINE Trainings speziell für Jugendlichen konzipiert und für die Durchführung einen Pool an neuen Zivil.Courage.Online TrainerInnen ausgebildet. Die zentralen Ziele des ZIVIL.COURAGE.ONLINE-Trainings sind die Sensibilität für zivilcouragiertes Handeln im Internet zu erhöhen, das eigene Verhaltensrepertoire auf unterschiedlichen Ebenen zu erweitern sowie Zivilcourage und gesellschaftliche Verantwortung auf Online-Plattformen zu stärken. Außerdem wird das Erkennen von Fake News und Verschwörungstheorien gefördert und Media Literacy vermittelt. Die ZIVIL.COURAGE.ONLINE Trainings tragen zur Sensibilisierung, Förderung und Stärkung sozialer Kompetenzen bei und helfen bei der Erarbeitung individueller Handlungsoptionen in schwierigen Situationen auf diversen Onlinekanälen. Fokussiert wird auf Sensibilisierung, Förderung und Stärkung sozialer Kompetenzen, Erarbeitung individueller Handlungsoptionen in schwierigen Situationen: z.B. Umgang mit Angriffen, Ohnmacht, Aggression und Angst im Internet und diskriminierenden Situationen: speziell hier wird eine Übung zum Trainieren von Counter Speech angeboten.

Die verschiedensten Arten der Vermittlung zielen ab eine ständige Interaktion mit den Jugendlichen und auf langfristige Wissensvermittlung ab. Auf Stärken und Schwächen der Jugendlichen wird mit unterschiedlichen methodischen Ansätzen reagiert. Die Trainings basieren auf Prinzipien und Praktiken des non-formalen Lernens. Angeboten wird ein Methodenmix aus theoretischen Inputs, interaktiven Übungen, Diskussionen sowie Arbeiten in Kleingruppen. Das Training soll die Potentiale und Handlungsmöglichkeiten von Mädchen und Jungen möglichst breit – ohne geschlechtsspezifische Rollenvorgaben – fördern. In der Durchführung des Workshops werden diverse Formen selbstgesteuerten und partizipativen Lernens eingesetzt um die Beteiligungschancen aller teilnehmenden Jugendlichen zu sichern.

Im Sinne des Bildungsauftrags und der Sensibilisierung in Bezug auf aktuelle Tendenzen des Rechtspopulismus in Österreich richtet sich das Projekt besonders an die jugendliche Zielgruppe. Schwerpunktmäßig werden mit dem Vermittlungsangebot Schüler*innen aus Polytechnischen Schulen, Berufsschüler*innen, Lehrlinge, Jugendliche aus außerschulischen Einrichtungen, jugendliche ArbeitnehmerInnen sowie Auszubildende angesprochen. Die ZIVIL.COURAGE.ONLINE-Trainings

werden in ganz Österreich durchgeführt. Die Inhalte der Trainings werden auf die jeweiligen Gruppen maßgeschneidert und dauern im Onlineformat 2 Stunden und im Präsenzformat insgesamt 4 Stunden.

Die folgenden Abbildungen vermitteln kurze Einblicke dieses Projekt: Angeboten werden hier konkrete Übungen, als auch Geschichten zu Online-Held*innen.

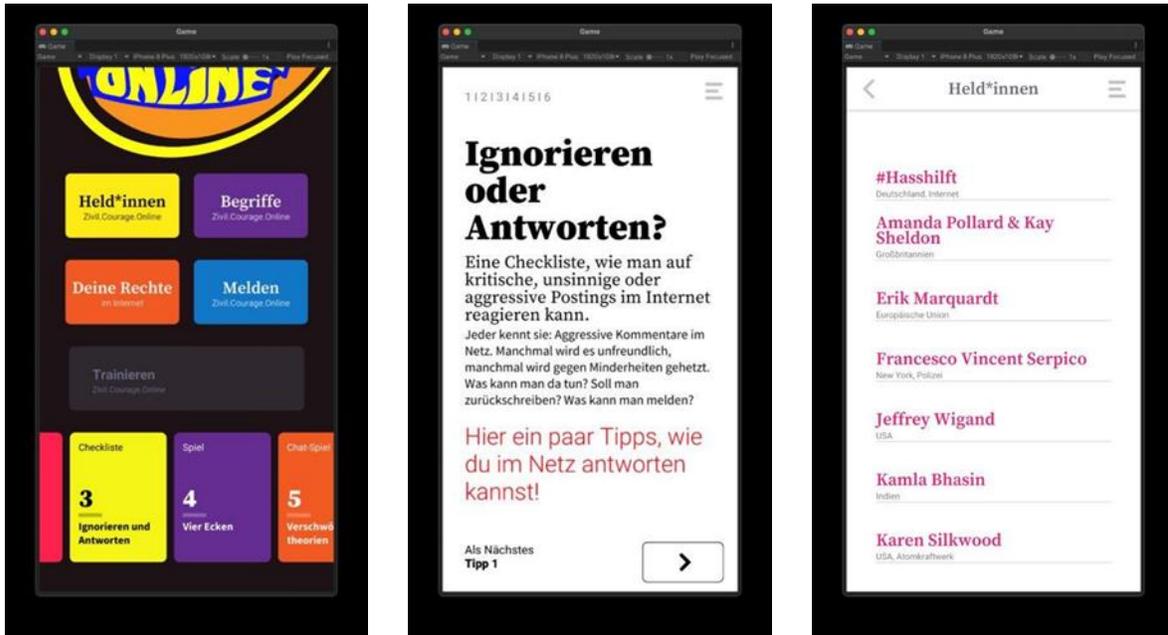


Abbildung 7

2) Integration der Ergebnisse in kostenlos verfügbare Online Materialien:

Das MKÖ bietet außerdem kostenlose Online Materialien in Form von Informations-, Übungs- und Trainingsmaterial für alle interessierten Personen an. Für Multiplikator*innen gibt es zusätzlich eine interaktive Materialienbox zur eigenständigen Vermittlung von Online-Zivilcourage: mit Hilfe der Ergebnisse wurden die Materialien auch hier ergänzt.

BM.I / Bundeskriminalamt / Büro 1.6 Kriminalprävention und Opferhilfe

1) Integration der Studienergebnisse im Rahmen von Überarbeitungen der Schulungsprogramme:

Die Studienergebnisse werden im Rahmen von Überarbeitungen der Schulungsprogramme für Jugendliche miteinbezogen. Dabei ist im Rahmen des Jugendpräventionsprogramms UNDER18 die Programmschiene „Clik&Check“ von hoher Relevanz. Dieses wird immer wieder adaptiert und weiterentwickelt, wodurch auch die gewonnen Forschungserkenntnisse in Form von neuen Arbeitsblättern und Übungen integriert werden können.

2) Einbindung der Ergebnisse in die Fortbildung von Präventionsbediensteten

Die tätigen Präventionsbediensteten erhalten regelmäßig Schulungen und Weiterbildungen. Auch hier können die gewonnen Erkenntnisse gut weitervermittelt werden.

Bundeskanzleramt: Referat VI/5/a – Kompetenzzentrum Jugend

1) Einbringen der Ergebnisse beim Nationalen Komitee No Hate Speech:

Die Ergebnisse wurden und werden im Rahmen des Nationalen Komitee No Hate Speech (www.nohatespeech.at/komitee/) präsentiert und diskutiert. In diesem Rahmen können zusammen mit vielen relevanten Stakeholdern (Bundesministerien, Landesdienststellen, außerschulische Kinder- und Jugendarbeit, einschlägige Beratungs- und Informationsstellen sowie anderweitige Initiativen) entsprechende Fragestellungen erörtert und entsprechende Vorschläge erarbeitet werden und eventuell auch die „Empfehlungen des Nationalen No Hate Speech Komitees an politische Entscheidungsträgerinnen und Entscheidungsträger“ ergänzt werden (www.nohatespeech.at/wp-content/uploads/2020/06/Empfehlungen_No_Hate_Speech_Komitee_06_2020.pdf).

2) Integration der Ergebnisse in bestehende Initiativen

Die Erkenntnisse regten zu Diskussion mit dem Österreichischen E-Sport-Verband (ESVÖ) an bzw. wurden etwaige Schlussfolgerungen für den Bereich des e-Gamings diskutiert. Außerdem können die Erkenntnisse in die Aufarbeitung im Rahmen der Bundesstelle für die Positivprädikatisierung von digitalen Spielen (z.B. über deren Website www.bupp.at) einfließen. Die Ergebnisse werden weiters in das Komitee für Kinder- und Jugendgesundheit eingebracht, dass die Umsetzung der Kinder- und Jugendgesundheitsstrategie begleitet. Diese Strategie wird derzeit weiterentwickelt und daher kann die Frage, inwieweit Hate Speech, Counter Speech und psychische Gesundheit im digitalen Bereich zu berücksichtigen wäre, entsprechend erörtert werden. Zusätzlich wird mit den Ressortkoordinationen der österreichischen Jugendstrategie diskutiert, inwieweit die Ergebnisse in die entsprechenden Jugendziele der Österreichischen Jugendstrategie einfließen können. Angestrebt wird weiters die Einbindung der Erkenntnisse in Schulungen, die vom BKA angeboten oder mitfinanziert werden.

Innovationsgehalt des Projekts aus Sicht der Praxispartner*innen

Zentrales Ziel des Forschungsprojektes war es, die gewonnen Erkenntnisse für die Praxis nutzbar zu machen. Dabei stand auch der Innovationsgehalt des gesamten Projektes im Vordergrund, um Ergebnisse zu generieren, die die Praxispartner*innen in ihrer täglichen Arbeit tatsächlich bereichern und neue Einsichten und Ansatzpunkte eröffnen. Die Praxispartner*innen wurden daher gebeten jene Aspekte zu schildern, die ihnen als besonders innovativ erschienen:

Die Praxispartner*innen hoben die deutlichen **Bildungs-Unterschiede in der Art und Weise, wie Jugendliche auf Hasskommentare reagieren** hervor, die sich sehr deutlich in den Ergebnissen abbilden: In den gemeinsamen Diskussionen und insbesondere in der gemeinsamen Durchführung der digitalen Aktionsforschung wurden diese Unterschiede konkret erlebbar: So wurde gesehen, dass Jugendliche mit höherem Bildungshintergrund eher sachlich und auch strategisch überlegend reagiert haben und bildungsfernere Jugendliche eher emotional, manchmal auch moralisierend und manchmal auch sehr ohnmächtig, ohne eigene Handlungsmöglichkeiten zu erkennen – allerdings sind das genau jene Jugendlichen, die viel Erfahrung mit Hass-Kommentaren haben. Die beteiligten Jugendlichen aus den AHS-Bereich hatte damit dagegen weniger Erfahrung im eigenen Leben.

Auch die im Projekt sichtbar gewordenen **genderspezifischen Unterschiede** sind für die Praxispartner*innen eine wichtige Erkenntnis: Ganz generell sehen und kennen Mädchen mehr an negativen Übergriffen im Internet als Burschen. Es lässt sich aber aus der Praxiserfahrung nicht genau ableiten, ob Mädchen tatsächlich mehr mit solchen Übergriffen konfrontiert sind, oder ob diese diesen gegenüber einfach aufmerksamer sind. Auffällig ist jedenfalls die Erkenntnis, dass Mädchen generell mehr Counter Speech Handlungsoptionen sehen und kennen als Burschen.

Die Ergebnisse machen auch klar, wie wichtig es ist, **Hasskommentare erkennen zu können**. In der praktischen Arbeit mit den Jugendlichen zeigte sich, dass es nicht immer einfach ist, Hass-Kommentare auch wirklich zu erkennen. Umso mehr braucht es Sensibilisierung und Diskussion mit jungen Menschen, wie sie solche Einordnungen treffen und was, wie auf sie wirkt. Auch Trusted-Flagger-Organisationen berichten, dass die Abgrenzung nicht immer einfach und eindeutig ist. Bei Jugendlichen kommt aber hinzu, dass aufgrund dieser Unsicherheit in der Zuordnung auch die Motivation niedrig bleibt, etwas dagegen zu unternehmen. Dies bestärkt die Praxispartner*innen in ihrer Arbeit, Jugendliche für das Erkennen von Hasskommentaren zu sensibilisieren und dazu zu ermutigen, zumindest das Melden von Kommentaren und Beiträgen vorzunehmen und sich nicht durch mangelnde Erfolge davon abhalten zu lassen.

Die Projektergebnisse verdeutlichen aber auch den **Mangel an Counter Speech Erfolgserlebnisse für die beteiligten Jugendlichen**. Oft gingen die eigenen Aktionen in der Masse der negativen Kommentare unter, manchmal waren sie gar nicht wahrnehmbar, was insgesamt demotivierend wirkte, selbst wenn die Umsetzung von Counter Speech als Teil einer Übung als motivierende Rahmenbedingung umgesetzt wurde. Es ist anzunehmen, dass ohne einer solchen unterstützenden Rahmendbedingung, Jugendliche noch weniger dazu motiviert sind, Counter Speech zu betreiben. Nach Ansicht der Praxispartner*innen braucht es die Begleitung und Unterstützung durch Erwachsene, die motivierend unterstützen und auch aufzeigen, dass es trotz eines Gefühls der Wirkungslosigkeit im Moment Sinn macht: Je mehr Internet-Nutzer*innen zeigen, dass sie mit negativen Inhalten nicht einverstanden sind, desto besser können Soziale Netzwerke dazu lernen und eine allgemeine Stimmungsänderung bewirken. Zudem darf auch die

Wirkung auf mitlesende Peers nicht außer Acht gelassen werden, woran letztlich auch der Erfolg von Counter Speech gemessen werden könnte.

Seitens des ÖIAT/Saferinternet.at, wurde insbesondere auch die Erkenntnis über die **Betroffenheit junger Kinder** betont. Die Bedeutung, die das Vorleben / die „Vorbildwirkung“ der Erwachsenen sowohl in sozialen Netzwerken wie auch im Umgang mit Hasskommentaren hat, war vor den Aktionsforschungsdiskussionen mit den Jugendlichen nicht in dieser Dimension bekannt. Denn auch erst die Beschäftigung mit den möglichen Personen, die gemeine Kommentare schreiben, zeigte auf, dass es sich – möglicherweise Pandemie-bedingt – um wirklich junge Nutzende handeln könnte. „Es muss also schon sehr früh anfangen, mit der Aufklärung“, so die Meinung der Jugendlichen. Da viele der Angebote zu diesem Thema erst mit 14+ Jahren beginnen, so ist dies aus Sicht des ÖIAT ein neuer Zugang und damit auch eine neue Notwendigkeit, die durch das Forschungsprojekt entstand.

Die Praxispartner*innen wiesen zudem darauf hin, dass die einzelnen **Detailergebnisse sehr wichtig** waren, um diese auch in der praktischen Arbeit mit den Jugendlichen im Rahmen der Trainings und auch in der Konzeption der interaktiven Übungen mitzudenken. Im Speziellen die unterschiedlichen Voraussetzungen von Motivation und Anreiz zu Counter Speech. Insbesondere das MKÖ wird hier laufend weitere **Übungen aus den Projektergebnissen** erarbeiten bzw. werden Materialien bzw. **geplante Chat-Übung zu Counter Speech** auf Basis der Ergebnisse adaptiert.

Von den Projektpartner*innen wurde auch hervorgehoben, dass im Zuge der Projektumsetzung bereits **vorhandene Counter Speech Unterstützungsangebote** (z.B. <https://www.schnellerkonter.at/>; <https://no-hate-speech.de/de/kontern/fuer-gegen-alle-hate-speech/>) immer wieder **integriert** wurden und somit das Forschungsprojekt selbst aktiv zu deren Dissemination beigetragen hat.

Schließlich stellte die aktive Mitarbeit der teilnehmenden Jugendlichen im Projekt für viele eine Ermutigung dar, auch außerhalb des Projekts aktiv zu werden. Dabei wurde bewusst, wie wichtig die **aktive Auseinandersetzung mit Counter Speech und das Üben des Konterns ist**, um die Wahrscheinlichkeit zu erhöhen, selbst aktiv zu werden. Dieser Übungsgedanke wird von den Praxispartner*innen v.a. in Trainings mit Jugendlichen vermehrt eingebracht.

Nachhaltige Nutzungsstrategien und Visionen für die Zukunft

Die entwickelten Unterstützungsangebote fließen in die tagtägliche Arbeit der beteiligten Organisationen ein und werden interessierten Personen, wie z.B. Pädagog*innen zur Verfügung gestellt.

Die **detaillierten Ergebnisse** der Studie werden aber auch zukünftig in die **Weiterentwicklung der Trainings sowie Übungen und Materialien** der beteiligten Praxispartner*innen integriert und dabei auch

Beispiele aus der Studie für die Nutzung in Trainings eingesetzt. Dabei sollen bei der Entwicklung entsprechender Methoden explizit auch Bildungsunterschiede und damit einhergehende vorhandene Kompetenzen stärker berücksichtigt werden bzw. Überlegungen zur Steigerung der Motivation für Counter Speech bzw. das Triggern von entsprechenden Anreizen.

Die entwickelten **Tools bieten zum Teil auch Potenzial, diese für weitere Studien** zu verwenden: So gibt es etwa in der vom MKÖ Chat-Übung im Rahmen der App zum Trainieren von Counter Speech die Möglichkeit, die Ergebnisse anonymisiert zu dokumentieren, was für weitere Studien nutzbar sein könnte.

Zukünftiges Ziel für die Praxis ist es, Ergebnisse der Studie direkt in die „**Digitale Grundbildung**“ von Schulen einfließen zu lassen. Dazu zählt auch der Ansatz, bereits bei deutlich jüngeren Kindern, wie etwa **Volksschüler*innen, mit Aufklärung und Bewusstseinsbildung** zu beginnen als das bislang der Fall ist. Gerade in Volksschule haben solche Themen bislang noch wenig Platz und Aufmerksamkeit gefunden und brauchen eine alterstypische Herangehensweise.

Ein weiteres **wichtiges Thema wird zukünftig auch das gemeinsame Auftreten gegen Online Hass** sein: Für Jugendliche ist es eine besondere Herausforderung, aktiv gegen Hasskommentare zu werden, da sie sich oft alleine damit alleine gelassen fühlen. Daher ist das Bilden von Gruppen, die gemeinsam gegen Hasskommentare vorgehen und die sich in einem Messenger oder Gruppe absprechen und gegenseitig unterstützen, eine wichtige Maßnahme, wie Jugendliche viel eher aktiv werden könnten. Auch das kann im Rahmen der digitalen Grundbildung unterrichtet werden und so den Jugendlichen als Option gezeigt werden.

Darüber hinaus muss zukünftig **von Sozialen Netzwerken** selbst verstärkt angedacht werden, **Counter Speech Aktionen gezielt zu unterstützen**, um das Machtungleichgewicht zwischen Hate- und Counter Speech klar entgegen zu treten. Es ist auch für Jugendliche unklar, welche Rolle hier Algorithmen spielen, um bestimmte Inhalte stärker anzuzeigen als andere.

Praxisrelevante Verwertung und Verbreitung der Studienergebnisse

Die bestehenden Netzwerke und Verwertungsstrukturen der Projektpartner*innen konnten auch für die Verbreitung der Studienergebnisse genutzt werden: Da die beteiligten Projektpartner*innen zentrale Positionen für Zivilcourage bzw. für den sicheren Umgang Jugendlicher mit digitalen Medien in einem österreichweiten und internationalen Netzwerk einnehmen, ist ein hohes und breites Nutzungs- und Einsatzpotenzial vorhanden.

Die **Dissemination der entwickelten Unterstützungsangebote der Praxispartner*innen** erfolgt über die vorhandenen, vielfachen Kanäle und Netzwerke:

- Dissemination über das Saferinternet.at-Netzwerk (www.saferinternet.at/broschuerenservice) und Saferinternet.at-Social Media-Kanäle und das Saferinternet.at-Veranstaltungsservice (www.saferinternet.at/veranstaltungsservice), über das MKÖ-Netzwerk und MKÖ Social Media Kanäle (<https://www.zivilcourage.at/zivilcourageonline>), sowie Verbreitung der Ergebnisse über entsprechenden Kanäle des BKA im Bereich Jugend und Familie (Newsletter, Websites)
- Nutzung der Saferinternet.at-Lehrendendatenbank und des MKÖ-Netzwerkes zur Bewerbung bei Lehrenden in Österreich
- Bewerbung über diverse Schulmedien und -initiativen rund um digitale Medien sowie im Rahmen der regelmäßig stattfindenden „Safer Internet“-Vernetzungstreffen für die offene und verbandliche Jugendarbeit (in Kooperation mit BOJA und BJV), sowie Verbreitung über den MKÖ-Newsletter
- Vorstellung beim Saferinternet.at-Beirat und dem MKÖ-Vorstand

Dissemination im Rahmen einer Abschlussveranstaltung

Die für Februar 2022 geplante Abschlussveranstaltung musste Corona-bedingt auf 24. Juni 2022 verschoben werden. Im Rahmen dieser Veranstaltung, die in der Sky Lounge am Oscar Morgenstern Platz stattfinden wird, werden sowohl Ergebnisse der Studie öffentlich vorgestellt, als auch die praktische Nutzung der Studienergebnisse durch die Praxispartner*innen. In Form von Informationsständen und workshopartigen Diskussionsrunden werden sich alle Praxispartner*innen präsentieren und Einblicke in die Arbeitsweisen der einzelnen Institutionen zu geben. Die Abschlussveranstaltung wird sowohl über die verfügbaren Kanäle der Universität Wien, als auch der Praxispartner*innen beworben und richtet sich an alle interessierten Personen aus dem Jugendbereich (z.B. Bildungsbereich, Soziale Arbeit, Jugendberatung, Jugendbetreuung, Präventionsarbeit mit Jugendlichen, psychologische Einrichtungen für Jugendliche, Jugendpolitik) bzw. Professionist*innen der Jugendarbeit.

Literaturverzeichnis

- Aslanidis, Paris (2017):* Measuring populist discourse with semantic text analysis: an application on grassroots populist mobilization. *Quality & Quantity*, 52(3), 1241-1263.
- Atzmüller, Christiane; Zartler, Ulrike; Kromer, Ingrid (2019a):* Zivilcourage 2.0 - Zivilcourage von Jugendlichen im Umgang mit wahrgenommener Gewalt im Internet. Forschungsendbericht. KIRAS-Sicherheitsforschung, FFG.
- Atzmüller, Christiane; Zartler, Ulrike; Kromer, Ingrid (2019b):* Online-Held*innen gibt es nicht? Was 14- bis 19-jährige Jugendliche an Zivilcourage im Internet hindert. *Sozialwissenschaftliche Rundschau*, 59. Jg., Heft 1, S. 87-109.
- Barlińska, J.; Szuster, A.; Winiewski, M. (2013):* Cyberbullying among adolescent bystanders: role of the communication medium, form of violence, and empathy. *Journal of Community and Applied Social Psychology*, 23, 37-51.
- Bastiaensens, Sara; Pabian, Sara; Vandebosch, Heidi; Poels, Karolien; van Cleemput, Katrien; DeSmet, Ann; Bourdeaudhuij, Ilse de (2016):* From normative influence to social pressure. How relevant others affect whether bystanders join in cyberbullying. In: *Social Development* 25 (1), S. 193–211.
- Bastiaensens, Sara; Vandebosch, Heidi; Poels, Karolien; van Cleemput, Katrien; DeSmet, Ann; Bourdeaudhuij, Ilse de (2014):* Cyberbullying on social network sites. An experimental study into bystanders' behavioural intentions to help the victim or reinforce the bully. In: *Computers in Human Behavior* 31, S. 259–271.
- Benesch, Susan; Ruths, Derek; Dillon, Kelly P.; Saleem, Haji Mohammad; Wright, Lucas (2016a):* Counterspeech on Twitter. A Field Study. A report for Public Safety Canada under the Kanishka Project. Online verfügbar unter <https://dangerousspeech.org/counterspeech-on-twitter-a-field-study/>,
- Benesch, Susan; Ruths, Derek; Dillon, Kelly P.; Saleem, Haji Mohammad; Wright, Lucas Wright (2016b):* Considerations for Successful Counterspeech. Online verfügbar unter <https://dangerousspeech.org/wp-content/uploads/2016/10/Considerations-for-Successful-Counterspeech.pdf>
- Blanchard, F. A.; Crandall, C. S.; Brigham, J. C.; Vaughn, L. A. (1994):* Condemning and condoning racism: A social context approach to interracial settings. *Journal of Applied Psychology*, 79, 993–997.
- Brody, N.; Vangelisti, A. L. (2015):* Bystander Intervention in Cyberbullying, *Communication Monographs*.
- Brown, R. (2016):* Defusing hate: A strategic communication guide to counteract dangerous speech. United States Holocaust Memorial Museum.

- Butler, Leah C.; Graham, Amanda; Fisher, Bonnie S.; Henson, Billy; Reynolds, Bradford W. (2021):* Examining the Effect of Perceived Responsibility on Online Bystander Intervention, Target Hardening, and Inaction. *Journal of Interpersonal Violence*, 0(0), 1-26.
- Butler, Leah C.; Fisher, Bonnie S. (2020):* Who Are the Mavens of Bystander Intervention? Implications for the Social Diffusion of Intervention Norms. *Journal of Interpersonal Violence*, 37(3-4), 2126-2146.
- Chandrasekharan, E.; Pavalanathan, U.; Srinivasan, A.; Glynn, A.; Eisenstein, J.; Gilbert, E. (2017):* You Can't Stay Here. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 1–22. <https://doi.org/10.1145/3134666>
- Cobbe, J. (2020):* Algorithmic Censorship by Social Platforms: Power and Resistance. *Philosophy & Technology*, 8(3), 270. <https://doi.org/10.1007/s13347-020-00>
- Czopp, A. M.; Monteith, M. J.; Mark, A. Y. (2006):* Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology*, 90, 784–803.
- Czopp, A. M.; Monteith, M. J. (2003):* Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin*, 29, 532–544.
- DeSmet, Ann; Bastiaensens, Sara; van Cleemput, Katrien; Poels, Karolien; Vandebosch, Heidi; Cardona, Greet; Bourdeaudhuijs*, Ilse De (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. In: *Computers in Human Behavior* 57, S. 398–415.
- Dickter, Cheryl L.; Newton, Virginia A. (2013):* To confront or not to confront. Non-targets' evaluations of and responses to racist comments. In: *J Appl Soc Psychol* 43 (2), E262-E275.
- Dickter, C. L.; Kittel, J. A.; Gyurovski, I. I. (2012):* Perceptions of non-target confronters in response to racist and heterosexist remarks. *European Journal of Social Psychology*, 42, 112–119.
- Dillon, Kelly P.; Bushman, Brad J. (2015):* Unresponsive or unnoticed? Cyberbystander intervention in an experimental cyberbullying context. *Computers in Human Behavior*, 45, 144–150.
- Ernst, Julian; Schmitt, Josephine B.; Rieger, Diana; Beier, Ann Kristin; Vorderer, Peter; Bente, Gary; Roth, Hans-Joachim (2017):* Hate Beneath the Counter Speech? A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos. In: *Journal for Deradicalization* (10)
- Fazio, R. H.; Hilden, L. E. (2001):* Emotional reactions to a seemingly prejudiced response: The role of automatically activated racial attitudes and motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, 27, 538–549.
- Franzosi, Roberto (2004):* From words to numbers: narrative, data and social science. Cambridge: Cambridge University Press

- Frenett, R.; Dow, M. (2015):* One to one online interventions: A pilot CVE methodology. Institute for Strategic Dialogue.
- Frischlich, Lena; Rieger, Diana; Morten, Anna; Gary, Bente (Hg.) (2017):* Videos gegen Extremismus? Counter-Narrative auf dem Prüfstand. In Kooperation mit der Forschungsstelle Terrorismus/Extremismus des Bundeskriminalamtes (Polizei + Forschung, 51).
- Gini, G. et al. (2008):* The role of bystanders in student's perception of bullying and sense of safety. *Journal of School Psychology*, 46, 617-638.
- Gradinger, P.; Strohmeier, D. (2013):* Cyberbullying – Mobbing mit neuen Medien: Prävalenzen, Komorbidität und Motive. *Psychologie in Österreich*, 33 (1), 7-12.
- Gröndahl, T.; Pajola, L.; Juuti, M.; Conti, M.; Asokan, N. (2018):* All You Need is "Love": Evading Hate-speech Detection. <https://arxiv.org/pdf/1808.09115.pdf>
- Halpern, D.; Gibbs, J. (2013):* Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior*, 29(3), 1159–1168.
- Henson, Billy; Fisher, Bonnie S.; Reyns, Bradford W. (2020):* There Is Virtually No Excuse: The Frequency and Predictors of College Students' Bystander Intervention Behavior Directed at Online Victimization. *Violence Against Women*, 26(5), 505-527.
- Herry, Emily; Gönültas, Seçil; Mulvey, Kelly L. (2021):* Digital era bullying: An examination of adolescent judgments about bystander intervention online. *Journal of Applied Developmental Psychology*, 76, 1-11.
- Lange, Andreas; Reiter, Herwig; Schutter, Sabine; Steiner, Christine (2018):* Handbuch Kindheits- und Jugendsoziologie. Wiesbaden: Springer VS.
- Lee, E.-J.; Jang, Y. J. (2010):* What Do Others' Reactions to News on Internet Portal Sites Tell Us? Effects of Presentation Format and Readers' Need for Cognition on Reality Perception. *Communication Research*, 37(6), 825–846.
- Levine, M.; Crowther, S. (2008):* The responsive bystander. How social group membership and group size can encourage as well as inhibit bystander intervention. *Journal of Personality and Social Psychology*, 95, 1429-1439.
- Machackova, H., Dedkova, L., Sevcikova, A., & Cerna, A. (2018). Bystanders' Supportive and Passive Responses to Cyberaggression. Journal of School Violence*, 17(1), 99–110.
- Mathew, B., Saha, P., Tharad, H., Rajgaria, S., Singhania, P., Maity, S. K., Goyal, P. & Mukherjee, A. (2019). Thou Shalt Not Hate: Countering Online Hate Speech, 369–380 (Proceedings of the Thirteenth International AAAI Conference on Web and Social Media*

- Medienpädagogischer Forschungsverbund Südwest (Hg.). (2020):* JIM Studie 2020: Jugend, Information, Medien. https://www.mpfs.de/fileadmin/files/Studien/JIM/2020/JIM-Studie-2020_Web_final.pdf
- Obermaier, Magdalena; Fawzi, Nayla; Koch, Thomas (2016):* Bystanding or standing by? How the number of bystanders affects the intention to intervene in cyberbullying. In: *new media & society* 18 (8), S. 1491–1507.
- Olenik-Shemesh, Dorit; Heiman, Tali; Eden, Sigal (2017):* Bystanders' Behavior in Cyberbullying Episodes. Active and Passive Patterns in the Context of Personal-Socio-Emotional Factors. In: *Journal of interpersonal violence* 32 (1), S. 23–48.
- Patterson, Lisa J.; Allan, Alfred; Cross, Donna (2016):* Adolescent perceptions of bystanders' responses to cyberbullying. In: *new media & society* 19 (3), S. 366–383.
- Rasinski, H. M.; Czopp, A. M. (2010):* The effect of target status on witnesses' reactions to confrontations of bias. *Basic and Applied Social Psychology*, 32,8– 16.
- Rösner, L.; Krämer, N. C. (2016):* Verbal Venting in the Social Web: Effects of Anonymity and Group Norms on Aggressive Language Use in Online Comments. *Social Media + Society*, 2(3).
- Schultz, J. R.; Maddox, K. B. (2013):* Shooting the messenger to spite the message? Exploring reactions to claims of racial bias. *Personality and Social Psychology Bulletin*, 39, 346–358.
- Shultz, Emily; Heilman, Rebecca; Hart, Kathleen J. (2014):* Cyberbullying: An exploration of bystander behavior and motivation. *Cyberpsychology*, Vol. 8 Issue 4.
- Slonje, R.; Smith, P. K. (2007):* Cyberbullying: Another main type of bullying? *Scandinavian Journal of Psychology*, 49, 147–154.
- Tarablus, Tamar; Heiman, Tali; Olenik-Shemesh, Dorit (2015):* Cyber bullying among teenagers in Israel. An examination of cyber bullying, traditional bullying, and socioemotional functioning. In: *Journal of Aggression, Maltreatment & Trauma* 24 (6), S. 707–720.
- Thomas, Laura; Falconer, Sarah; Cross, Donna; Monks, Helen; Brown, Deborah (2012):* Cyberbullying and the Bystander. Australian Human Rights Commission, Sydney.
- Thornberg, R.; Jungert, T. (2013):* School bullying and the mechanisms of moral disengagement. *Aggressive Behavior*, 40,99–108.
- Van Cleemput, Katrien; Vandebosch, Heidi; Pabian, Sara (2014):* Personal characteristics and contextual factors that determine "helping," "joining in," and "doing nothing" when witnessing cyberbullying. In: *Aggressive behavior* 40 (5), S. 383–396.
- Wächter, N.; Atzmüller, C; Hollauf, I. (2016):* Einstellungen und Verhalten von SchülerInnen auf sozialen Netzwerkseiten mit den Schwerpunkten Bilder, Selbstdarstellung, Sicherheit und Videospieldkultur.

Qualitative Forschungsergebnisse. Forschungsbericht der Universität Graz, gefördert von Sparkling Science (BMWF)

Wagner, U.; Brüggem, N.; Gerlicher, P.; Schemmerling, M. (2012): Wo der Spaß aufhört ... Jugendliche und ihre Perspektive auf Konflikte in Sozialen Netzwerkdiensten. Zweite Teilstudie im Projekt „Das Internet als Rezeptions- und Präsentationsplattform für Jugendliche“ im Auftrag der Bayerischen Landeszentrale für neue Medien (BLM). München: JFF – Institut für Medienpädagogik in Forschung und Praxis.

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. In: *Journal of homosexuality* 62 (10), S. 1289–1312.