universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Plattform-Governance als Akteur-Netzwerk
Die Rolle community-basierter Content-Moderation in antifeministischen Reddit-Foren

verfasst von | submitted by

Franziska Otremba BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 841
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Publizistik- und Kommunikationswissenschaft
Betreut von | Supervisor:	Univ.-Prof. Dr. Sophie Lecheler

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Verwendung von geschlechtergerechter Sprache

In dieser Arbeit wurde versucht, durch geschlechtsneutrale Formulierungen sowie dem sprachlichen Einbezug aller Geschlechter durch die Verwendung von Gendersternchen (*) eine geschlechtergerechte Sprache zu verwenden. Da das Forschungsinteresse dieser Arbeit sich auf exklusiv männliche Communities richtet, innerhalb derer weibliche und queere Nutzer*innen marginalisiert werden, erschien die Verwendung von inklusiver Sprache in der Beschreibung dieser Gruppen als wenig sinnvoll. Da geschlechtergerechte Sprache nicht dazu führen sollte, die geschlechtliche Spezifität sozialer Phänomene zu verdecken, wurde entschieden, in der Beschreibung der Akteure innerhalb maskulinistischer Gruppen entsprechend maskuline Formulierungen zu verwenden.

1. Männlich-zentrierter Antifeminismus online: Maskulinistische Online-Communities als ‚*digitale Gegenrevolution*'?

Feministische Bewegungen können als einer der einflussreichsten zivilgesellschaftlichen Akteure des 20. und 21. Jahrhunderts bezeichnet werden. Die Erfolge dieser Assoziationen bestehen hierbei nicht nur aus der wesentlichen Verbesserung der Lebensumstände von Mädchen und Frauen im jeweiligen nationalen Kontext, auch auf internationaler Ebene konnten durch ihre Bestrebungen diverse Gleichstellungspolitiken implementiert werden (Gergorić et al., 2025, S. 2-3). Seit Anbeginn des Erstarkens emanzipatorischer Forderungen und der Etablierung der Frauenbewegung bildeten sich unterschiedlichste Formen des Backlashs als Reaktion auf progressive Mobilisierung bezüglich des Strebens nach einer Gleichstellung der Geschlechter. Diese können sowohl auf kultivierte Frauenfeindlichkeit als auch auf gesellschaftliche Ängste vor Verlust von Ordnung und Struktur zurückgeführt werden (Schmincke, 2018, S. 28). Gegenwärtig wird dieser Backlash zudem erheblich durch Akteure der neuen Rechten intensiviert, die in vielen liberalen Demokratien einen enormen Zuwachs erleben (Sanders & Jenkins, 2022, S. 369).

Seit den 2000er-Jahren lässt sich die Verbreitung unterschiedlicher Formen von Anti-feminismen auch in der digitalen Öffentlichkeit vermehrt feststellen (Aigner & Lenz, 2023, S. 722; Holm, 2023, S. 423). Dabei zeichnen sich Debatten um die Themen Feminismus und Geschlechtergerechtigkeit in der Netzöffentlichkeit durch besonders intensive Formen der Inzivilität aus, welche eine konstruktive Auseinandersetzung mit Geschlechterungleichheit stark einschränkt (Ganz & Meßmer, 2015, S. 60). Obwohl soziale Interaktionen im digitalen Raum mit der Etablierung des Web 2.0 wesentlich durch populäre soziale Plattformen ermöglicht und strukturiert werden, die, um ihre Dienstleistung erbringen zu können, auf diverse Formen der inhaltlichen Moderation setzen müssen, da ihre Plattformen sonst auf Grund unüberschaubarer Mengen von Spam und problematischen Inhalten nicht nutzbar wären (Gillespie, 2018, S. 5), sind toxische Verhaltensweisen hier weit verbreitet. Insbesondere die Alltäglichkeit von frauenfeindlichen und antifeministischen Inhalten auf ebendiesen Plattformen hat zur Folge, dass Frauen sich aus politischen Debatten zurückziehen und exklusiv männliche und antifeministische Räume sich stark ausbreiten können (Reich & Bachl, 2025, S. 12-13; Nagle, 2018, S. 105).

In Hinblick auf die Gegenwärtigkeit von Frauenfeindlichkeit im Netz konnte durch vergangene Forschung bereits aufgezeigt werden, dass digitale Räume von antifeministischen Akteuren genutzt werden, um frauenfeindliche Inhalte zu verbreiten und um antifeministische Agitation zu betreiben. Zudem können antifeministische Gegenöffentlichkeiten gebildet werden, indem

Online-Foren dazu dienen, Werte und Ziele antifeministischer Gruppen auszuhandeln und Inhalte als Wissensressourcen bereitzustellen (Ganz & Meßmer, 2015, S. 66).

In der wissenschaftlichen Auseinandersetzung mit digitalem Antifeminismus stand zuletzt primär die inhaltsanalytische Beschäftigung mit antifeministischen Postings im Fokus, weniger aber, welche Rolle die technischen Rahmenbedingungen der Plattformen spielen, auf denen derartige Inhalte gepostet und verbreitet werden können (Holm, 2023, S. 423). Dabei stehen populäre Social-Media-Plattformen in der Öffentlichkeit vermehrt unter Druck, ihre Systeme der Content-Moderation zu verbessern, da Frauen, queere und rassifizierte Personen überproportional von digitaler Inzivilität betroffen sind (Gillespie, 2018, S. 24; Ging & Siapera, 2018, S. 515). So zeigte insbesondere die Reddit-Hasskampagne *#gamergate*, in deren Rahmen Frauen aus der Gaming-Community gezielter Belästigung durch frauenfeindliche Nutzer ausgesetzt waren, wie Plattformen durch ihre Architektur als fördernder Faktor für ebensolche misogynen Attacken wirken können (Massanari, 2017, S. 334, 336-341). Durch das Beispiel *#gamergate* wurde zudem sichtbar, dass digitaler Antifeminismus nicht nur durch frauenfeindliche Binnenkommunikation in eigens dafür angelegten Räumen besteht, sondern darüber hinaus gezielte Angriffe auf öffentlich sichtbare Frauen organisiert werden, welche nicht nur im Cyberspace stattfinden, sondern oftmals weit in das analoge Leben der Betroffenen hineinreichen können (Marwick & Caplan, 2018, S. 543-544).

Um derartige Angriffe in Zukunft zu verhindern, können digitale Plattformen auf verschiedene Formen der Content-Moderation zurückgreifen. Die Mehrheit populärer Plattformen baut hierbei auf einen Top-down-Ansatz in der Filterung und Löschung von digitalen Inhalten. Dieser zeichnet sich dadurch aus, dass die Grenzziehung zwischen erlaubten und verbotenen Inhalten von den Plattformbetreibenden selbst getroffen wird (meist kommuniziert durch sog. *community guidelines*) und auch die Entscheidung zur Löschung von Inhalten seitens der Plattform vorgenommen wird (Gillespie, 2018, S. 45; Seering, 2020, S. 2-3). Nutzer*innen können in diesem Modell nur durch das Melden von problematischen Inhalten (sog. *flagging*) versuchen, Einfluss auf das System der Top-down-Moderation zu nehmen (Are, 2025, S. 3580).

Aus feministischer Perspektive wurde in der Vergangenheit vor allem die mangelnde Transparenz, die innerhalb von Top-down-Moderationssystemen herrscht, kritisiert. Hierbei fehlt es zumeist an Einsicht in Handbücher und Regelwerke der professionellen Content-Moderation, nachvollziehbaren und zuverlässigen Beschwerdewegen sowie Möglichkeiten, Einblicke in die Prozesse der automatisierten Filterung zu gewinnen (Gerrard, 2020, S. 748-749). Im Fokus öffentlicher, feministischer und wissenschaftlicher Debatten über Content-

Moderation stehen meist die Schwachstellen ebendieser Top-down-Modelle, da diese zentralisierte Herangehensweise in der Content-Moderation durch die meisten populären Plattformen angewandt wird (Seering, 2020, S. 3).

Eine Alternative zu Top-down-Ansätzen bieten Modelle der community-basierten Moderation. Hierbei erhalten Nutzer*innen wesentlich größere Spielräume, um die digitalen Räume, in denen sie sich aufhalten, selbstbestimmter filtern zu können. Allerdings ist das dezentrale Modell der Community-Moderation bislang wesentlich weniger erforscht als klassische Ansätze der Content-Moderation und weist zudem eine höhere Anfälligkeit für Missbrauch durch radikalisierte Gruppen auf, da community-basierte Systeme auf interne Regulierung bauen (Seering, 2020, S. 2-4, 8).

Die Entwicklung digitaler Plattformen als zentrale Orte für soziale Interaktionen im Netz ist jeher begleitet von breiten gesellschaftspolitischen und rechtlichen Debatten um die Möglichkeiten sinnvoller Herangehensweisen an Content-Moderation, die deviantes Verhalten frühzeitig erkennen, problematische Inhalte zuverlässig filtern und gleichzeitig die Freiheiten zu Meinungsäußerung und -bildung, Möglichkeiten zu Partizipation sowie Kreativität nicht übermäßig einschränken (Gillespie, 2018, S. 1-5). In diesen Aushandlungsprozessen bleiben community-basierte Modelle meist unberücksichtigt, da sie derzeit nur von wenigen Plattformen angeboten werden und bislang nur begrenzte Erkenntnisse über ihre Funktionalität vorliegen. Dies bedeutet gleichzeitig, dass der Diskurs die Verantwortung der Plattformen stark zentriert. Forderungen nach größeren Handlungsspielräumen und mehr Mitbestimmung für Nutzer*innen, um digitale Räume, die ihr soziales Leben prägen, selbstbestimmt mitgestalten zu können, stehen hierbei weniger im Vordergrund. Um derartige Perspektiven zu fördern, braucht es die Grundlage wissenschaftlicher Erkenntnisse, die mögliche Vor- und Nachteile der Beteiligung von User*innen an Moderationsprozessen aufzeigen.

Auf Grund der geringen Dichte wissenschaftlicher Erkenntnis zu Modellen der community-basierten Moderation befasst sich das vorliegende Forschungsprojekt daher mit eben solchen Prozessen der Content-Filterung. Da die Gefahr des potenziellen Missbrauchs der den User*innen bereitgestellten Werkzeugen zur Filterung und Löschung eine der wesentlichen Schwächen des Modells darstellt und vergangene Forschung die zentrale Rolle von Plattforminfrastrukturen für die Verbreitung und Reproduktion von Frauenfeindlichkeit herausarbeiten konnte, werden in dieser Arbeit Moderationsprozesse innerhalb von antifeministischen Foren der Plattform Reddit (sog. *Subreddits*) untersucht. Im Rahmen des Forschungsprojekts wird durch die Analyse von Handlungsweisen innerhalb des

Moderationssystems und Formen der Aneignungen der durch Reddit bereitgestellten Handlungsmöglichkeiten zur Moderation innerhalb der Communities untersucht werden, welche Rolle das System der community-basierten Moderation auf Reddit für den Erhalt der antifeministischen Gruppen einnimmt. Hierbei soll herausgefunden werden, ob und wie die Struktur der Plattform in Hinblick auf das Filtern von Inhalten genutzt werden kann, um das Fortbestehen der antifeministischen Communities zu sichern.

2. Digitale Plattformen aus politischer und feministischer Perspektive

Seit der zunehmenden Popularisierung des Internets in den 1990er- und 2000er-Jahren gibt es einen breiten Diskurs rund um die Zunahme von Möglichkeiten politischer Partizipation. In der feministischen Auseinandersetzung mit den emanzipatorischen Potenzialen des digitalen Raums entwickelten sich Techno-Utopien von sozialen Interaktionen, die losgelöst von den Zwängen einer Festschreibung auf einen vergeschlechtlichten Körper stattfinden (Luckman, 1999, S. 41). Insbesondere durch die Perspektiven des Cyberfeminismus wurden die Potenziale der Internetnutzung in Hinblick auf das Aufbrechen eindeutiger Identitäten und Binaritäten beschrieben (Haraway, 1995, S. 34-35; Van Zoonen, 2001, S. 68). Durch die weite Verbreitung sexistischer und rassistischer Inhalte im Internet mussten diese Perspektiven allerdings schon bald um eine tiefgreifende feministische Medienkritik ergänzt werden (Van Zoonen, 2001, S. 68).

Die Chancen der Körperlosigkeit und uneindeutiger Identität konnten im Cyberspace bislang nur eingeschränkt realisiert werden, da seit Anbeginn des Internets spezifische Aspekte analoger Verkörperung in den digitalen Raum übernommen bzw. in das Digitale übersetzt wurden. Bereits vor der Entwicklung des Web 2.0 zur Zeit des textbasierten Internets wurden Chatroom-Avatare genutzt, um rassistische und sexistische Stereotypen zu reproduzieren (Nakamura, 2002, S. 41). Zudem zeigten sich die Schwierigkeiten der User*innen mit der potenziellen Körperlosigkeit im Netz umzugehen und auf eindeutige Bekanntgaben der Geschlechtlichkeit anderer zu verzichten. So sahen sich Nutzer*innen, die keine eindeutige Geschlechtsangabe zur Beschreibung ihres digitalen Avatars verwendeten, vermehrt damit konfrontiert, dass andere Mitglieder der textbasierten Foren gezielt versuchten ihr analoges Geschlecht herauszufinden (Turkle, 1998, S. 343).

Mit der Entstehung vieler einflussreicher sozialer Medien im Zuge des Web 2.0 in den 2000er-Jahren haben sich neue Möglichkeiten der digitalen Konnektivität herausgebildet.

Privatwirtschaftlich organisierte Plattformen bündeln und strukturieren seither einen Großteil sozialer Interaktionen im Internet (Gillespie, 2018, S. 14; Van Dijck, 2013, S. 4). Plattformen können hierbei verstanden werden als digitale Infrastrukturen, die usergenerierte Inhalte speichern, sortieren, verbreiten und moderieren, während sie selbst – anders als traditionelle Medien – keinen primären Fokus auf die professionelle Generierung von Medieninhalten legen (Habermas, 2022, S. 44; Gillespie, 2018, S. 18-21). Da soziale Interaktionen zahlreich auf Social-Media-Plattformen stattfinden, kann davon ausgegangen werden, dass auch Aushandlungsprozesse rund um die gesellschaftliche Bedeutung der Geschlechterverhältnisse zunehmend im digitalen Raum vorgefunden werden können.

Für die Analyse vergeschlechtlichter Interaktionen im digitalen Raum ist es hierbei zentral, dass Plattformen nicht als neutrale Werkzeuge verstanden werden können. Jede Form der Öffentlichkeit prägt die in ihr verhandelten Aspekte durch ihre Struktur mit und kann durch diese Struktur einige Positionierungen besser abbilden als andere (Fraser, 1990, S. 69). So wurden beispielsweise die Möglichkeiten zu einem geschlechtsneutralen Auftreten im digitalen Raum spätestens durch die Verbreitung kommerzialisierter Plattformen stark eingeschränkt. Plattformen verfolgen ein ökonomisches Interesse an personenbezogenen und damit auch geschlechtsbezogen Konsument*innendaten, um personalisierte Vorschläge von Werbeanzeigen und Postings zu optimieren. Dementsprechend verlangt ihre algorithmische Struktur meist eine eindeutige Zuordnung in eine vorgegebene geschlechtliche Kategorie (Bivens, 2017, S. 885; Bucher, 2018, S. 5). So war es im Jahr 2004 noch möglich, einen Facebook-Account ohne jegliche Angabe eines Geschlechts zu erstellen. Auch wenn es seit dem Jahr 2014 eine zunehmende Zahl an Auswahlmöglichkeiten für die Angabe des eigenen Geschlechts auf der Plattform gibt, ist die Angabe seither verpflichtend und wird auch bei der Auswahl einer nicht-binären Geschlechtsangabe im Hintergrund in eine binäre Kategorie umcodiert (Bivens, 2017, S. 885).

Darüber hinaus werden die Partizipationsmöglichkeiten von Userinnen in Online-Diskursen durch die starke Präsenz von Frauenfeindlichkeit im Internet bis in die Gegenwart stark eingeschränkt (Shaw, 2014, S. 275). So verstärken sexistische Kommentare in politischen Online-Debatten die Angst von Frauen, durch eine Teilnahme an Diskussionen Belästigung zu erfahren, wodurch ihre Sicherheit, sich an der Diskussion beteiligen zu können, sinkt (Reich & Bachl, 2025, S. 12-13). Darüber hinaus können Erfahrungen mit Sexismus online auch zu einem Rückzug von Frauen aus Offline-Debatten führen, da sie hier mit ähnlichen negativen Reaktionen auf ihre Beteiligung rechnen (Ortiz, 2024, S. 124-125).

Diese Befunde weisen darauf hin, dass das Internet insbesondere durch den Einzug digitaler Plattformen – entgegen den anfänglichen feministischen Techno-Utopien – ein vergeschlechtlichter Raum ist, in dem gesellschaftspolitische Aushandlungsprozesse über Geschlechterverhältnisse stattfinden. Hierbei spielt nicht nur die stark ausgeprägte Präsenz von Sexismus und Misogynie eine Rolle, sondern auch digitaler Antifeminismus, unter dem im konkreteren Sinne Formen der Frauenfeindlichkeit zu verstehen sind, die als direkte Reaktion auf feministische Forderungen oder Errungenschaften gewertet werden können (Schmincke, 2018, S. 28-29).

3. Entwicklungen des Antifeminismus on- und offline

Antifeminismus beschreibt die Mobilisierung gegen Feminismus und Geschlechtergerechtigkeit und besteht als Reaktion auf feministische Forderungen und gleichstellungspolitische Maßnahmen seit Beginn des 20. Jahrhunderts (Aigner & Lenz, 2023, S. 722). Seither hat sich die antifeministische Gegenbewegung weiter ausdifferenziert und kann heute schwerlich unter einem Begriff zusammengefasst werden. Unter anderem bildete sich so der *männlich-zentrierte Antifeminismus* (oder auch *Maskulinismus*) heraus, dessen Ursprünge in der vormals progressiven Männerrechtsbewegung der 1970er-Jahre liegen (Kaiser, 2021, S. 37). Diese spezifische Form des Antifeminismus zeichnet sich vor allem durch die Annahme eines hegemonialen Feminismus aus, der vermeintlich dazu führe, dass Männer in der Gegenwart durch Gleichstellungspolitiken gesellschaftlich benachteiligt würden, wobei ‚der Feminismus' durch die Akteure der Gegenmobilisierung zum zentralen Antagonisten erklärt wird (Nagle, 2018, S. 107). So verweisen Maskulinsten auf die Benachteiligung von Jungen innerhalb des Schulsystems, die höheren Suizidraten unter Männern im Vergleich zu Frauen sowie fehlende öffentliche Infrastruktur zur Unterstützung von männlichen Gewaltopfern und machen gleichstellungspolitische Entwicklungen hierfür verantwortlich (Blais & Dupuis-Déri, 2012, S. 23-24). Das Kernanliegen des männlich-zentrierten Antifeminismus ist es, den Status hegemonialer Männlichkeit zu erhalten und feministische Errungenschaften aufzuheben (Aigner & Lenz, 2023, S. 722). Dieses Anliegen bleibt bis heute erhalten, allerdings hat sich das Phänomen des männlich-zentrierten Antifeminismus insbesondere durch politische und ökonomische Veränderungen sowie durch seine massive Verlagerung in den digitalen Raum verändert.

3.1 Maskulinismus als dominante Form des gegenwärtigen Antifeminismus

Antifeminismus entwickelt sich nicht in einem Vakuum, sondern interagiert immer auch mit den bestehenden ökonomischen und politischen Verhältnissen (Blais & Dupuis-Déri, 2012, S. 25). In der wissenschaftlichen Auseinandersetzung mit männlich-zentriertem Antifeminismus bzw. Maskulinismus gilt es daher zu berücksichtigen, dass die Entwicklung der Geschlechterverhältnisse in westlichen Gesellschaften gegenwärtig von einer Vielzahl von Wandlungs- und Prekarisierungsprozessen geprägt ist. Gesellschaftliche Aushandlungsprozesse über eine geschlechtsspezifische Rollen- und Arbeitsverteilung und damit einhergehender Ungleichheiten sind wesentlich beeinflusst durch das Schwinden des Modells des männlichen Alleinernährers, die Förderung von weiblicher Erwerbstätigkeit und Diversität auf dem Arbeitsmarkt bei gleichzeitiger Prekarisierung der Erwerbssphäre und einer zunehmenden sozialen Abwärtsmobilität (Wimbauer et al., 2015, S. 43-45, 50-51). Diese Entwicklungen bedeuten für viele Männer einen zunehmenden Verlust von Privilegien bei bislang ausbleibenden alternativen Rollenangeboten und können zu starken Verunsicherungen führen (Sauer, 2021, S. 67).

Die Desorientierung, welche mit dem Wandel hegemonialer Männlichkeit einhergeht, kann insbesondere durch rechtspopulistische Akteure besetzt und zu einer „*Krise der Männlichkeit*“ umgedeutet werden (Sauer, 2021, S. 67). Hieraus entsteht eine doppelte Funktion des Maskulinismus, die diesen wesentlich charakterisiert. In erster Linie werden Frauen und insbesondere Feministinnen als Schuldige für politische und ökonomische Missstände benannt. Diese Schuldzuweisung ermöglicht wiederum die Rechtfertigung der Aufrechterhaltung bzw. Wiederherstellung männlicher Privilegien (Blais & Dupuis-Déri, 2012, S. 25). Der maskulinistische Diskurs ist demnach der Versuch einer (Re-)souveränisierung hegemonialer Männlichkeit, durch den verschiedene Dimensionen gegenwärtiger männlicher Prekarisierungserfahrungen im Sinne einer antifeministischen Gegenmobilisierung zu einem Krisennarrativ verknüpft werden. Auf diese Weise wird auf eine vermeintliche Umkehr der Machtverhältnisse zwischen Männern und Frauen verwiesen, aus der die Schlussfolgerung abgeleitet wird, dass ‚der Feminismus‘ und dessen Errungenschaften in Form von Gleichstellungsmaßnahmen für eine Benachteiligung von Männern im Allgemeinen verantwortlich zu machen seien. Das Berufen auf ein solches Krisennarrativ und das Hervorbringen einer „*männlichen Opfer-Identität*“ (Träbert, 2017, S. 273) sowie die Benennung ‚des Feminismus‘ als zentralen Antagonisten stellen eine (Re-)sourveränisierungsstrategie hegemonialer Männlichkeit dar (Forster, 2006, S. 200).

Für rechtskonservative und rechtspopulistische Akteure funktionieren antifeministische und antigender Diskurse als eine Art symbolischer Kleber und dienen strategisch dazu, verschiedenste Kräfte aus dem entsprechenden politischen Spektrum zusammenzubringen. Progressive Kräfte sind hierbei darauf angewiesen, neue Strategien zu entwickeln, um einen Umgang mit der starken Zunahme von antifeministischen Diskursen zu finden (Petö, 2015, S. 127-139). Dabei muss zudem berücksichtig werden, dass Forderungen des liberalen Feminismus und die neoliberale Wirtschaftsweise, die in vielen liberalen Demokratien dominiert, sich teilweise stützen, da der Bedarf nach mehr günstigen Arbeitskräften nun mit pseudofeministischer Rhetorik in Bezug auf die Notwendigkeit der Erwerbstätigkeit von Frauen ausgekleidet werden kann (Fraser, 2013). Diese logische Vereinbarkeit erschwert die Abgrenzung progressiver Standpunkte zusätzlich und kann von antifeministischen Akteuren genutzt werden, indem beispielsweise das Bild der klassischen Hausfrau als idyllisches Privileg neu besetzt wird (Leidig, 2023, S. 68).

3.2 Die Rolle von Online-Communities in der Verbreitung von Maskulinismus im digitalen Raum

Männlich-zentrierter Antifeminismus verbreitet sich gegenwärtig insbesondere über Online-Communities innerhalb von themenspezifischen Internetforen und Webseiten (Kaiser, 2021, S. 34-35). Online-Communities können definiert werden als Gruppen von Menschen, die regelmäßig über Computersysteme und digitale Infrastrukturen – allen voran über das Internet – in Interaktion miteinander treten. Die Gruppen verfolgen einen festgelegten Zweck und verfügen über Rahmenbedingungen in Form von explizit kommunizierten Regeln und Gesetzen oder implizit angenommenen Normen, die innerhalb der Gruppe gelten und eingefordert werden (Preece, 2000, S. 10). Die Anliegen von digitalen Communities sind hierbei vielfältig und reichen von der Bereitstellung von themenspezifischen Informationen bis hin zu sozial eng verknüpften Selbsthilfegruppen (Ren et al., 2012, S. 842). Auch maskulinistische Online-Communities können verschiedenste Absichten verfolgen. Allerdings stellen die Bereitstellung von Informationen entsprechend dem Themenfokus der Community sowie Aushandlungsprozesse und Debatten rund um Deutungsweisen der Problem- und Konfliktstellungen hinsichtlich der Geschlechterverhältnisse und das Teilen von Überzeugungen und Wertvorstellungen der Gruppe zentrale Funktionen dar (Ganz & Meßmer, 2015, S. 66).

Aus feministischer Perspektive stellt sich im Hinblick auf die Ausbreitung antifeministischer Online-Communities die Frage, in welchem Verhältnis diese zu hegemonialen Diskursen um Geschlechterverhältnisse und Gleichstellungspolitiken stehen. Grundsätzlich weist der digitale Raum Potenziale für eine weitere Demokratisierung der Öffentlichkeit und einer Zunahme politischer Partizipationsmöglichkeiten auf. Gleichzeitig lässt sich anhand der Fragmentierungsthese aufzeigen, dass das Internet ebenso dazu einlädt, die öffentliche Sphäre in immer kleinere Nischen zu unterteilen, auf Grund derer sich Individuen unterschiedlicher Ansichten nicht mehr begegnen müssen (Habermas, 2022, S. 45-47; Papacharissi, 2002, S. 15). So konnte bereits zur Zeit des textbasierten Internets in der Untersuchung von politischen Online-Communities aufgezeigt werden, dass diese oftmals aus einer eher homogenen Zusammensetzung von Mitgliedern bestehen und die dort stattfindenden politischen Debatten wenig Reziprozität aufweisen (Wilhelm, 1998, S. 329-331). Auch für Reddit-Communities konnte bereits ein Hang zu Homogenisierung nachgewiesen werden (Cinelli et al., 2021, S. 5). Dieser Umstand kann die produktive Auseinandersetzung mit gesellschaftlichen Konfliktstellungen in digitalen Öffentlichkeiten einschränken, da die homogene Zusammensetzung von Communities dazu führen könnte, dass ein deliberativer Austausch unterbunden wird und die innerhalb der Community dominierende Deutungsweise größtenteils unwidersprochen in das Zentrum der internen Debatte gestellt wird.

Für die maskulinistischen Communities würde die Annahme der zunehmenden Fragmentierung bedeuten, dass diese sich mit einer gruppenspezifischen, antifeministisch geprägten Perspektive auf Geschlechterverhältnisse von anderen Diskursen abkapseln und im Sinne des Communityerhalts versuchen, Interaktionen außerhalb dieser Gruppen zu meiden. In vergangener Forschung zu digitalen Communities hat sich diesbezüglich gezeigt, dass eine Homogenität unter den Mitgliedern und die Vermeidung von Konflikten zur Langlebigkeit einer solchen Gruppe beitragen können (Preece, 2000, S. 81). Für die Moderatoren könnte sich hieraus – entgegen der normativen Ansprüche an Inhaltsfilterung, deren zentrales Ziel es ist, toxisches Verhalten im digitalen Raum zu verringern – ergeben, dass es für den Erhalt der Community sinnvoller ist, Inhalte zu entfernen, die das Fortbestehen der Gruppe gefährden. Eine solche Moderationspraxis würde die maskulinistischen Gruppen stützen und gleichzeitig die Chancen auf konstruktive Geschlechterdebatten im Netz verringern.

Online-Communities aller Arten sind wesentlich auf ihre Mitglieder und deren aktiver Beteiligung an der Interaktion innerhalb der Gruppe angewiesen, da die Community ohne Inhalte und sozialen Austausch nicht bestehen kann. Die Communities müssen daher eine kritische Menge an aktiven Mitgliedern erreichen und langfristig binden, die groß genug ist,

um die Gruppe durch interessante Interaktionen zu erhalten. Gleichzeitig darf die Mitgliederanzahl nicht zu hoch sein, damit die Interaktionen nicht unübersichtlich und unpersönlich werden (Preece, 2000, S. 171). Diese Voraussetzung stellt sich in der Forschung immer wieder als wesentliche Hürde für den Erhalt von Online-Communities heraus, da Gruppen im digitalen Raum meist ohne große negative Konsequenzen verlassen werden können (Ren et al., S. 842). Den Akteuren einer solchen Community muss es also gelingen, Mitglieder langfristig durch den Aufbau von Verbundenheit zu halten. Dies gelingt durch die Förderung der Ausprägung einer Gruppenidentität innerhalb der Community (Ren et al., 2012, S. 842). Grundlegend für eine solche Identifizierung der Mitglieder mit der Gruppe ist die Wahrnehmung, dass die Individuen, die das Kollektiv bilden, einer gleichen sozial bedeutsamen Kategorie (beispielsweise Wohnort, Alter, Geschlecht) angehören (McPherson et al., 2001, S. 420-429). Gleichzeitig finden Prozesse der Depersonalisierung statt, indem individuelle Merkmale einzelner Partizipierender in den Hintergrund gerückt und hauptsächlich über ihre Partizipation an der Gruppe adressiert werden. Förderlich für die Herausbildung einer Gruppenidentität ist es zudem, die Homogenität der Gruppe kommunikativ hervorzuheben, während andere Gruppen in einem Konkurrenz- oder Konfliktverhältnis zur eigenen Gruppe positioniert werden (Ren et al., 2012, S. 842). Die erfolgreiche Herstellung einer solchen Identifizierung stabilisiert digitale Communities, intensiviert die Partizipation und trägt so wesentlich zum langfristigen Erhalt der Gruppe bei (Ren et al., 2012, S. 852-856).

In Hinblick auf die Herstellung einer Gruppenidentität in Online-Communities muss zudem ihr digitaler Kontext berücksichtigt werden. Die Akteure sehen sich bei der Arbeit zur Herstellung des Kollektivs auf digitalen Plattformen verschiedenen Rahmenbedingungen ausgesetzt, die in Hinblick auf ihre Arbeit sowohl als handlungsfördernd als auch -einschränkend wirken können. Diese Rahmenbedingungen werden sich von den Akteuren angeeignet und es wird versucht, sie im Sinne des Gruppenerhalts einzusetzen, wodurch der Aufbau, die Entwicklung und der Erhalt von Online-Communities wesentlich mit den technischen Rahmenbedingungen verwoben sind, in denen sie bestehen (Latour, 1990, S. 103). Dies bedeutet, dass Online-Communities nicht von ihrem digitalen Kontext getrennt verstanden werden können (Latour, 1990, S. 105).

Die Anzahl maskulinistischer Communities im Internet ist in den letzten Jahren stark angestiegen, sodass Nagle von einer *digitalen Gegenrevolution* spricht (ebd., 2018, S. 105). Die betreffenden Foren, Communities und Webseiten werden in der wissenschaftlichen Literatur oft als Manosphäre (*manosphere*) zusammengefasst, womit ein loser Zusammenschluss

digitaler Räume beschrieben wird, „*where men's perspectives, needs, gripes, frustrations and desires are explicitly explored*" (Farrell et al., 2019, S. 87). Userinnen sind in diesen Räumen meist implizit, durch sexistische Handlungsweisen innerhalb des Forums, oder explizit, d.h. durch offen einsehbare Regeln, die mitteilen, dass weibliche Perspektiven unerwünscht sind, ausgeschlossen (Farrell et al., 2019, S. 87). Auch in der Manosphäre lassen sich Communities mit unterschiedlichen Themenschwerpunkten finden. Hierbei reicht das Spektrum von Gruppen, die sich mit der mentalen Gesundheit von Männern befassen, bis hin zu höchst frauenfeindlichen Foren, die bereits in Verbindung zu Gewaltverbrechen standen (Nagel, 2018, S. 105; Solska & Sawicki, 2024, S. 172).

Der Fokus dieser Arbeit liegt hierbei auf nicht-radikalisierten Communities. Dieser Schwerpunkt wurde auf der Grundlage zweier wesentlicher Überlegungen gewählt. Zum einen gibt es in der Forschung zu Online-Communities bereits einen vielfältigen Erkenntnisstand (siehe Cameron-McKee, 2020; Copland, 2020; Massanari, 2017). Zum anderen führt die starke wissenschaftliche Auseinandersetzung mit Radikalisierung im Netz, die einen Extremfall innerhalb existierender Interaktionen mit digitalen Inhalten darstellt, bei gleichzeitiger Vernachlässigung von Nicht-Radikalisierung dazu, dass Radikalisierung als zentrale Umgangsform mit extremen Inhalten gesehen wird (Pilkington, 2025, S. 28). Hierbei wird ausgeblendet, dass auch nicht-radikalisierte Communities politisch wirkmächtig sein können, beispielsweise indem durch die vermehrte Interaktion mit antifeministischen Inhalten sexistische Einstellungen kultiviert werden (Fox et al., 2015, S. 440). Zudem stellt Radikalisierung keinen plötzlich eintretenden Zustand dar, sondern muss als Prozess begriffen werden, dessen zentraler Ausgangspunkt die Interaktion mit frauenfeindlichen Inhalten in nicht-radikalisierten Gruppen sein kann (Evans, 2018).

Die verschiedenen Communities, die unter dem Begriff der Manosphäre zusammengefasst werden, weisen sowohl Unterschiede als auch starke Gemeinsamkeiten auf. So teilen sie ihre maskulinistische Positionierung, insbesondere die antagonistische Haltung in Bezug auf ‚den Feminismus'. Zu dieser Positionierung gehört eine gemeinsame Sprache, die es ermöglicht, trotz verschiedener Schwerpunkte innerhalb der Gruppen eine gemeinsame Identität herzustellen (Marwick & Caplan, 2018, S. 553; Kracher, 2020, S. 238-251). Diese Homogenisierung maskulinistischer Rhetorik kann unter anderem auf die Affordanz des Teilens von Hyperlinks zurückgeführt werden, durch die verschiedene Communities der Manosphäre miteinander verknüpft wurden (Ging, 2019, S. 644-645). So wird Feminismus in mehreren Communities der Manosphäre als Misandrie bezeichnet, um darauf zu verweisen, dass Frauen aus maskulinistischer Sicht nicht strukturell benachteiligt sind, sondern

emanzipatorische Anliegen vermeintlich als Vorwand nutzen, um männerfeindliches Verhalten zu legitimieren (Marwick & Caplan, 2018, S. 554).

Trotz der vielen Gemeinsamkeiten können digitale Communities, die dem männlich-zentrierten Antifeminismus zugeordnet werden, anhand ihrer übergeordneten Zielsetzungen in zwei Kategorien unterteilt werden. Foren, die eine antifeministische Gegenbewegung anstreben, orientieren sich am öffentlichen Diskurs und zielen darauf ab, diesen zu beeinflussen, während maskulinistische Foren mit individualistischem Fokus sich mit konkreten Erfahrungen und Lebensstilentscheidungen von Männern befassen (Han & Yin, 2023, S. 1925). Die Handlungen ersterer sind hierbei eher auf reichweitenstarke Kommunikation und dem Eingehen strategischer Allianzen mit Akteuren der neuen Rechten ausgelegt. Individualistisch orientierte Gruppen sind eher von einzelnen Handlungen zur Abwertung inklusiver Geschlechternormen, Selbstoptimierung, aktiver oder passiver Belästigung von Frauen und der Verherrlichung von Gewalt geprägt (Han & Yin, 2023, S. 1935).

4. *Who's acting sexist here?:* Das Problem der agency in der Analyse antifeministischer Online-Communities

In ihrer Analyse sexistischer und rassistischer Hasssprache im Internet beschreibt Nakamura eine Tendenz in der alltagstheoretischen Auseinandersetzung mit diesen Phänomenen, innerhalb derer die Ursachen für hasserfüllte Handlungsweisen in der Struktur des digitalen Raums verortet werden. Nach diesem Erklärungsansatz führen die Affordanzen, die eine jeweilige Plattform ihren Nutzer*innen bietet, zu der weiten Verbreitung von Inzivilität im Netz. Affordanzen können hier als *„the range of functions and constraints that an object provides for, and places upon, structurally situated subjects"* verstanden werden (Davis & Chouinard, 2017, S. 1). So wird insbesondere die Möglichkeit zur Anonymität in der Kommunikation auf Plattformen als Grund dafür gesehen, dass viele Nutzer*innen inzivile Äußerungen tätigen (Nakamura, 2013). Nakamura kritisiert, dass die Perspektive „*it's not the actor, it's the network*" Nutzer*innen ihre agency nehmen würde (Nakamura, 2013). Zudem entstehe durch diese Argumentation die Illusion, dass Gelegenheiten, rassistisch oder frauenfeindlich zu agieren, hauptsächlich im Internet bestünden, wobei die Gegenwärtigkeit dieser Phänomene im analogen Alltag von Betroffenen ausgeblendet wird (Shaw, 2014, S. 274).

Entgegen dieser Kritik konnte für die digitalen Affordanzen der Anonymität und der Interaktivität auf Twitter tatsächlich gezeigt werden, dass diese eine enthemmende Wirkung

haben und Nutzer*innen unter diesen Handlungsbedingungen eher dazu neigen, sexistische Haltungen zu beziehen (Fox et al., 2015, S. 440). Zudem trägt die stärkere Sichtbarmachung ziviler Kommentare durch die Affordanzen einer Plattform dazu bei, dass weitere User*innen sich an den wahrgenommenen Normen orientieren (Engelmann et al., 2024, S. 931). Auf diese Weise können die Affordanzen einer Plattform sexistische Handlungen sowohl fördern als auch unterbinden.

Angesichts dieser zwei Perspektiven auf die Ursachen von Sexismus auf digitalen Plattformen ist eine Voraussetzung für eine fruchtbare wissenschaftliche Auseinandersetzung mit dieser Problemstellung die Berücksichtigung sowohl der technologischen als auch der sozialen Anteile eines Netzwerks. Auf diese Weise sollten Einflüsse von Affordanzen spezifischer Plattformen auf den dort präsenten Sexismus analysiert werden, ohne in einen technologischen Determinismus zu verfallen. Dies bedeutet, dass die Architektur digitaler Plattformen weder als Garantie für die Entwicklung einer egalitären Debattenkultur – wie es die Perspektive des Cyberfeminismus nahelegt – gesehen werden kann noch als alleinige Ursache für die Allgegenwärtigkeit von Sexismus und Frauenfeindlichkeit im Internet (Harvey, 2020, S. 119-120). Auf Grund der bereits angeführten wissenschaftlichen Erkenntnisse bezüglich der Auswirkungen von Plattform-Affordanzen auf die digitale Debattenkultur sollte auch das Verfolgen eines sozialen Determinismus vermieden werden, „*where technologies are seen as empty tools imbued with power solely in their use*“ (Harvey, 2020, S. 123). Alternativ zu diesen beiden Positionen erscheint es in der Erforschung eines soziotechnologischen Phänomens wie digitalen Plattformen sinnvoll, die wechselseitige Bedingung zwischen menschlichen und nicht-menschlichen Anteilen dieser Netzwerke in den Blick zu nehmen.

4.1 Plattform-Affordanzen als Teil des Akteur-Netzwerks

Plattformen stellen nicht einfach neutrale Werkzeuge dar, die zum Zweck der Kommunikation gebraucht werden können. Jede Plattform bietet spezifische Funktionen und Leistungen, um die Kommunikation zwischen User*innen zu ermöglichen (Van Dijck, 2013, S. 6). Durch die Interaktion zwischen User*innen und der jeweiligen Plattform werden in einer Co-Produktion Sozialität und Technologie hervorgebracht (Bucher, 2018, S. 2). Dies bedeutet, dass sowohl die Handlungen der Nutzer*innen wesentlich durch die Affordanzen einer Plattform bedingt werden als auch die User*innen durch ihr Handeln auf der Plattform die Affordanzen formen. So fördert die Architektur einer Plattform bestimmte Formen von Sozialität, während sie andere erschwert (Bucher, 2018, S. 4). Gleichzeitig ist es Nutzer*innen bis zu einem gewissen Grad

möglich, sich die technologischen Rahmenbedingungen zu eigen zu machen und die Plattform-Affordanzen auch entgegen ihrer vorgesehenen Nutzungsweise anzuwenden (Van Dijck, 2013, S. 6). Im Folgenden wird dieser Umstand durch das Konzept der programmierten Sozialität (*programmed sociality*) beschrieben. Der Begriff bezeichnet das wechselseitige Formen und Bedingen zwischen menschlichen und nicht-menschlichen Entitäten in der Kommunikation auf digitalen Plattformen, ohne von einem deterministischen Verhältnis zwischen diesen auszugehen (Bucher, 2018, S. 5). So wird es möglich, soziale Interaktionen in dieser Umgebung als soziotechnologisches Phänomen zu begreifen.

Durch diesen ständig ablaufenden Prozess des Formens, der Aneignung und Aushandlung in der Sozialität auf digitalen Plattformen kann davon ausgegangen werden, dass diese nie als abgeschlossen verstanden werden können. Digitale Plattformen unterliegen einem ständigen Prozess des Werdens (Van Dijck, 2013, S. 6). Auf diese Weise entsteht eine dialektische Beziehung zwischen den User*innen, ihren Nutzungsweisen und den damit einhergehenden Interessen in Bezug auf die technologische Beschaffenheit der Plattform und denjenigen, die über die Affordanzen der Plattform verfügen und technologische Veränderung umsetzen können (Squirrell, 2019, S. 1913). Die programmierte Sozialität, die auf Plattformen beobachtet werden kann, wird dementsprechend geprägt durch eine Vielzahl an menschlichen und nicht-menschlichen Entitäten, die auf das Netzwerk einwirken und es ständig verändern. Social-Media-Plattformen sind demnach zu verstehen als Ansammlungen von verschiedenen menschlichen und nicht-menschlichen Entitäten, die das jeweilige Netzwerk ausmachen, sich also gegenseitig bedingen, da sie ohne die jeweils anderen Entitäten selbst keine oder eine andere Position und Funktion innerhalb des Netzwerks hätten (Bucher, 2018, S. 50). Für die wissenschaftliche Auseinandersetzung mit Plattformen bedeutet diese Grundannahme, dass es nicht möglich ist, eine Handlung innerhalb des Netzwerks auf einen konkreten menschlichen oder nicht-menschlichen Anteil der Plattform zurückzuführen. Wenn sich die Entitäten gegenseitig bedingen, gehören immer schon mehrere Anteile des Netzwerks zu einer Handlung, sodass der Ursprung von agency niemals eindeutig zurückverfolgt werden kann. So kann auch nicht eindeutig geklärt werden, ob eine Handlung rein auf die technologischen Anteile des Netzwerks zurückzuführen ist oder ausschließlich auf die Nutzer*innen, da immer schon beides an einer Handlung beteiligt ist (Latour, 1990, S. 110; Bucher, 2018, S. 51).

Vielversprechender für wissenschaftliche Analysen soziotechnologischer Phänomene auf digitalen Plattformen ist es daher, danach zu fragen, welche menschlichen und technischen Anteile des Netzwerks als Akteure an einer Handlung beteiligt sind. Hierbei kann geklärt werden, ob die jeweilige Entität nur Impulse von Akteuren weiterleitet oder selbst einen

Unterschied macht und damit zum Handlungsträger wird (Latour, 2015, S. 123). Zudem ist es möglich herauszufinden, ob spezifische Anteile des Netzwerks bestimmte Handlungen nahelegen, während andere Handlungsweisen unterbunden werden. So beschreiben Gerrard und Thornham die Akteure der Content-Moderation auf den Plattformen Tumblr und Pinterest als sexistische Zusammenschlüsse (ebd., 2020, S. 1279). Sie argumentieren, dass die von ihnen untersuchten technologischen Anteile der beiden Plattformen normativ geprägt sind und somit implizit sexistische Wirkungen entfalten, indem sie in ihrem Zusammenspiel Handlungen fördern, die ein hierarchisches Geschlechterverhältnis stützen, da unter anderem durch technische Rahmenbedingungen weibliche Perspektiven weniger sichtbar werden oder Frauen durch die verstärkte Sichtbarkeit von Sexismus die Teilnahme an digitalen Diskursen erschwert wird (Gerrard & Thornham, 2020, S. 1280; Reich & Bachl, 2025, S. 12-13).

Die Affordanzen digitaler Plattformen konnten bislang allerdings nicht nur in Zusammenhang mit der Reproduktion von Sexismus sowie der Verringerung der Sichtbarkeit von Frauen gebracht werden. Wie Holm in ihrer Untersuchung maskulinistischer Blogs aufzeigen konnte, fördert die technologische Struktur der analysierten Plattformen antifeministische Positionierungen (ebd., 2023). Die Plattformen ermöglichen es Nutzern und den von ihnen verbreiteten maskulinistischen Inhalten Sichtbarkeit zu erfahren, ohne dem gatekeeping traditioneller Medien ausgesetzt zu sein. Zudem können sie durch die Verwendung von Hyperlinks dafür sorgen, dass ihr Netzwerk sich stetig und über Ländergrenzen hinweg vergrößert und sich eine gemeinsame Sprache zwischen den maskulinistischen Communities verbreitet und etabliert (Holm, 2023, S. 435-346; Ging, 2019, S. 644-645). Zudem trägt die Affordanz der Anonymität dazu bei, dass trotz eines wahrgenommenen sozialen Stigmas für antifeministische Positionierungen ebensolche in einer größeren Öffentlichkeit preisgegeben werden, da durch die Möglichkeit des anonymen Auftretens weniger Angst vor sozialen Sanktionen besteht. Die Affordanz der Pseudonymität ermöglicht es zudem, trotz Anonymität unter einem spezifischen Namen aufzutreten und auf diese Weise eine Persona innerhalb der antifeministischen Blogsphäre etablieren zu können (Holm, 2023, S. 437).

Sexismus und antifeministische Diskurse wurden bislang mit einem starken Fokus auf die Analyse der Medieninhalte derartiger Debatten untersucht (Holm, 2023, S. 423). Die in diesem Kapitel vorgestellten Befunde weisen allerdings darauf hin, dass das Phänomen des digitalen Antifeminismus durch die wechselseitige Prägung von Nutzungsverhalten und technologischen Rahmenbedingungen beeinflusst wird. Ziel dieser Arbeit ist es daher, das Zusammenwirken verschiedener menschlicher und nicht-menschlicher Anteile in der plattformbasierten Kommunikation sichtbar zu machen.

4.2 Affordanzen der Content-Moderation

Durch den großen Erfolg digitaler Plattformen im Rahmen der Etablierung des Web 2.0 findet seither auch ein Großteil sozialer Interaktionen im Internet innerhalb ebendieser digitalen Räume statt. Da die Plattformbetreibenden ein ökonomisches Interesse daran haben, dass diese für viele Menschen brauchbar bleiben, müssen sie verhindern, dass die Plattform und deren vorgesehener Anwendungszweck durch destruktive und inzivile Nutzungsweisen lahmgelegt wird (Seering, 2020, S. 2). Moderation von geposteten Inhalten ist demnach keine Leistung, die Plattformen anbieten können, wenn sie den dort stattfindenden Diskurs gesund und inklusiv halten wollen, sondern etwas, dass sie vornehmen müssen.

> „*Platforms must, in some form or another, moderate: both to protect one user from another, or one group form its antagonists, and to remove the offensive, vile, or illegal – as well as to present their best face to new users, to their advertisers and partners, and to the public at large*“ (Gillespie, 2018, S. 5).

Durch diese Notwendigkeit haben alle populären Plattformen Systeme entwickelt, mithilfe derer sie mit unerwünschten Inhalten umgehen können. Content-Moderation kann im klassischen Sinne wie folgt definiert werden: „*Moderation typically amounts to removing the content, making it inaccessible, and/or suspending or terminating the accounts of those who posted the content*“ (Goanta & Ortolani, 2021, S. 2). In diesem Sinne wird die Arbeit der Inhaltsmoderation erst im Nachhinein ausgeübt, d.h., sie wird immer dann angewandt, wenn ein unerwünschter Inhalt auf der Plattform gepostet wurde (Gillespie, 2018, S. 75).

Diese Perspektive schließt bereits wesentliche Teile des Systems der Content-Moderation ein. Allerdings lässt sich argumentieren, dass vorab festgelegte Regeln und die Art und Weise, wie sich eine Plattform in Bezug auf ihre gesamte Öffentlichkeitarbeit präsentiert, immer schon mit beeinflussen, welche Inhalte eher gepostet werden als andere. So werden community guidelines zwar auch genutzt, um Moderationsentscheidungen seitens der Plattform zu legitimieren, sie bieten allerdings auch eine Möglichkeit, im Vorhinein bekannt zu geben, welche Inhalte gewünscht sind und welche nicht (Gillespie, 2018, S. 45, 48-51). Des Weiteren konnte Duguay aufzeigen, wie die Präsentation von Plattformen in App-Stores, deren Aufbau, Angebote zu cross-posting sowie die Affordanzen zum Zweck der Produktion von Inhalten, die eine App anbietet, spezifische Nutzungserwartungen an die User*innen kommunizieren und wie diese unter queeren Nutzer*innen zu einer Anpassung an heteronormative Standards in ihren Handlungen auf der Plattform führen können (Duguay, 2016, S. 5-9). Dies bedeutet, dass in der

Analyse von Systemen der Content-Moderation auch Anteile des Netzwerks mit einbezogen werden sollten, die ihre Wirkung auf das Nutzungsverhalten möglicherweise ex ante entfalten.

In Hinblick auf Plattform-Affordanzen, die nachträgliche Moderation ermöglichen, konnte Are in ihrer Studie zu Flagging-Attacken auf Instagram und TikTok die Benachteiligung von Frauen und die verringerte Sichtbarkeit von feministischen Inhalten durch ebendiese aufzeigen (ebd., 2025). Flagging beschreibt hierbei eine Form der Content-Moderation, bei der Nutzer*innen selbst Inhalte an die Plattform melden, von denen sie denken, dass diese gegen die inhaltlichen Richtlinien der Plattform verstoßen (Crawford & Gillespie, 2016, S. 411). Im Rahmen von Flagging-Attacken melden Nutzer*innen allerdings entgegen dem vorgesehenen Anwendungszeck der Affordanz gezielt Inhalte an die jeweilige Plattform, nicht weil sie von einem Verstoß gegen die Content-Policies ausgehen, sondern weil sie die Sichtbarkeit spezifischer Inhalte aus politischen Motiven verringern möchten. Auf diese Weise eignen sie sich die Affordanz des Flagging an, um weibliche Perspektiven weniger sichtbar zu machen. Da die Plattformen nur wenige professionelle Content-Moderator*innen beschäftigen und für die meisten User*innen nicht persönlich erreichbar sind, können fehlerhafte Löschungen auf Grund von Massen-Flagging kaum angefochten werden (Are, 2025, S. 3589). Betroffen sind von diesen Attacken unter anderem vor allem Frauen und insbesondere Feministinnen. Somit stellt auch Flagging eine Affordanz dar, die genutzt werden kann, um Frauen und feministischen Inhalten Sichtbarkeit zu nehmen. Auch in diesem Fall kann von einer sexistischen Ansammlung aus menschlichen und technischen Anteilen der Plattformen gesprochen werden (Are, 2025, S. 3591-3592).

5. Maskulinistische Communities und die Affordanzen der Content-Moderation auf der Plattform Reddit

Innerhalb der Manosphäre sind es maskulinistische Unterforen auf Reddit (sog. *subreddits*), die besonders relevant für die digitale Mobilisierung gegen feministische Anliegen sind, da die Plattform zwischen stark top-down regulierten populären Angeboten wie Instagram und unregulierten Nischenangeboten wie 4chan positioniert ist (Massanari, 2017, S. 334). Antifeministische Subreddits sind in den letzten Jahren immer wieder durch ihre besonders hassgeprägte Sprache aufgefallen (Kaiser, 2021, S. 27, 38; Nagle, 2018, S. 113). Das hohe Ausmaß devianter Handlungsweisen auf Reddit sowie der geplante Börsengang des Unternehmens im Jahr 2024 haben dazu geführt, dass die Struktur der Plattform angepasst wurde. Insbesondere das Deplatforming stark toxischer Communities und ein überarbeitetes

System der Content-Moderation haben zuletzt für öffentliche Aufmerksamkeit gesorgt (Roose, 2024). Auf diese Weise ist es Reddit gelungen, besonders frauenfeindliche Communities zu verkleinern, wobei angenommen werden muss, dass es dabei zu einer Migration auf andere Plattformen kam, sodass nicht davon ausgegangen werden kann, dass die Maßnahmen dazu beigetragen haben, Hassrede im Internet generell zu verringern (Copland, 2020, S. 20-22). Darüber hinaus sind weiterhin antifeministische Gruppen auf Reddit aktiv. Im Gegensatz zu den stark radikalisierten Gruppen, welche vor dem Börsengang durch Reddit gesperrt oder stark reguliert wurden, zeigt sich bei den gegenwärtig noch aktiven Gruppen das Problem, dass sich hier eine wesentlich subtilere Form frauenfeindlicher Sprache etabliert hat, die selbst in der automatisierten Content-Moderation durch die Verwendung von large language models schwer von solchen Manosphäre-Foren unterschieden werden können, in denen sich in gemäßigter Form über Gleichstellungspolitiken ausgetauscht wird (Deligianni & Horne, 2023).

Massanari betont in ihrer Analyse misogyner Hasskampagnen auf Reddit, dass plattformspezifische Charakteristika zu einer stark frauenfeindlichen Kultur auf Reddit beitragen, diese sogar in vielen Fällen befördern (ebd., 2017, S. 336-337). In Bezug auf die Verstärkung misogyner Handlungen und Radikalisierung durch Reddits Affordanzen wurden bislang die Bewertungslogik von Postings, Reddits Standardeinstellungen in Hinblick auf die Sortierung von Inhalten, die Zusammenstellung der r/all-Startseite sowie das Sperren radikalisierter Communities für neue Nutzer*innen (Quarantäne) untersucht (Massanari, 2017; Copland, 2020). Hierbei konnte festgestellt werden, dass Reddits Affordanzen die Verbreitung antifeministischer Inhalte befördern. Die Quarantäne einzelner Communities unterbindet diese Entwicklung teilweise, indem Postings aus diesen Foren keinen neuen User*innen mehr vorgeschlagen werden, allerdings zeigen sich antifeministische Handlungen innerhalb der jeweiligen Subreddits mit gleichbleibender Intensität (Copland, 2020, S. 4).

Da antifeministische Subreddits das primäre Ziel eines Binnenaustauschs verfolgen (Ganz & Meßmer, 2015, S. 66), fehlt hierbei bislang die Betrachtung von plattformspezifischen Affordanzen, die sich primär auf die gruppeninterne Kommunikation beziehen, insbesondere des plattformspezifischen Systems der community-basierten Content-Moderation.

Das System der Content-Moderation auf Reddit unterscheidet sich wesentlich von den zentralisierten Top-down-Ansätzen anderer populärer Plattformen wie Instagram, Facebook oder YouTube, da die Betreibenden den einzelnen Foren in Bezug auf die Regulierung von Inhalten einen großen Spielraum einräumen. Anders als bei vergleichbaren Plattformen mit

einem ähnlich geringen Maß an Top-down-Moderation wie Telegram, basiert das Filtern von Inhalten auf Reddit auf einem System der community-basierten Moderation. Den Moderator*innen, die ihre Arbeit unbezahlt und auf freiwilliger Basis ausüben, werden seitens der Plattformbetreibenden mehrere Handlungsmöglichkeiten in Bezug auf den Umgang mit Fehlverhalten in den jeweiligen Foren bereitgestellt. Diese bestehen wesentlich aus der selbstständigen Ausarbeitung von Regeln für das Subreddit, dem Aussprechen von Warnungen sowie dem Entfernen von Inhalten und dem Reagieren auf User*innen-Anfragen sowie der Freigabe von Postings und der Anwendung von Mitteln der automatisierten Content-Moderation (Jhaver et al., 2019, S. 3). Für die Anwendung der Affordanzen der Content-Moderation konnten vergangene Untersuchungen aufzeigen, dass die genannten Maßnahmen meist in einer sich steigernden Form verwendet werden, beginnend mit sozialen Vorgehensweisen bis hin zu technischem Eingreifen (Seering, 2020, S. 10).

Die Community-Moderator*innen sind üblicherweise auch die Gründer*innen des jeweiligen Subreddits, oftmals unterstützt durch ein Team aus weiteren Mitgliedern des Forums, die sich ebenfalls freiwillig als Moderator*innen beteiligen (Squirrell, 2019, S. 1913). Diese stellen auch die Regeln für das Forum auf, welche als eine Art epistemischer Rahmen verstanden werden können, durch den eine mehr oder weniger stark definierte Deutungsweise des Inhalts des Forums sowie unterschiedlich strenge Vorgaben bezüglich des Posting-Verhaltens und des Umgangs innerhalb der Community festgelegt werden, die stark auf das Thema des jeweiligen Subreddits abgestimmt sind (Fiesler et al., 2018, S. 78; Squirrell, 2019, S. 1917).

Aus diesem Grund können die Handlungen der Content-Moderator*innen nicht primär als eine Maßnahme zur Verminderung von toxischem Verhalten verstanden werden, sondern stellen – wie im Weiteren ausgeführt wird – notwendige Arbeit zur Herstellung und zum Erhalt der Gruppe dar.

Da Reddits System der Content-Moderation aus einem Zusammenspiel der Community-Moderator*innen und den ihnen bereitgestellten manuellen und automatisierten Mitteln besteht, kann die Plattformstruktur und damit auch das System der Content-Moderation als Handlungsträger verstanden werden. Dafür soll im Rahmen dieser Arbeit die Perspektive der Akteur-Netzwerk-Theorie (ANT) nach Bruno Latour als zentrale Grundlage herangezogen werden.

Aus Sicht der ANT kann angenommen werden, dass Reddit-Communities ohne Handlungen wie das regelmäßige Posten und Kommentieren, festgelegte Themen und Regeln, die durch die Arbeit der Content-Moderator*innen aufrechterhalten werden müssen, Gruppengrenzen schnell

verschwimmen und die Gruppe nicht mehr als solche existiert. Der Gruppenerhalt durch ständige Arbeit kann insbesondere für digitale Communities eine Herausforderung sein, da mit der Popularisierung des Internets auch die Möglichkeit von destruktiven Handlungen wie *trolling* (gezieltes Stören und Provozieren anderer Nutzer*innen), *brigading* (koordinierte Aktionen, durch die eine organisierte Gruppe einzelne Nutzer*innen oder andere Gruppen belästigt) oder Hasskampagnen zugenommen hat, die unter anderem auch darauf abzielen können, Online-Communities aufzulösen oder zu unterwandern (Seering, 2020, S. 2).

Im Rahmen dieser Arbeit wird die community-basierte Content-Moderation als essenzieller Teil der gruppenerhaltenden Arbeit innerhalb von Reddit-Communities verstanden. Akteure können hierbei aus der Perspektive der ANT sowohl menschliche als auch nicht-menschliche Anteile des Netzwerks sein, solange ihr Beitrag zu dieser Gruppe einen Unterschied macht und sie nicht reine Zwischenglieder sind, die zwar Informationen weiterleiten, das Transportierte allerdings nicht verändern (Latour, 2015, S. 123). Hierbei kann also davon ausgegangen werden, dass die von Reddit bereitgestellten Affordanzen selbst entscheidende Handlungsträger sind, da sie einen Einfluss auf die Handlungen von Communities der Plattform haben und nicht nur den Austausch von Informationen zwischen den Mitgliedern der Subreddits ermöglichen (Massanari, 2017, S. 336 ff.). In der Untersuchung des Modells der community-basierten Content-Moderation auf Reddit kann hierbei im Sinne der ANT davon ausgegangen werden, dass nicht nur Moderator*innen als Mittler und somit als handlungstragende Akteure verstanden werden können, sondern auch die gesamte nicht-menschliche Struktur der Content-Moderation entscheidend zu Veränderungen des Netzwerks beiträgt. Anteile der Plattformstruktur können daher ebenso als Mittler verstanden werden.

Im vorliegenden Forschungsprojekt werden verschiedene Affordanzen in Hinblick auf ihre Rolle in der community-basierten Content-Moderation untersucht.

Ein wesentlicher Handlungsspielraum innerhalb des Systems der Content-Moderation ist die Möglichkeit, community-spezifische Regeln innerhalb des Subreddits aufzustellen. In Hinblick auf die Absicherung des Gruppenerhalts soll untersucht werden, inwiefern die thematischen Ausrichtungen der maskulinistischen Communities bereits in die Regeln der Subreddits eingeschrieben werden.

FF1: *In welchem Zusammenhang steht der thematische Fokus von maskulinistischen Communities auf Reddit mit den von ihnen im Rahmen der Community-Moderation verwendeten Regeltypen?*

Die Auswahl der Community-Regeln als zu untersuchende Affordanz erfolgte aus drei zentralen Erwägungen. Einerseits konnte in einer Untersuchung heterogener Subreddits festgestellt werden, dass es einen Zusammenhang zwischen der thematischen Ausrichtung von Reddit-Communities und deren verwendeter Regeltypen gibt (Fiesler et al., 2018, S. 78). Zudem kann anhand der Kategorisierung maskulinistischer Communities nach Han & Yin davon ausgegangen werden, dass individualistisch-orientierte Gruppen zu mehr Sexismus und Misogynie neigen (ebd., 2023, S. 1935). Geht man von einer normativen Perspektive in Hinblick auf die Funktionen von Content-Moderation aus, sollte diese Tendenz entsprechend in den Regeln der Foren berücksichtigt sein. Außerdem stellen die Regeln durch ihre prominente Platzierung innerhalb der Foren eine der ersten Affordanzen dar, durch die Nutzer*innen mit der communitybasierten Moderation in Kontakt treten.

Eine der Affordanzen, die in einer indirekten Beziehung zum System der Content-Moderation stehen könnte, ist das sog. Voting. Ähnlich wie die Like-Optionen, die andere populäre Plattformen anbieten, haben Nutzer*innen hierdurch die Möglichkeit, durch eine niederschwellige Interaktionsform ihre Zustimmung oder ihr Gefallen für einen Inhalt auszudrücken. In Hinblick auf Reddits Voting-System argumentiert Massanari, dass die Bevorzugung von Postings mit besonders hoher Anzahl von Upvotes durch den Reddit-Algorithmus fördert, dass Nutzer*innen dazu tendieren, vermehrt Haltungen auszudrücken, die in der jeweiligen Community bereits befürwortet werden während Gegenrede in Hinblick auf die im Subreddit hegemonialen Positionen durch das Zusammenspiel der Voting-Affordanz und der automatisierten Hierarchisierung von Inhalten in ihrer Sichtbarkeit stark eingeschränkt wird (Massanari, 2017, S. 337). Im Kontext des Forschungsinteresses dieser Arbeit ist es zusätzlich relevant zu klären, in welchem Verhältnis die Affordanz des Votings zum System der Content-Moderation steht. Da die Moderation von Communities von Mitgliedern des Forums ausgeführt wird, die gleichzeitig in Aushandlungsprozesse mit anderen Mitgliedern eingebunden sind und die Haltung dieser insofern berücksichtigen müssen, als dass die Existenz des Forums auf eine bestimmte Zahl aktiver Nutzer*innen angewiesen ist, könnte das Voting eine indirekte Möglichkeit zur Einflussnahme auf Moderationsprozesse darstellen.

FF2: *Welcher Zusammenhang besteht zwischen der Popularität eines Postings und der Löschentscheidung der Community-Moderatoren?*

Eine weitere zentrale Affordanz des Systems der Content-Moderation ist die Handlungsoption zur Entfernung von Kommentaren, welche Nutzer*innen in der Moderationsrolle innerhalb eines Forums zukommt. Hierbei handelt es sich im Vergleich zum Voting eher um ein

klassisches Werkzeug in der Regulierung von digitalen Inhalten, da im Gegensatz zum Voting ein unmittelbarer Einfluss auf die Verfügbarkeit der Inhalte besteht. Im Rahmen der Untersuchung ist von Interesse, ob dieser Handlungsspielraum genutzt wird, um den Fortbestand der Community abzusichern. Hierbei ist vor allem in Hinblick auf die normativen Ansprüche an Modelle der Content-Moderation relevant, welche Tendenzen sich für die Praxis der Community-Moderation in maskulinistischen Subreddits aufzeigen lassen und inwiefern diese normative Ansprüche an die Regulierung digitaler Inhalte erfüllt.

FF3: *Inwiefern besteht ein Zusammenhang zwischen der Löschung von Kommentaren durch Community-Moderatoren und dem Auftreten gruppenidentitätsgefährdender Inhalte in diesen Kommentaren?*

Durch die Beantwortung dieser Forschungsfragen soll herausgefunden werden, inwiefern Reddits Affordanzen im Kontext der community-basierten Moderation zum Erhalt maskulinistischer Communities genutzt werden können.

6. Methodische Vorgehensweise

Um die Verwendung der Affordanzen zum Zweck der Content-Moderation in maskulinistischen Reddit-Foren zu untersuchen, wurde eine quantitative Inhaltsanalyse durchgeführt, mittels derer die Regeln der Foren sowie das Voting-Verhalten und die Inhalte von gelöschten und moderierten Kommentaren untersucht wurden. Für die durchgeführte Analyse wurden zwei Stichproben gezogen. Einerseits wurden antifeministische Subreddits aus der Grundgesamtheit ausgewählt, um einen möglichen Zusammenhang zwischen der inhaltlichen Ausrichtung der Subreddits und deren Regeln zu finden. Für die tiefergehende inhaltliche Analyse von Kommentaren wurde ein besonders aktives Forum in Hinblick auf die tägliche Anzahl neuer Postings und Kommentare und einem hohen Zuwachs neuer Mitglieder aus der Grundgesamtheit ausgewählt. Für dieses Forum wurden für einen Zeitraum von drei Wochen alle Kommentare mittels Scraping heruntergeladen und in einem Datensatz gesammelt.

6.1 Stichprobe

Zur Durchführung der Analyse wurden zwei Stichproben gezogen, wobei die Stichprobe, die sich aus den Kommentaren eines spezifischen Forums zusammensetzte, für die inhaltliche Analyse aus forschungspragmatischen Gründen nochmals unterteilt werden musste. Für die

erste Stichprobe wurden antifeministische Subreddits aus der Grundgesamtheit ausgewählt, um deren Regeln herunterzuladen und diese zu codieren. Für die Untersuchung des Voting-Verhaltens und der inhaltlichen Analyse der Kommentare wurde dann ein spezifisches Forum ausgewählt, für das innerhalb eines Untersuchungszeitraums von drei Wochen alle geposteten Kommentare erhoben wurden. Für die statistischen Tests bezüglich des Votings wurden alle hierbei gewonnenen Kommentare in eine Stichprobe übernommen (N = 40.855). Da für die abschließende Untersuchung zur Beantwortung von Forschungsfrage 3 in einem wesentlich aufwendigeren Verfahren der Datengewinnung die Texte von entfernten Kommentaren regeneriert und codiert werden mussten, wurde für diesen letzten Schritt eine kleinere geschichtete Stichprobe aus dem gewonnenen Corpus der Kommentare gebildet. In dieses kleinere Sample wurden zufällig ausgewählte Kommentare aus der Gruppe der nicht-moderierten Inhalte und entfernte Kommentare aus der Gruppe der moderierten Inhalte, für die eine Regenerierung möglich war, aufgenommen (N = 432).

6.1.1. Sampling der antifeministischen Subreddits

Um eine repräsentative Stichprobe der relevanten Subreddits zu erstellen, wurde die Grundgesamtheit der Communities auf Reddit mit ähnlichem thematischem Fokus als eigenständiges digitales Netzwerk aufgegriffen.

Da digitale Communities die Charakteristika von small-world networks erfüllen, die durch ihre Nutzer*innen miteinander verbunden werden, und algorithmische Vorschläge von ähnlichen Communities dieses Phänomen noch verstärken, ist es möglich, durch exponentielles Schneeballsampling eine repräsentative Stichprobe zu erhalten, solange ein Netzwerk von Communities innerhalb der gleichen Plattform untersucht werden soll (Beauvais, 2023, S. 61). Für diese Sampling-Methode wurden relevante Communities innerhalb des antifeministischen Netzwerks als Ausgangspunkt verwendet, wobei die erste Auswahl alle von Han & Yin identifizierten inhaltlichen Ausrichtungen maskulinistischer Foren umfasste (ebd., 2023, S. 1935), damit sichergestellt werden konnte, dass auch Communities, die nicht ideologisch miteinander verknüpft sind und daher einen anderen Nutzerkreis aufweisen könnten, in die Grundgesamtheit aufgenommen wurden. Die ausgewählten Subreddits werden als Startpunkt für die Schneeballtechnik genutzt, um weitere ähnliche Subreddits zu finden. Hierfür schlägt Beauvais vor, eine Umfrage in den Foren zu posten, in der die User*innen ähnliche Communities angeben können. Dies ist bei einer Untersuchung maskulinistischer Foren keine geeignete Vorgehensweise, da Teilnahmeaufrufe zu wissenschaftlichen Untersuchungen in

diesen Subreddits meist unerwünscht sind und gezielt von den Moderatoren gelöscht werden. Stattdessen wurden innerhalb der Foren per Zufallsauswahl aktive User gewählt, deren öffentlich einsehbaren Mitgliedschaften auf weitere maskulinistische Subreddits untersucht wurden, um die Schneeballtechnik dann von diesen weiteren Foren aus fortzusetzen. Beendet wurde die Suche nach weiteren Communities, als keine neuen Foren mehr gefunden werden konnten (Beauvais, 2023, S. 62). Zusätzlich zur Orientierung an den von Han & Yin (2023) bereits identifizierten Communities und des exponentiellen Schneeballsamplings wurde eine manuelle Suche durch den Community-Filter der Suchleiste auf Reddit durchgeführt. Hierfür wurden Suchbegriffe verwendet, die mit der Manosphäre assoziiert sind, um die Vollständigkeit der Ergebnisse des Schneeballsamplings zu überprüfen und diese gegebenenfalls zu ergänzen (siehe Anhang, S. 71).

Durch diese Vorgehensweise konnten insgesamt 42 maskulinistische Subreddits gefunden werden, die eine Mindestgröße von 200 Mitgliedern aufweisen und in denen mindestens einmal pro Woche neue Postings hochgeladen werden. Auf Grund der geringen Anzahl wurden für die Analyse der Regeln maskulinistischer Communities alle gefundenen Foren in das Sample aufgenommen.

6.1.2. Sampling der Kommentare innerhalb einer ausgewählten Community (*r/shortguys*)

Da die Gewinnung einer größeren Anzahl gelöschter und moderierter Kommentare sich als eine der zentralen Hürden für die Durchführung des Forschungsprojekts darstellte, die durch eine aufwendige Form der Datengewinnung überwunden wurde, musste die Analyse der Moderationsentscheidungen und des Voting-Verhaltens auf ein ausgewähltes Subreddit beschränkt werden. Die Auswahl fiel hierbei auf das Reddit-Forum *r/shortguys*, da die Community innerhalb des Untersuchungszeitraums das größte Wachstum an neuen Mitgliedern verzeichnen konnte und eine wissenschaftliche Auseinandersetzung mit maskulinistischen Foren, die sich mit dem Thema Körpergröße befassen, bislang ausgeblieben ist. Daher erschien die Community *r/shortguys* als eine relevante Fallstudie.

Das Subreddit versteht sich als Community, die sich mit einer angenommenen sozialen Marginalisierung von Männern mit geringer Körpergröße (von der Community als *heightism* bezeichnet) und den hieraus abgeleiteten sozialen Problemstellungen, insbesondere in Hinblick auf das Geschlechterverhältnis, beschäftigt. Die Selbstbeschreibung von *r/shortguys* lautet:

„*r/ShortGuys is a space for short men to discuss height, heightism, and how it affects us*“ (r/shortguys, 2013). Dieser Beschreibung kann bereits entnommen werden, dass es in der Community nicht nur darum geht, individuelle Erfahrungen mit dem Thema Körpergröße zu teilen. So wie durch Squirrell beschrieben, liegt innerhalb des Forums ein epistemischer Rahmen vor, der bereits beim Einstieg in das Forum durch die zentrale Platzierung der Selbstbeschreibung, der Community-Highlights sowie durch die Regeln des Forums bekanntgegeben wird (ebd., 2019, S. 1916). Auf diese Weise werden negative Erfahrungen in Hinblick auf das Thema Körpergröße und Männlichkeit innerhalb der Community in eine Deutungsweise eingebettet, durch die von einer grundsätzlichen gesellschaftlichen Unterdrückung von Männern mit kleinerer Körpergröße ausgegangen wird. Alle Personen, die nicht zu dieser Gruppe von Männern gehören und sich an dieser Unterdrückung beteiligen, werden als „*monoliths*“ beschrieben. In Hinblick auf das Geschlechterverhältnis wird in diesem Deutungsrahmen davon ausgegangen, dass Frauen grundsätzlich kein Interesse an romantischen und sexuellen Beziehungen zu Männern mit kleinerer Körpergröße haben (siehe Abbildung 1).

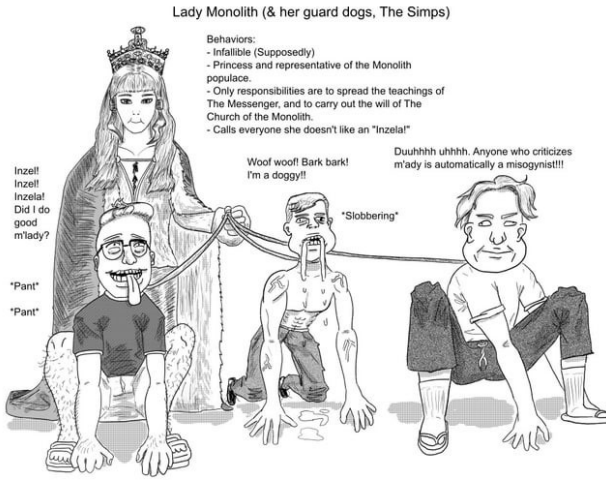

Abbildung 1: Beschreibung von „Lady Monolith“ aus dem *Shortguys Lorebook*

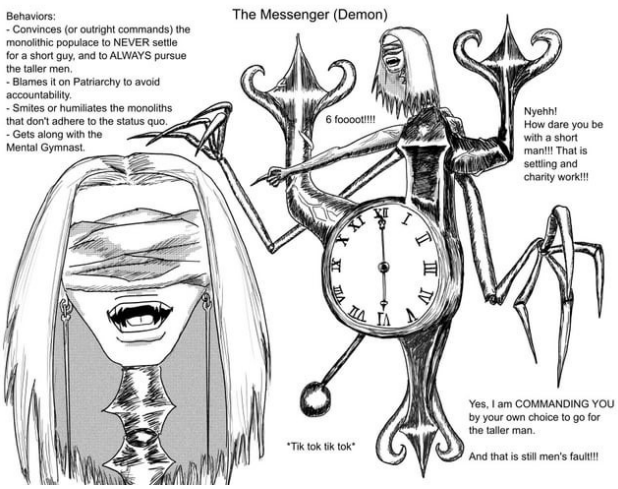

Abbildung 2: Beschreibung von „The Messenger“ aus dem *Shortguys Lorebook*

Dementsprechend gilt es, Frauen für diese Oberflächlichkeit zu kritisieren. Betroffenen Männern, die sich der in der Community vorherrschenden Perspektive nicht anschließen, wird vorgeworfen, dies vermeintlich ausschließlich aus dem Motiv heraus zu tun, positive Aufmerksamkeit von Frauen zu erhalten. Diese Deutungsweise kann unter anderem dem *Shortguys Lorebook* (siehe Abbildung 1 & 2) – einem durch die Community angefertigten

Comic, welches unter den Community-Highlights abgerufen werden kann und in dem die Deutung der Konfliktstellung, mit der sich die Community befasst, erläutert wird – entnommen werden (r/shortguys, 2025).

Als antifeministisch kann die Community *r/shortguys* bezeichnet werden, da das Framing und die Debatten innerhalb des Subreddits wesentlich von sexistischen Stereotypen geprägt sind, durch die davon ausgegangen wird, dass Frauen Beziehungen zu Männern primär aus oberflächlichen oder monetären Gründen eingehen würden. Zudem werden Sexismus und Misogynie nicht als reale Phänomene anerkannt, sondern als Abwehrmechanismen begriffen, die gebraucht werden, um kritische Einwände in Bezug auf das Verhalten von Frauen abzuwerten und zu entkräften (siehe Abbildung 2). Die Community kann zudem als maskulinistisch verortet werden, da sie darauf ausgerichtet ist, spezifisch männliche Interessen abzubilden.

In Bezug auf die theoretische Verortung der Community innerhalb der antifeministischen Foren auf Reddit lässt sich *r/shortguys* anhand der Kategorisierung von Han & Yin genauer als eine Community beschreiben, die an maskulinistischen Standpunkten orientierte, individualistische Diskurse hervorbringt, die einer community-eigenen Philosophie folgen und aus der ein spezifischer Lebensstil, insbesondere im Verhältnis zu und im Umgang mit Frauen, abgeleitet und propagiert wird, da die Mitglieder von *r/shortguys* dazu angehalten werden, Körpergröße als zentrale sozial-strukturierende Kategorie anzusehen und sich von Frauen zu distanzieren (Han & Yin, 2023, S. 1935).

6.2 Datengenerierung

Für die quantitative Inhaltanalyse wurden alle Regeln der 42 antifeministischen Subreddits mittels Scraping durch das R-Paket *jsonlite* heruntergeladen. So konnte ein erster Datensatz mit insgesamt 321 Regeln erstellt werden, wobei manuell im Datensatz vermerkt wurde, wenn ein Forum über keine Regeln verfügte.

In Hinblick auf die Analyse des Voting-Verhaltens sowie der Kommentare in *r/shortguys* wurde im Zeitraum 2. Januar 2025 bis 24. Januar 2025 viermal täglich ein Scrape der Kommentare des Forums vorgenommen. Dies erfolgte mit Hilfe des R-Pakets *RedditExtractor* (siehe Anhang, S. 76; Rivera, 2022; Hintz & Betts, 2022, S. 126-127). Auf diese Weise konnten alle während des Untersuchungszeitraums neu geposteten Kommentare für die Auswertung in einem Datensatz gesammelt werden, sodass die URL des Kommentars,

die Anzahl seiner Up- und Downvotes sowie der Text des Kommentars und seine Platzierung innerhalb des Threads (Diskussionsstrang aus mehreren aufeinander bezogenen Kommentaren) einsehbar waren. Wurde ein Kommentar durch den Kommentator selbst gelöscht, durch den automatisierten Reddit-Filter oder einen Community-Moderator entfernt, wurde dies im Datensatz sichtbar, da dies automatisiert vermerkt wurde. Im Textfeld des Kommentars wurde so anstelle eines Textes „[deleted]“ verzeichnet, wenn der Post von seinem Autor entfernt wurde, „[removed]“, wenn die Löschung durch einen Moderator herbeigeführt oder „[Removed by Reddit]“, wenn der Inhalt durch den automatisierten Filter von Reddit entfernt wurde. Da dieser Status allerdings nur Auskunft über die gegenwärtige Verfügbarkeit des Kommentars auf Reddit gibt und somit nicht ausreicht, um die inhaltlichen Charakteristika eines Kommentars zu erfassen und systematisch auswerten zu können, wurde eine Vorgehensweise entwickelt, die es ermöglicht, die Texte von gelöschten und entfernten Kommentaren zu regenerieren. Hierfür wurden die Kommentare der in den drei Wochen entstandenen 84 Datensätze (jeder Datensatz enthielt den Scrape des Forums zu einem bestimmten Tag und Uhrzeit) anhand ihrer Time-Stamps automatisiert durch R miteinander abgeglichen, um herauszufinden, ob für einen Kommentar, der in einem älteren Datensatz noch vollständig erhalten ist, zu einem späteren Zeitpunkt die Meldung „[deleted]“, „[Removed by Reddit]“ oder „[removed]“ verzeichnet wurde (siehe Anhang, S. 77-80). War dies der Fall, so konnte angenommen werden, dass der entsprechende Kommentar innerhalb des Zeitraums der Datengenerierung gelöscht oder entfernt wurde, sein Inhalt allerdings durch das intensive Scrapen des Forums rechtzeitig erfasst werden konnte. So war es möglich, Texte von insgesamt 460 gelöschten und entfernten Kommentaren zu rekonstruieren. Für die Analyse des Zusammenhangs zwischen den Inhalten der Kommentare und den Löschentscheidungen von Nutzern und Moderatoren wurde durch diese rekonstruierten Daten ein kleiner Datensatz mit insgesamt 450 verfügbaren, entfernten und gelöschten Postings erstellt. Diese Reduzierung erfolgte, da die Inhalte der Kommentare für die statistische Auswertung aufwendig codiert werden mussten, wodurch keine größere Stichprobe ermöglicht werden konnte. Die Untersuchung des Zusammenhangs zwischen dem Voting-Score (Summe aus Up- und Downvotes) eines Kommentars und seinem Moderationsstatus wurde separat anhand eines größeren Datensatzes durchgeführt, in dem alle 40.855 gesammelten Kommentare zusammengeführt wurden, da deren statistische Auswertung keine Codierung voraussetzte.

6.3 Entwicklung des Codebuchs und Codier-Prozess

Für das Codebuch wurden die verschiedenen thematischen Ausrichtungen der antifeministischen Subreddits, deren Verwendung von unterschiedlichen Regeltypen sowie die Inhalte der User-Kommentare unter Forum-Postings in Hinblick auf deren Aufweisen von gruppenidentitätsgefährdenden Inhalten und toxischem Verhalten operationalisiert, indem verschiedene Ausprägungen und die dazugehörigen Indikatoren festgelegt wurden. Anschließend wurden durch die Operationalisierung vier Leitfäden zur Codierung der jeweiligen Variablen entwickelt (siehe Anhang, S. 72-75).

Thematische Ausrichtung der Communities

Die Codierung der Themen der 42 verschiedenen Subreddits wurde mit Hilfe der Kategorisierung von Han & Yin (2023) vorgenommen. Diese untergliedert antifeministische Communities wesentlich in zwei dominante Gruppen, indem festgestellt wird, inwiefern sich die jeweilige Community um Diskurse organsiert, die darauf abzielen, die Meinungsbildung in Hinblick auf politische und rechtliche Themen, die Geschlechterverhältnisse betreffen, zu beeinflussen. Diese erste Gruppe der *antifeministischen Gegenbewegung* (Han & Yin, 2023, S. 1935) wird zudem nochmals in *organisierte* oder *unabhängige* Communities unterteilt, wobei erstere sich an antifeministischen Bewegungen und Organisationen orientieren, während letztere von diesen unabhängig agieren und nach Han & Yin wesentlich durch einzelne Journalisten oder Influencer angetrieben werden. Nach einer ersten Probecodierung wurde diesem Schema zudem die Kategorie der *unabhängigen antifeministischen Gegenbewegung* hinzugefügt, die sich weder an einer Organisation noch an Personen des öffentlichen Lebens orientieren.

Das Gegenstück zu den Diskursen der *antifeministischen Gegenbewegung* stellen maskulinistische Foren dar, in denen individualistisch geprägte Debatten verbreitet sind. Hierbei geht es darum, die individuelle Ebene der Geschlechterverhältnisse zu thematisieren. Auch diese Gruppe kann in zwei Untergruppen untersteilt werden. Die eher philosophisch geprägten Typen dieser Kategorie bewerben einen spezifischen Lebensstil und die Optimierung dessen. Es wird zudem zur Distanzierung von Frauen aufgerufen. Die eher pragmatisch ausgerichteten maskulinistischen Foren befassen sich dahingegen mit konkreten Lebenserfahrungen im Kontext der Geschlechterverhältnisse sowie dem Austausch von Ratschlägen in Hinblick auf diese spezifischen Erfahrungsberichte (Han & Yin, 2023, S. 1935).

Um alle Communities passend zuordnen zu können, wurde nach Abschluss der Probecodierung noch eine Gruppe für *gemischte Communities* hinzugefügt. Dieser wurden Subreddits zugeordnet, die maskulinistisch ausgerichtet sind, allerdings zu ähnlichen Anteilen gesellschaftspolitische und individualistische Diskurse integrieren.

Anhand dieser sechs Subkategorien und den im Codebuch festgelegten Indikatoren (siehe Anhang, S. 72) wurden die 42 Communities den entsprechenden Codes zugeordnet. Zur Orientierung dienten hierfür die auf Reddit angegebene Selbstbeschreibung der Community sowie die Materialien aus den jeweiligen Community-Highlights, welche den thematischen Fokus sowie den Zweck der Gruppen meist in detaillierterer Form schildern.

Regeln der Communities

Um die verschiedenen Regeln der Subreddits zu codieren, wurden ebenfalls Überkategorien gebildet, um Regeln mit gleichen Anliegen zusammenzufassen. Für die erste Ausarbeitung des Codierleitfadens diente die Einteilung von Fiesler et al. (2018, S. 77) zur Orientierung, die in ihrer Analyse von Regeln verschiedener Reddit-Foren bereits mehrere Regeltypen herausarbeiten konnten. Diese Typen wurden im Rahmen der Probecodierung um fehlende Typen ergänzt. Hierbei handelte es sich insbesondere um Regelformulierungen, die dazu dienen, die Gruppe nach außen abzugrenzen, sodass das Mitteilen einer Haltung, die von der in der Community vorherrschenden Deutungsweise abweicht, bereits als Regelverstoß gewertet werden kann. Ein Beispiel für eine derartige Regel stellt das Verbot von „*simping*“ oder „*white knighting*“ dar. Hierdurch soll das Fürsprechen oder die Verteidigung von Frauen innerhalb des Forums unterbunden werden. Zudem wurden Regeltypen ergänzt, welche die Verwendung von Neologismen anderer antifeministischer Gruppen verbieten, um die Community von diesen (insbesondere von Incels) abzugrenzen.

Abschließend wurden die gesammelten Regeltypen im Sinne des Forschungsinteresses ihrem Zweck nach in acht Kategorien zugeteilt, indem unterschieden wurde, ob die Regel die gewünschte Qualität der Postings, deren Format, die Orientierung an Reddits Content-Policy, das Vermeiden von Inzivilität oder gruppenspezifischer Hassrede, das spezifische Thema des Forums oder seine Abgrenzung zu anderen Gruppen vorgibt. Zuletzt wurde zudem eine Kategorie für Regeln gebildet, die die Rolle der Moderation in den Foren thematisieren (siehe Anhang, S. 73).

Codierung der Kommentare bezüglich störender Inhalte in Hinblick auf die Gruppenidentität der Community

Wie bereits im theoretischen Rahmen der Arbeit erläutert wurde, basieren Online-Communities auf einer Identität, die bis zu einem bestimmten Grad unter den Mitgliedern geteilt wird. Wird diese Identifizierung mit der Gruppe gestört, ist die Existenz der Community gefährdet. Da es für diese Arbeit von Interesse ist herauszufinden, inwiefern die Affordanzen der Content-Moderation auf Reddit genutzt werden, um das Bestehen maskulinistischer Communities abzusichern, wurden die mittels Scraping gewonnenen Kommentare auf gefährdende Inhalte in Hinblick auf die Gruppenidentität des digitalen Kollektivs untersucht und dementsprechend codiert. Zur Orientierung diente die Übersicht der Merkmale identitätsbasierter Zugehörigkeit zu Communities von Ren et al. (2012). Hierfür werden verschiedene theoretische Ansätze zu Gruppenidentitäten zu fünf zentralen Merkmalen von Gruppenidentitäten innerhalb von Online-Communities zusammengefasst (Ren et al., 2012, S. 843-844). Drei dieser Merkmale erwiesen sich während der Probecodierung für die Untersuchung antifeministischer Reddit-Communities als sinnvoll und konnten daher übernommen werden (siehe Anhang, S. 74). Da die Moderationsentscheidungen im Kontext des Gruppenerhalts analysiert wurden, wurden die Merkmale nach Ren et al. aufgegriffen und in Störungen, d.h. in Abweichungen von den Voraussetzungen für kollektive Identifizierung, umformuliert, sodass die Kommentare anhand von Eigenschaften kategorisiert wurden, die sich den Grundlagen der Gruppenidentität entgegenstellen und auf diese Weise eine Gefährdung für die Community darstellen.

Da die Mehrheit der Kommentare auf Grund des fehlenden Kontexts nicht für sich alleinstehend zugeordnet werden konnten, wurde bei der Zuordnung zu einem Code für jeden Kommentar mittels der im Datensatz enthaltenen URL und den Angaben zur Identifizierung der Platzierung eines Kommentars innerhalb eines Threads (insbesondere relevant für den Fall, wenn mehrere Kommentare innerhalb eines Threads gelöscht wurden) das dazugehörige Posting eingesehen sowie die innerhalb des gleichen Threads vorangegangenen und folgenden Kommentare einbezogen, um eine sinnvolle Einordnung vornehmen zu können.

Codierung der Kommentare bezüglich ihres Aufweisens von toxischem Verhalten

Um die Community-Moderation nicht nur in Hinblick auf ihren Umgang mit gruppenidentitätsstörenden Inhalten beurteilen zu können, sondern auch ihren Umgang mit Inhalten, die toxisches Verhalten aufweisen, zu erfassen, wurden die Kommentare aus der

geschichteten Stichprobe auch in Bezug auf ihr Aufweisen von Inzivilität und Unhöflichkeit codiert. Hierzu wurde ein Codier-Schema aus einer Studie von Papacharissi übernommen (ebd., 2004, S. 273-274, siehe Anhang, S. 75). Ein Fall von Inzivilität lag demnach vor, wenn ein Kommentar die Demokratie ablehnte, die Rechte anderer gefährdete oder Stereotype in der Beschreibung einer Gruppe von Personen verwendete. Ein Kommentar wurde als unhöflich bewertet, wenn der Inhalt eine persönliche Beleidigung, Übertreibung, eine Bezichtigung des Lügens oder vulgäre Formulierungen beinhaltete. Ziel dieser zusätzlichen Codierung ist es, die Praxis der Community-Moderation in ihrem Umgang mit gruppenidentitätsstörenden Inhalten und mit Inhalten, die toxisches Verhalten aufweisen, vergleichen zu können und ein zu starkes Überlappen dieser beiden inhaltlichen Dimensionen ausschließen zu können. Auf diese Weise sollte sichergestellt werden, dass die Ergebnisse tatsächlich ermöglichen spezifische Aussagen über den Umgang mit gruppenidentitätsstörenden und toxischen Inhalten als distinkte Phänomene treffen zu können.

Intra- und Intercoderreliabilität

Um die Zuverlässigkeit der Messung zu gewährleisten, wurde jeweils ein Drittel der Daten von der Hauptcodiererin mehrfach codiert sowie zusätzlich von einem zweiten Codierer zugeordnet, um die Intra- und Intercoderreliabilität berechnen zu können. Für die Codierung der Regeltypen konnte nach der ersten Probecodierung bereits ein Wert für CohensKappa von 1 für die Intracoderreliabilität und eine Intercoderrelibilität von 0.7 festgestellt werden, sodass der Codierleitfaden in dieser Fassung beibehalten wurde.

Nach der Probecodierung der Kommentare bezüglich ihres Aufweisens von gruppenidentitätsstörenden Inhalten, wurde zuerst nur eine Intercoderreliabilität von 0.31 errechnet, wonach eine Konkretisierung des Codebuchs vorgenommen wurde und Beispiele für jede Ausprägung angefügt wurden. Nach einer zweiten Probecodierung ergab sich eine Intercoderreliabilität von 0.54, was einer moderaten Übereinstimmung entspricht. Die Intracoderreliabilität belief sich auf 0.92. In Hinblick auf die Codierung der Kommentare nach ihrem Gehalt an toxischem Verhalten wurde eine Intracoderreliabilität von 0.95 und eine Intercoderreliabilität von 0.82 errechnet.

6.4 Positionalität der Forschenden

Die Forschende nimmt eine feministische Position ein, von der aus Geschlechtergerechtigkeit als ein zentrales und erstrebenswertes gesellschaftspolitisches Ziel angesehen wird. Das Forschungsprojekt ist auf dieser Haltung aufbauend entstanden, die Arbeit verfolgt daher ein emanzipatorisches Erkenntnisinteresse. Dies bedeutet, dass eine zentrale Zielvorstellung dieses Projekts ist, Wissen zu generieren, welches als Grundlage für die produktive Kritik der Architekturen populärer Plattformen genutzt werden kann sowie zur Entwicklung von Lösungen von Geschlechterungleichheit im Netz und zur Förderung konstruktiver Diskurse um geschlechtliche Gleichstellung beitragen soll.

Die Forschende ist selbst kein Teil der hier untersuchten Communities und bezieht daher eine außenstehende Position. Da die hier analysierten Gruppen entgegen der Positionierung der Forschenden antifeministische Interessen verfolgen, besteht die Gefahr, unbegründete Vorannahmen unbewusst in die wissenschaftliche Untersuchung zu übernehmen. Um dieser möglichen Tendenz entgegenzuwirken, wurde versucht, eine machtkritische Reflektion und Einordnung bezüglich des Phänomens des digitalen Maskulinismus in der Erläuterung des theoretischen Rahmens vorzunehmen (siehe Kapitel 3.1, S. 11-12). Zentral war es hierbei, eine Verortung der Phänomene Antifeminismus, Sexismus und Frauenfeindlichkeit als strukturell bedingte Problemstellungen vorzunehmen, um eine Verbindung zwischen den antiemanzipatorischen Handlungen individueller Akteure innerhalb der Communities mit der Makro-Ebene herzustellen. Die normative Bewertung der maskulinistischen Perspektive einzelner Nutzer oder Moderatoren ist kein Ziel dieser Arbeit und soll durch die Verbindung mit der strukturellen Dimension der untersuchten Problemstellung vermieden werden. Um dies insbesondere während der Codierarbeit, die vor allem eine intensive Auseinandersetzung mit den Inhalten der Kommentare aus der Reddit Community *r/shortguys* bedeutete, zu gewährleisten, wurde eine maximale Anzahl von Kommentaren, die am Stück codiert werden durften, vorab festgelegt.

Die außenstehende Position der Forschenden erwies sich zudem bei der inhaltlichen Interpretation der Kommentare als Einschränkung, da maskulinistische Communities vielfach gruppenspezifische Neologismen verwenden. Konnte ein Kommentar auf Grund derartiger Formulierungen nicht zugeordnet werden, wurde auf das Glossar von Kracher aus dem Jahr 2020 zurückgegriffen (Kracher, 2020, S. 238-251). Da maskulinistische Begriffe sich ständig verändern, mussten insgesamt 17 Kommentare aus der Stichprobe als nicht zuordenbar markiert werden.

6.5 Nutzung von künstlicher Intelligenz

Künstliche Intelligenz wurde als Hilfsmittel für zwei zentrale Arbeitsschritte in der Realisierung des Projekts angewendet. Der Chatbot *ChatGPT* konnte für die Erstellung von Codes für die Verwendung in R unterstützend eingesetzt werden. Hierbei handelte es sich insbesondere um die Korrektur von Fehlern innerhalb des Codes, der Erstellung von Listen, in denen eine größere Anzahl von Datensätzen mit ähnlichen Datei-Benennungen integriert werden musste sowie die Erstellung der Anweisung zur richtigen Gewichtung innerhalb der geschichteten Stichprobe (siehe Anhang, S. 86).

Zudem wurde nach einer ersten intensiven manuellen Literaturrecherche das auf künstlicher Intelligenz basierende Recherchetool *ResearchRabbit* verwendet, um aufbauend auf der bereits identifizierten relevanten Literatur weitere ähnliche Texte finden zu können. Auf die Verwendung von künstlicher Intelligenz zur Textgenerierung wurde gänzlich verzichtet, da die geringe Transparenz und die in KI-Sprachmodellen enthaltenen Biases nicht mit den Gütekriterien der Nachvollziehbarkeit und Reflexivität sowie dem emanzipatorischen Erkenntnisinteresse dieser Arbeit vereinbar sind.

7. Ergebnisse

Im Folgenden Kapitel werden die Ergebnisse der quantitativen Inhaltsanalyse vorgestellt. Ziel der Erhebung war es dabei herauszufinden, ob und wie Reddits Affordanzen der Content-Moderation den Erhalt maskulinistischer Communities ermöglichen. Hierzu wurden die Regeln der Foren, das Voting-System im Kontext der Moderation von Inhalten sowie die Inhalte von entfernten Kommentaren analysiert. Der Aufbau der Ergebnispräsentation orientiert sich hierbei an den Forschungsfragen und thematisiert zudem deskriptive Statistiken bezüglich der relevanten Variablen, um ein tiefergehendes Verständnis der betreffenden Online-Communities zu ermöglichen.

7.1 Deskriptive Analyse maskulinistischer Communitites

Um die Regeln der Communities zu untersuchen, wurde ein Sample aus insgesamt 42 maskulinistischen Reddit-Foren generiert. Damit hierauf folgend ein erster Überblick über die Gruppen gewonnen werden konnte, wurden vor der Erhebung deren thematische Schwerpunkte, die von den Communities verwendeten Regeltypen sowie Daten über die Anzahl deren Mitglieder, Regeln und Moderatoren und der Intensität ihrer Aktivität gewonnen

(siehe Tabelle 1). Hierbei kann festgestellt werden, dass die Anzahl der Mitglieder der Communities stark variiert, da die Spannweite von 200 bis 700.000 Mitgliedern reicht. Auch in der Intensität der Aktivität der Gruppen liegen bei einer Spannweite von einem Posting pro Woche bis zu 13.000 Beiträgen pro Woche große Unterschiede vor.

	Mittelwert	**SD**	**Median**	**Min**	**Max**
Anzahl Mitglieder	53070.48	128351.39	10500.0	200	700000
Aktivität (Postings pro Woche)	1678.19	3216.81	96.5	1	13000
Anzahl Regeln	8.10	4.32	8.0	0	15
Anzahl Moderatoren	5.95	4.87	4.0	1	26

Tabelle 1: Deskriptive Werte der maskulinistischen Subreddits (N = 42)

Die untersuchten Reddit-Foren weisen im Durchschnitt M = 8.10 Regeln auf (SD = 4.32, Md = 8), wobei die Anzahl der Regeln zwischen 0 und 15 liegt, und weisen durchschnittlich 5.95 Moderatoren auf (SD = 4.87, Md = 4), wobei die Gruppe mit einer Anzahl von 26 Moderatoren einen extremen Fall darstellt.

Die am längsten bestehende maskulinistische Community, die durch das Schneeballverfahren zur Suche relevanter Subreddits gefunden werden konnte, wurde im Jahr 2008 gegründet. Anhand der Gründungsjahre wird deutlich, dass seit diesem Zeitpunkt kontinuierlich maskulinistische Gruppen auf Reddit hinzugekommen sind (siehe Abbildung 3).

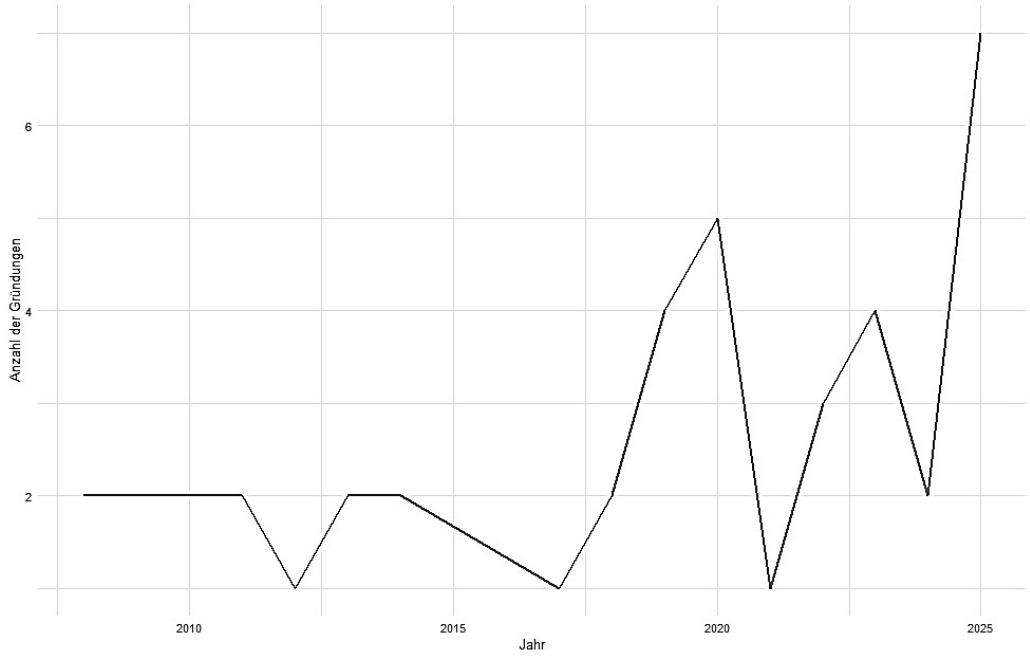

Abbildung 3: Übersicht der Gründungsjahre maskulinistischer Subreddits

Zu erkennen ist zudem, dass es um das Jahr 2017 einen ersten Anstieg gibt und die meisten der im Sample enthaltenen Communities im Jahr 2025 gegründet wurden.

7.2 Untersuchung des Zusammenhangs zwischen thematischer Ausrichtung und verwendeter Regeltypen

Um zu analysieren, ob die Affordanz der Gestaltung community-eigener Regeln den Erhalt antifeministischer Foren auf Reddit stützt, wurden die Themen der Communities sowie deren verwendete Regeltypen erhoben. Der Zusammenhang der Variablen wurde mittels einer Kreuztabelle untersucht. Hierfür wurden die Regeln anhand ihres Zwecks in acht Typen unterteilt (siehe Abbildung 4).

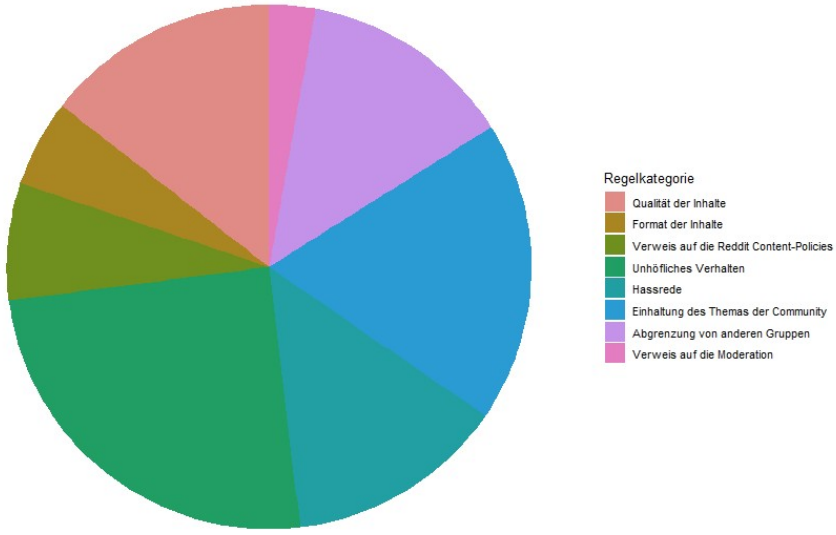

Abbildung 4: Vorkommen der verschiedenen Regeltypen (N=318)

Der am häufigsten vertretene Regeltyp sind solche Regeln, die höfliche Umgangsformen thematisieren (24,84 %). Regeln, die das Kernthema der Community erklären und auf die Einhaltung des Themenfokus plädieren, machen mit 18,55 % den zweitgrößten Anteil der Stichprobe aus. Hierauf folgen Vorgaben zu den Standards, welche die Community bezüglich der Qualität von Inhalten anwendet (14,47 %), Regeln, welche die Verwendung verschiedener Arten von Hassrede verbieten (13,52 %) sowie die Abgrenzung zu anderen Gruppen (13,21%). Am wenigsten verwendet werden Regeln, die auf die Content-Policies von Reddit verweisen (7,23 %), die Formate für gepostete Inhalte vorgeben (5,34 %) und solche, die die Befugnisse und die Funktionen der Community-Moderation bekanntgeben (2,83 %). Werden die Regeltypen, die sich mit der Regulierung des Verhaltens innerhalb der Community befassen, zusammengenommen (unhöfliches Verhalten und Verwendung von Hassrede), machen diese etwas mehr als ein Drittel der gesamten verwendeten Regel aus.

Die untersuchten maskulinistischen Communities wurden anhand ihres Themenschwerpunkts, welcher der Selbstbeschreibung des jeweiligen Reddit-Forums entnommen wurde, sechs Kategorien zugeordnet. Dabei fokussieren *Communities der antifeministischen Gegenbewegung* gesellschaftspolitische Strategien im Umgang mit Geschlechterverhältnissen und befassen sich eher mit dem öffentlichen Diskurs um Feminismus und Gleichstellungspolitik. Sie sind entweder mit einer sozialen Bewegung assoziiert, fokussieren ihre Debatten auf eine spezifische Person des öffentlichen Lebens (meist maskulinistische Influencer) oder sind von diesen beiden Akteuren unabhängig zu verstehen. Die

individualistisch orientierten Gruppen befassen sich mit persönlichen Erfahrungen in Bezug auf Geschlechterverhältnisse und entwickeln Strategien im Umgang mit vergeschlechtlichten Interaktionen. Diese sind entweder an einer community-eigenen Philosophie orientiert und verweisen auf einen spezifischen Lebensstil, der innerhalb der Gruppe angestrebt wird, oder sind pragmatisch ausgerichtet, wobei sich die Debatten eher an den spezifischen geschilderten Erfahrungen ausrichten.

Thema der Community	**Absoluter Anteil**	**Relativer Anteil**
Antifem_organisiert	44	0.13836478
Antifem_personenzentriert	8	0.02515723
Antifem_unabhängig	115	0.36163522
Individualistisch_phil	29	0.09119497
Individualistisch_prag	79	0.24842767
Gemischte Themen	43	0.13522013

Tabelle 2: Deskriptive Werte Themenfokus der Communities (N = 42)

Mit etwa 52,52% Anteil an der Stichprobe stellen die auf einen antifeministischen Gegendiskurs ausgerichteten Communities die größte Gruppe innerhalb der untersuchten Foren dar. Die individualistisch geprägten Communities machen mit 33,96% etwa ein Drittel des Samples aus. Foren wurden den gemischten Themen zugeordnet, wenn die Selbstbeschreibung nicht eindeutig auf eine antifeministische oder individualistische Ausrichtung hinwies. Diese Zuordnung trifft für 13,52% des Samples zu (siehe Tabelle 2).

Nachdem für die Stichprobe alle deskriptiven Werte betrachtet wurden, wurde eine Kreuztabelle für die Variablen des Themenfokus der Communities und deren verwendeter Regeltypen erstellt und die Voraussetzungen für einen Chi^2-Test getestet. Da zu viele der erwarteten Häufigkeiten innerhalb der Kreuztabelle einen Wert ≤ 5 aufwiesen, waren die Voraussetzungen für diesen Test nicht erfüllt. Alternativ wurde daher ein Fisher-Test durchgeführt. Dieser wies für den Zusammenhang zwischen dem Themenfokus und den verwendeten Regeltypen $p = 0.1874$ auf, womit von keinem Zusammenhang zwischen diesen Variablen ausgegangen werden kann, da der Fisher-Test ein nicht-signifikantes Ergebnis zeigt.

7.3 Deskriptive Statistiken bezüglich der Voting-Scores in der Community *r/shortguys*

Die Votes eines Postings stellen für Inhalte auf Reddit einen ähnlichen Wert dar, wie die Anzahl der Likes auf anderen populären digitalen Plattformen. Da es auf Reddit sowohl die Möglichkeit gibt, einen Post mit positiven Votes (Upvotes) oder negativen Votes (Downvotes) zu bewerten, wurde in der statistischen Auswertung des Voting-Verhaltens mit dem Gesamtscore, also der Summe aus Up- und Downvotes, gerechnet. Der Durchschnitt des Voting-Scores des Samples von N = 40.855 Kommentaren liegt bei M = 6.66, SD = 12.03 und einem Median von 3. Die Spannweite der Gesamtscores liegt zwischen -73 für den Kommentar mit den meisten Downvotes und 220 für den Kommentar mit der höchsten Anzahl an Upvotes.

Um die Beziehung zwischen dem Status der Verfügbarkeit und dem Voting-Score eines Postings zu analysieren, wurde die Korrelation zwischen der Variable „Verfügbarkeit des Postings“ mit den Ausprägungen „available“ für Postings, die noch über das Forum abgerufen werden können, „deleted“ für Postings, die von den Erstellern des Postings selbst gelöscht wurden, „Removed by Moderator“ für Postings, die durch Community-Moderatoren entfernt wurden, sowie „Removed by Reddit“ für Postings, die durch die automatisierte Moderation von Reddit entfernt wurden, und der Variable „total score“ für die Summe der Up- und Downvotes eines Postings mittels Varianzanalyse untersucht (siehe Tabelle 3). Die Sample-Größe betrug auch hier weiterhin N = 40.855 Kommentare.

	N	**Mittelwert**	**SD**	**Min**	**Max**
available	39470	6.82	11.97	-73	220
deleted	691	5.27	12.27	-33	91
Removed by Moderator	654	-0.74	12.79	-44	161
Removed by Reddit	40	1.88	4.37	0	28

Tabelle 3: Deskriptive Werte der Voting-Scores nach Status der Verfügbarkeit der Kommentare (N = 40.855)

Der größte Anteil der Stichprobe besteht aus verfügbaren Postings, deren Anteil 96,61 % des Samples ausmacht. Darauf folgen die von den Nutzern selbst gelöschten Kommentare mit 1,69% und die durch Community-Moderatoren entfernten Inhalte mit 1,60 % Anteil des Samples. Die kleinste Gruppe stellen die Kommentare dar, die durch die automatisierte Reddit-Moderation gelöscht wurden. Ihr Anteil macht 0,10 % der Stichprobe aus. Den höchsten Voting-Score weisen die verfügbaren Kommentare mit einem Mittelwert für den Gesamtscore von 6.82 auf, wobei die von den Nutzern selbst gelöschten Inhalte mit einem Score von M =

5.27 den Votes der verfügbaren Inhalte noch relativ nahe liegen. Ein deutlich größerer Unterschied wird allerdings zu den Gesamtscores der moderierten Inhalte sichtbar. Die von Reddit automatisiert entfernten Inhalte weisen nur noch einen durchschnittlichen Gesamtscore von 1.88 Votes auf. Der niedrigste Mittelwert des Voting-Scores zeigt sich bei den Kommentaren, die von Community-Moderatoren entfernt wurden, mit einer durchschnittlichen Summe aus ihren Up- und Downvotes von -0.74. Ein durchschnittlicher Kommentar, der von den Community-Moderatoren entfernt wird, weist demnach mehr Downvotes als Upvotes auf.

7.4 Analyse der Interaktion zwischen dem Voting-Score und der Verfügbarkeit der Kommentare in *r/shortguys*

Um die Beziehung zwischen den Variablen des Voting-Scores und des Status der Verfügbarkeit eines Kommentars zu analysieren, wurde ein Levene-Test zur Überprüfung der Annahme der Varianzhomogenität durchgeführt. Dieser fiel mit p = 0.0068 ** signifikant aus, wodurch die Voraussetzungen für eine Varianzanalyse nicht gegeben waren. Alternativ wurde daher ein Welch-Test durchgeführt. Dieser ergab eine p-Wert von < 2.2e-16 ****. Demnach liegt ein höchst signifikanter Unterschied zwischen den Mittelwerten der Voting-Scores und der verschiedenen Gruppen des Moderationsstatus vor.

Bei der Berechnung der Effektgröße mittels Eta-Quadrat ergab sich ein Wert von $\eta^2 = 0.61$, womit von einem starken Effekt zwischen den Variablen Voting-Score und Verfügbarkeit des Kommentars ausgegangen werden kann.

Anhand von Abbildung 5 wird sichtbar, dass alle Gruppen, die entfernte Kommentare zusammenfassen, durchschnittlich einen geringeren Voting-Score aufweisen als die Gruppe der Kommentare, die nicht entfernt wurden und weiterhin verfügbar sind. Die größte Diskrepanz zwischen den durchschnittlichen Voting-Scores liegt zwischen der Gruppe der verfügbaren Kommentare und den Kommentaren, die von Community-Moderatoren entfernt wurden, vor.

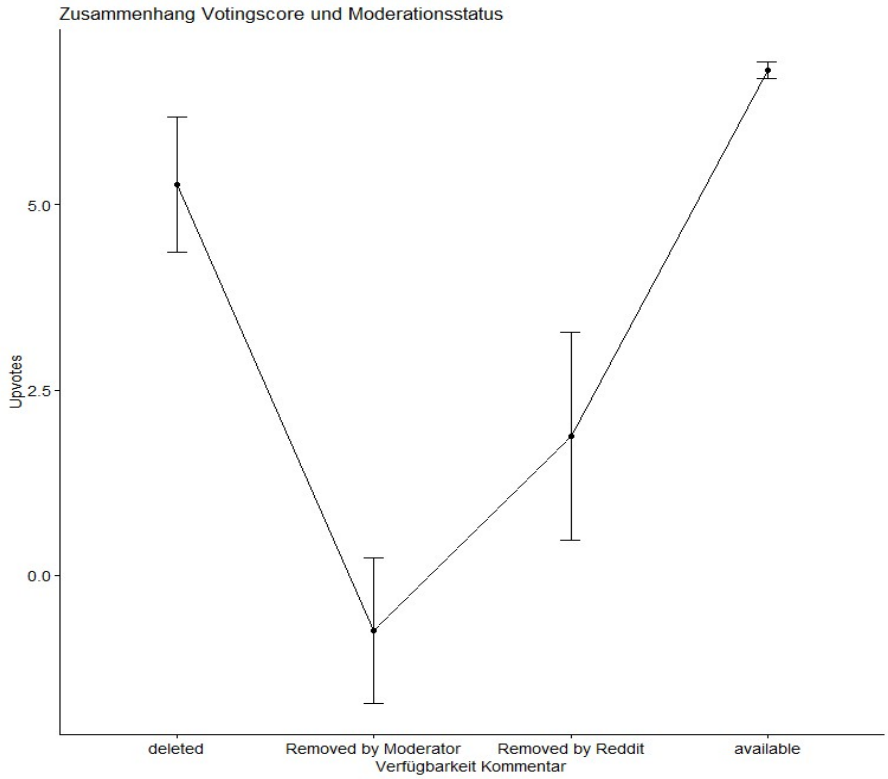

Abbildung 5: Verteilung der Mittelwerte der Voting-Scores nach Moderationsstatus (N = 40.855)

Nach dem auf Grund des Welch-Tests Games-Howell als Post-hoc-Test berechnet wurde, da nicht von Varianzhomogenität ausgegangen werden konnte, zeigte sich, dass zwischen allen Gruppen signifikante Unterschiede für den Voting-Score vorliegen (siehe Tabelle 4).

Moderationsstatus Gruppe I	**Moderationsstatus Gruppe II**	**Mittelwert-differenz**	**Signifikanz (p-Wert)**	**95%-Konfidenzintervalle**	
				Untergrenz.	**Obergrenz.**
Removed by Moderator	Deleted	-6.012213	< .001 ****	-7.7717945	-4.252631
Removed by Moderator	Available	7.557298	< .001 ****	6.2597658	8.854831
Removed by Moderator	Removed by Reddit	2.616590	0.015 *	0.3816593	4.851521
Available	Removed by Reddit	4.940708	< .001 ****	3.0788277	6.802589
Available	Deleted	1.545086	0.006 **	0.3334858	2.756686
Deleted	Removed by Reddit	-3.395622	< .001 ***	-5.5839993	-1.207245

Tabelle 4: Post-hoc-Test (Games-Howell) für die Gruppenunterschiede nach Verfügbarkeit der Kommentare bezüglich ihres Voting-Scores (N = 40.855)

Die Gruppen, für welche die größten Differenzen zwischen ihren jeweiligen Mittelwerten und mit p-Werten von < .001 extrem signifikante Ergebnisse angegeben werden können, sind die Unterschiede zwischen den Gruppen der verfügbaren Postings und solchen, die entweder von Community-Moderatoren oder dem automatisierten Reddit-Filter entfernt wurden, sowie dem Unterschied zwischen der Gruppe der von den Nutzern entfernten Postings und der von Community-Moderatoren entfernten Kommentaren. Zudem zeigt der Unterschied zwischen den selbst gelöschten Kommentaren und den von Reddit moderierten Inhalten einen hochsignifikanten Effekt. Anhand dieser Ergebnisse kann davon ausgegangen werden, dass Kommentare aus den beiden moderierten Gruppen signifikant geringere Voting-Scores aufweisen als Kommentare aus den beiden anderen Gruppen. Zudem zeigte sich selbst zwischen den moderierten Inhalten mit $p = 0.015$ ein signifikanter Effekt, d.h., dass die von den Community-Moderatoren entfernten Inhalte einen signifikant niedrigeren Voting-Score aufweisen als solche, die von Reddit gelöscht wurden. Ein geringerer Effekt zeigt sich zwischen den verfügbaren und den durch Selbstlöschung entfernten Kommentaren, wobei die selbstgelöschten Inhalte durchschnittlich einen kleinen Unterschied von einem um 1,55 Votes geringeren Score aufweisen als die verfügbaren Inhalte. Der kleinste Effekt zeigte sich zwischen den von Reddit und den Community-Moderatoren entfernten Inhalten mit einem durchschnittlichen Unterschied von 2,62 Votes.

Insgesamt lässt sich anhand dieser Ergebnisse konstatieren, dass insbesondere solche Kommentare, die von Moderatoren entfernt wurden, und die von Reddit automatisiert gelöschten Kommentare einen deutlich geringeren durchschnittlichen Voting-Score aufweisen als solche, die weiterhin über das Forum abgerufen werden können.

7.5 Deskriptive Statistiken zu den inhaltlichen Charakteristika moderierter Kommentare aus *r/shortguys*

In einem letzten Schritt der Analyse wurden die Variablen des Moderationsstatus der Kommentare sowie deren Aufweisen von gruppenidentitätsgefährdenden Inhalten und des Vorhandenseins toxischen Verhaltens untersucht. Hierfür wurde auf eine geschichtete Stichprobe zurückgegriffen. Dieser Schritt wurde vorgenommen, da wie in Tabelle 3 sichtbar wurde, die durch Moderatoren entfernten Kommentare insgesamt nur 1,60% der Kommentare ausmachen. Durch eine Analyse mittels Stichprobe ohne Schichtung wäre es somit nicht möglich, Aussagen über diese Kommentare zu treffen, da nur sehr wenige von ihnen in der Stichprobe enthalten wären. Die geschichtete Stichprobe setzt sich aus zwei Dritteln solcher

Kommentare, die zufällig aus der Gruppe der nicht-moderierten Postings aufgenommen wurden, und einem Drittel zufällig gewählter Postings aus der Gruppe der moderierten Kommentare zusammen. Die Schichtung wurde bei der statistischen Analyse berücksichtigt und entsprechend des Vorkommens der Gruppen in der Grundgesamtheit (98,40% nicht-moderierte und 1,60% moderierte Kommentare) gewichtet.

Die Variable des toxischen Verhaltens wurde in zwei Ausprägungen unterteilt. Toxizität weisen demnach Kommentare auf, die entweder eine Form der Inzivilität durch Ablehnung der Demokratie, der Rechte anderer oder Generalisierungen und Stereotype in Bezug auf Personengruppen enthalten oder unhöfliches Verhalten durch Beleidigungen, Bezichtigungen des Lügens, Übertreibungen oder vulgäre Sprache aufwiesen. Kommentare, die weder Inzivilität noch Unhöflichkeit aufwiesen, wurden der Ausprägung „kein toxisches Verhalten“ zugeordnet.

	Absoluter Anteil	**Relativer Anteil**
Toxisches Verhalten	183	0.4236111
Kein toxisches Verhalten	249	0.5763889

Tabelle 5: Deskriptive Werte der Variable „toxisches Verhalten“ (N = 432)

Die Stichprobe setzt sich aus 57,64% Kommentaren, die kein toxisches Verhalten aufweisen und 42,36% Kommentaren, die toxisches Verhalten enthalten, zusammen (siehe Tabelle 5). In Hinblick auf die Variable der gruppenidentitätsgefährdenden Inhalte weist die größte Anzahl der Kommentare keine störenden Inhalte auf (69,21%). Unter den störenden Inhalten treten mit 18,98% am häufigsten solche auf, die der Annahme der Homogenität innerhalb der Community widersprechen (siehe Tabelle 6).

	Absoluter Anteil	**Relativer Anteil**
Störung der Kategorisierung	13	0.03009259
Störung der Homogenität	82	0.18981481
Störung der Abgrenzung	38	0.08796296
Keine Störung	299	0.69212963

Tabelle 6: Deskriptive Werte der Variable gruppenidentitätsgefährdende Inhalte (N = 432)

7.6 Untersuchung der Interaktion zwischen dem Moderationsstatus, toxischem Verhalten und gruppenidentitätsstörender Inhalte der Kommentare aus *r/shortguys*

Um die Zusammenhänge zwischen dem Moderationsstatus, dem toxischen Verhalten und den gruppenidentitätsstörenden Inhalten zu untersuchen, wurde zunächst eine mögliche Dreifachinteraktion zwischen diesen Variablen überprüft. Zu diesem Zweck wurde ein log-lineares Modell mittels dreidimensionaler Kontingenztabelle gerechnet. Dieses Modell wies mit $p = 0.10948$ keinen signifikanten Zusammenhang für die Interaktion zwischen den drei Variablen auf, weshalb die Beziehungen zwischen den Variablen in einem nächsten Schritt durch Chi^2-Tests anhand von separierten Kreuztabellen untersucht wurden. Die Voraussetzung für diese Berechnung wurden mittels Einsicht der jeweiligen erwarteten Häufigkeiten überprüft und waren gegeben.

Für die Analyse des Zusammenhangs zwischen dem Vorkommen von toxischem Verhalten in Kommentaren und deren Moderationsstatus zeigte sich mit $p = 0.26$ kein statistisch signifikanter Zusammenhang zwischen diesen beiden Variablen. Kommentare, die toxisches Verhalten aufweisen, werden somit nicht öfter oder seltener von den Community-Moderatoren entfernt als solche, die kein toxisches Verhalten aufweisen.

In Abbildung 6 wird für das Verhältnis zwischen dem Moderationsstatus und toxischem Verhalten deutlich, dass die toxischen und nicht-toxischen Inhalte sehr ähnlich über die Gruppen der moderierten und der nicht-moderierten Kommentare verteilt sind. Moderierte Kommentare weisen daher nicht häufiger toxische Inhalte auf als solche, die weiterhin über das Forum abgerufen werden können.

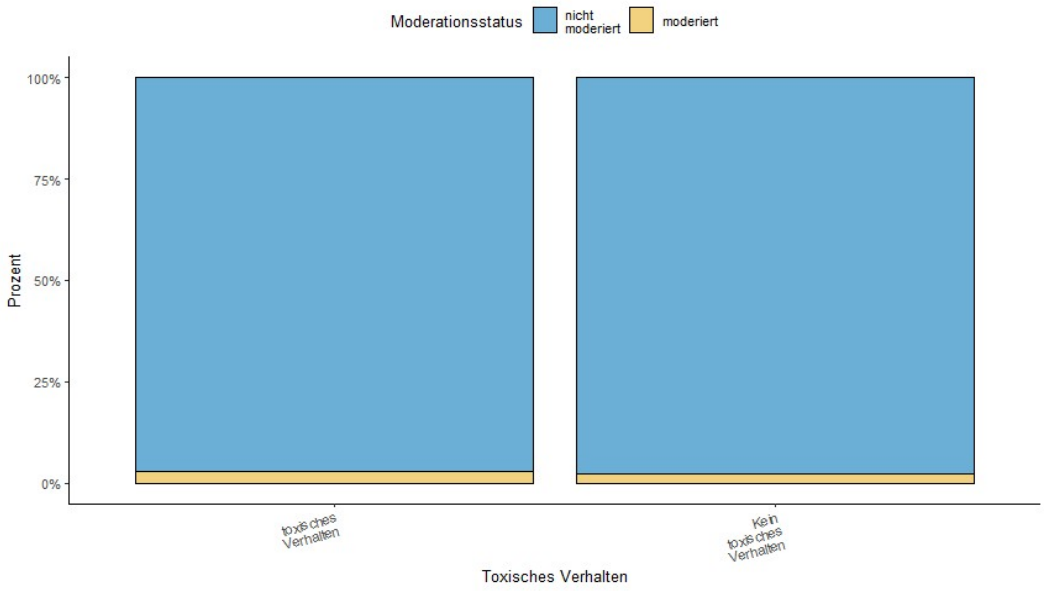

Abbildung 6: Toxisches Verhalten nach Moderationsstatus in Prozent (N = 432)

In einem zweiten Schritt der Analyse wurde, nachdem für die toxischen Inhalte und dem Moderationsstatus kein signifikanter Zusammenhang festgestellt werden konnte, auf einen Zusammenhang zwischen den identitätsstörenden Inhalten und dem Moderationsstatus getestet.

Anhand der Darstellung des Zusammenhangs in Abbildung 7 lässt sich zunächst erkennen, dass vor allem solche Kommentare, die einen identitätsgefährdenden Inhalt in Hinblick auf die Annahme der Homogenität der Community enthalten, einen deutlich größeren relativen Anteil der moderierten Kommentare ausmachen als in der Gruppe der nicht-moderierten Kommentare. Insgesamt machen die Kommentare mit gruppenidentitätsstörenden Inhalten mit 50,35 % etwa die Hälfte der Gruppe der moderierten Inhalte aus. Diese unterteilt sich in die Gruppe der Störung der Homogenisierung, die mit 34,27 % die größte Gruppe der moderierten Inhalte mit gruppenidentitätsstörendem Inhalt ausmacht. Hierauf folgt die Gruppe der Störung der Abgrenzung der Gruppe zu anderen Gruppen mit einem Anteil an den moderierten Kommentaren von 10,49 % und die Störungen der Kategorisierung, die mit 5,59 % den kleinsten Anteil der durch Moderatoren entfernten Kommentaren darstellt (siehe Abbildung 7).

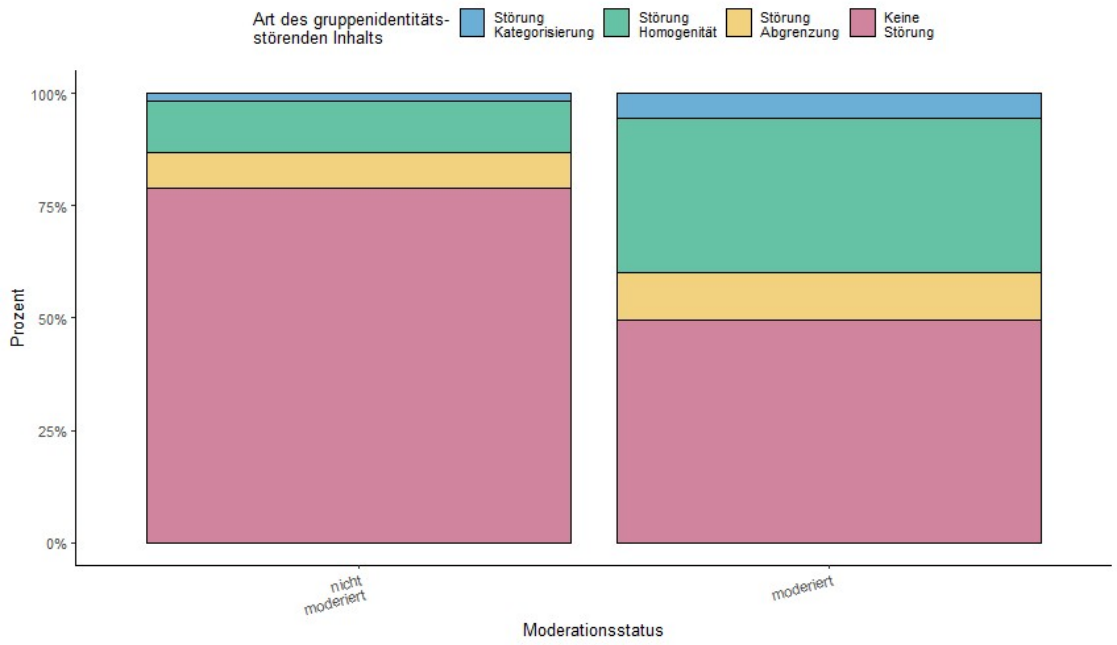

Abbildung 7: Moderationsstatus nach Art des gruppenidentitätsstörenden Inhalts in Prozent (N = 432)

Im Vergleich zur Gruppe der moderierten Kommentare machen die gruppenidentitätsstörenden Inhalte mit 21,11 % einen wesentlich kleineren Anteil der Kommentare innerhalb der verfügbaren Inhalte aus. Auch innerhalb der nicht-moderierten Kommentare stellen die

Störungen der Homogenität die größte Gruppe der Identitätsstörungen dar, diese ist aber mit einem Anteil von 11,42 % wesentlich kleiner als der Anteil der Homogenitätsstörungen an den moderierten Kommentaren. Auch die Gruppe der Störungen der Kategorisierung fällt mit 1,73 % Anteil der nicht-moderierten Kommentare deutlich kleiner aus (siehe Abbildung 7).

Die Analyse des Zusammenhangs zwischen den gruppenidentitätsstörenden Inhalten und dem Moderationsstatus der Kommentare ergab nach der angepassten Gewichtung einen Wert von $p < 0.01$ **** und ist somit höchst signifikant. Für die Berechnung der Effektstärke ergab sich Cramér's $V = 0.197$, wodurch von einem kleinen Effekt ausgegangen werden kann.

Um herauszufinden, für welche Kombinationen innerhalb der Kreuztabelle tatsächlich signifikante Unterschiede vorliegen, wurde ein Post-hoc-Test durchgeführt, dessen Ergebnisse in Tabelle 7 eingesehen werden können. Die hier angegebenen Werte sind unter Berücksichtigung der Gewichtung der geschichteten Stichprobe zu verstehen, weshalb die Häufigkeiten in Werten mit Dezimalstellen angegeben sind. Eine derartige Anpassung wurde auch für die standardisierten Residuen vorgenommen, um entsprechend der Gewichtung der geschichteten Stichprobe passende p-Werte zu erhalten.

Art der Störung	**Moderationsstatus**	**Beobachtete Häufigkeiten**	**Erwartete Häufigkeiten**	**Residuen**	**Signifikanz (p-Wert)**
Störung d. Kategorisierung	Nicht-moderiert	6.55	6.92	-0.89	0.37428604
Störung d. Homogenität	Nicht-moderiert	43.25	45.42	-2.17	0.03034699 *
Störung d. Abgrenzung	Nicht-moderiert	30.15	30.39	-0.29	0.77418560
Keine Störung	Nicht-moderiert	298.83	296.06	2.18	0.02930889 *
Störung d. Kategorisierung	Moderiert	0.54	0.18	0.89	0.37428604
Störung d. Homogenität	Moderiert	3.33	1.17	2.17	0.03034699 *
Störung d. Abgrenzung	Moderiert	1.02	0.78	0.29	0.77418560
Keine Störung	Moderiert	4.83	7.60	-2.18	0.02930889 *

Tabelle 7: Ergebnisse des Post-hoc-Tests der Interaktion zwischen gruppenidentitätsstörenden Inhalten und dem Moderationsstatus (N = 432)

Insgesamt zeigen sich für den Chi²-Test der Variablen Moderationsstatus und der gruppenidentitätsstörenden Inhalte nur vier Kombinationen mit signifikanten Ergebnissen. Da sich für die Effektstärke nach Cramér's V nur ein leichter Effekt zeigte und die hier als

signifikant zu identifizierenden Ergebnisse nicht sehr weit unter der Schwelle von $p = 0.05$ liegen, sind die vorliegenden Ergebnisse vorsichtig zu interpretieren. Etwas häufiger moderiert als erwartet werden solche Kommentare, die einer Störung der Homogenität aufweisen. Hierzu passend zeigt sich zudem, dass die homogenitätsstörenden Inhalte nicht nur in der Gruppe der moderierten Inhalte etwas häufiger repräsentiert sind als nach den erwarteten Häufigkeiten anzunehmen wäre, sie kommen zudem auch in der Gruppe der nicht-moderierten Inhalte seltener vor als erwartet. Dies bedeutet, dass ein leichter Effekt dafür angenommen werden kann, dass homogenitätsstörende Inhalte signifikant häufiger der Löschung durch Community-Moderatoren ausgesetzt sind als solche, die andere Arten von gruppenidentitätsstörenden Inhalten aufweisen oder keine Störung enthalten (siehe Tabelle 7).

Zudem kommen Inhalte, die keine Störung aufweisen, signifikant seltener als angenommen unter den moderierten Inhalten vor und zeigen sich häufiger als erwartet unter den nicht-moderierten Inhalten. Dies stellt insofern ein interessantes Ergebnis dar, da sich auch hier zeigt, dass Kommentare, die keine gruppenidentitätsstörenden Inhalte aufweisen, seltener als erwartet der Löschung durch die Community-Moderation ausgesetzt sind (siehe Tabelle 7).

Abschließend kann daher für die Zusammenhänge zwischen identitätsstörenden Inhalten, toxischem Verhalten und den Moderationsentscheidungen festgehalten werden, dass Kommentare, die potenziell die Identität der Community in Hinblick auf ihre Homogenität stören könnten, eher moderiert werden als solche, die toxisches Verhalten in Form von Unhöflichkeit oder Inzivilität aufweisen.

8. Diskussion

Im Rahmen dieser Arbeit wurden die Assoziationen verschiedener Akteure innerhalb von maskulinistischen Reddit-Communities untersucht. Hierbei wurde die Perspektive der Akteur-Netzwerk-Theorie eingenommen, um die Rolle von menschlichen und nicht-menschlichen Anteilen der Online-Communities gleichberechtigt in die Analyse mit einzubeziehen. Im Fokus der Analyse standen die community-spezifischen Regelsysteme, die Verwendung des Voting-Systems im Kontext der Content-Moderation sowie die Löschung von Inhalten durch Community-Moderatoren in Hinblick auf inhaltliche Charakteristika der entsprechenden Kommentare. Durch die Untersuchung dieser Affordanzen als Akteure innerhalb der Foren sollte zudem herausgefunden werden, welche Rolle diese für den Gruppenerhalt von maskulinistischen Reddit-Communities tragen.

(1) Die Analyse der Regelsysteme zeigte keinen signifikanten Zusammenhang zwischen den verwendeten Regeltypen der Communities und dem jeweiligen angestrebten Themenfokus innerhalb des maskulinistischen Spektrums. Da auf der Grundlage des Forschungsstandes zu Gruppen, die dem männlich-zentrierten Antifeminismus zugerechnet werden können, davon auszugehen ist, dass individualistisch orientierte Communities mit mehr Inzivilität und ausgeprägteren Formen von Sexismus und Misogynie in Verbindung stehen, da diese inhaltlich auf individuelle Erfahrungen mit spezifischen Frauen ausgerichtet sind (Han & Yin, 2023, S. 1935), bedarf dieser Befund weiterer Erklärung.

Da kein signifikanter Zusammenhang zwischen den Themen der Foren und deren verwendeter Regeltypen vorliegt, kann davon ausgegangen werden, dass sie sich diesbezüglich nicht von anderen Reddit-Communities unterschieden. Auch in Bezug auf die Häufigkeiten der verwendeten Regeltypen zeigen sich Verteilungen, die der Nutzung von Regeln in Communities außerhalb maskulinistischer oder antifeministischer Netzwerke stark ähneln. Fiesler et al. fanden in ihrer Untersuchung der Regeln innerhalb eines Samples von heterogenen Subreddits solche Regeltypen, die sich auf das erwünschte Verhalten innerhalb des Forums beziehen, und solche, in denen die gewünschten Inhalte vorgegeben wurden, mit insgesamt 74 % als am häufigsten verwendete Regeln vor (Fiesler et al., 2018, S. 77). Dieser Schwerpunkt zeigte sich auch für die im Rahmen dieser Arbeit untersuchten maskulinistischen Reddit-Communities. In ihrem besonderen Fokus auf das Einfordern einer zivilen Diskussionskultur und der Vermeidung von Hassrede können die Regeln der maskulinistischen Foren mit anderen politischen Diskussionsgruppen auf Reddit verglichen werden, für die ebendieser Fokus in ihren Regeln bereits aufgezeigt werden konnte. Demnach lässt sich, wie von Fiesler et al. gezeigt wurde, eine Ähnlichkeit in den verwendeten Regeltypen in Hinblick auf breit gefasste thematische Gruppen von Communities feststellen (Fiesler et al., 2018, S. 78). Durch eine feinere Untergliederung innerhalb der übergeordneten thematischen Gruppen, wie sie in dieser Arbeit für maskulinistische Communities vorgenommen wurde, lässt sich dieser Zusammenhang nicht mehr finden.

Die von einer Community verwendeten Regelsysteme erscheinen bei einer oberflächlichen Orientierung in den Foren als vermeintlich sehr relevanter Akteur innerhalb des Systems der Content-Moderation, da diese durch ihre prominente Platzierung innerhalb des Forums einer der ersten Aspekte ist, durch den Nutzer mit dem System der community-basierten Content-Moderation in Berührung kommen. Betrachtet man die Affordanz der Regelsysteme genauer, muss allerdings berücksichtigt werden, dass die Handlungsoptionen, die sich hieraus für Moderator*innen und Nutzer*innen ableiten lassen, nicht eindeutig sind, sondern Prozessen der

Interpretation unterliegen (Shaw, 2017, S. 594). Dies bedeutet, dass Moderierende in ihrer Formulierung der Regeln niemals abschließend berücksichtigen können, wie diese seitens der Nutzer*innen entschlüsselt werden. Zudem haben Nutzende immer auch die Möglichkeit, oppositionelle Standpunkte zu den Regeln einzunehmen und dementsprechende Nutzungsweisen zu zeigen. Da es zwischen Nutzer*in und Moderator*in kein hierarchisches Verhältnis gibt, sondern beide Gruppen auf die Arbeit und Partizipation der jeweils anderen angewiesen sind, kann der Umgang mit den Regeln der Community als ein Aushandlungsprozess verstanden werden, der in der Praxis der Moderation von Inhalten stark vom inhaltlichen Gehalt der Regeln abweichen kann (Squirrell, 2019, S. 1923). Dies bedeutet, dass, obwohl das Regelsystem als besonders sichtbarer Akteur der Content-Moderation verstanden werden kann, seine Rolle innerhalb des Moderationsprozesses entsprechend dieser Aushandlungsprozesse eingeordnet werden muss.

Durch diese Erkenntnis zeigt sich eine zentrale Schwäche der Akteur-Netzwerk-Theorie als theoretischem Rahmen für die Untersuchung der Verhältnisse zwischen verschiedenen Akteuren in einem Netzwerk. Da die ANT-Perspektive alle Akteure als gleichberechtigt in ihrem Einfluss auf die Entstehung sozialer Interaktion ansieht, ist es nicht möglich, besonders zentrale Beziehungen zwischen Akteuren herauszuarbeiten und andere Akteure als weniger relevant für spezifische Prozesse zu verstehen (Mladenović, 2025, S. 255). In Hinblick auf die Aushandlung der Regelsysteme in maskulinistischen Reddit-Communities scheint das dialektische Verhältnis zwischen dem Programm der Moderatoren und dem der verschiedenen Nutzer allerdings eine zentralere Position einzunehmen als die verschriftlichte Form der Regelsysteme. Demnach kann der Befund des nicht-signifikanten Verhältnisses zwischen den Regelsystemen und ihrer Ausrichtung auf den Gruppenerhalt mittels Akteur-Netzwerk-Theorie nicht zufriedenstellend eingeordnet werden. Für zukünftige Forschungsprojekte, die sich mit den Regeln innerhalb von Online-Communities befassen, scheint es daher besonders fruchtbar, die Affordanzen einer Plattform bezüglich ihrer Möglichkeiten zu Aushandlungsprozessen zwischen Moderator*innen und Nutzer*innen in Hinblick auf die Bildung impliziter Regeln und Normen zu untersuchen.

(2) Im Rahmen der Analyse der Voting-Affordanz als möglichem Akteur innerhalb des Systems der Content-Moderation konnte ein signifikanter Zusammenhang zwischen dem Moderationsstatus und dem Voting-Score der Kommentare festgestellt werden. Diese Assoziation könnte darauf hinweisen, dass das Voting-System als Akteur innerhalb der community-basierten Content-Moderation gesehen werden kann. Da durch die statistische

Analyse in dieser Arbeit allerdings keine Kausalität nachgewiesen werden konnte, ist der signifikante Zusammenhang ein Hinweis darauf, dass sich tiefergehende Untersuchungen bezüglich dieser Beziehung anbieten würden. Die Befunde stützen zudem die Erkenntnis, dass die Voting-Affordanz dazu beiträgt, toxisches Verhalten in antifeministischen Communities zu fördern.

Wie Massanari in ihrer Untersuchung der Förderung toxischer digitaler Kulturen zeigen konnte, tragen Reddits Affordanzen zur Sortierung des Contents innerhalb eines Forums mit der Standarteinstellung auf „*best*“ (hier werden die Postings mit dem höchsten Voting-Score bevorzugt angezeigt) dazu bei, dass Posts mit vielen Upvotes eine größere Sichtbarkeit erreichen. Da Nutzer*innen mit vielen sehr sichtbaren Postings Bekanntheit innerhalb einer Community erlangen können, fördern die Affordanz des Votings und die darauf basierenden Sortierung der Inhalte, dass es einen Anreiz für Nutzer*innen gibt, Inhalte zu posten, die solchen ähneln, die in der Vergangenheit bereits viele Upvotes durch andere Mitglieder erlangen konnten (Massanari, 2017, S. 337). Für maskulinistische Communities kann daher angenommen werden, dass dieses Zusammenspiel der technischen Rahmenbedingungen auf Reddit ein Posting-Verhalten begünstigt, welches die hegemoniale Perspektive innerhalb der Gruppe bestärkt, wohingegen Formen der Gegenrede benachteiligt werden, da sie durch die Interaktionen zwischen Downvotes und der automatisierten Hierarchisierung von Inhalten kaum Sichtbarkeit erlangen können.

Wird zudem noch der Zusammenhang zwischen der Löschung von Inhalten durch die Community-Moderation sowie dem automatisierten Reddit-Filter und deren geringem Voting-Score, der durch die Ergebnisse dieser Arbeit aufgezeigt werden konnte, berücksichtigt, so spielt nicht nur die stark verringerte Sichtbarkeit von Postings mit wenigen Upvotes eine zentrale Rolle, sondern auch, dass derartige Postings signifikant häufiger entfernt werden als Postings mit einem höheren Voting-Score. Folgt man Massanaris Erkenntnis, dass das Voting-System nicht nur einen Einfluss auf die Regulierung der Sichtbarkeit verschiedener Postings hat, sondern auch dazu führt, dass sehr ähnliche, dem Themenfokus der Community entsprechende Postings hierdurch gefördert werden, lässt sich für den Zusammenhang zwischen Moderationsentscheidung und Voting-Score festhalten, dass Voting als Affordanz im Sinne der Akteur-Netzwerk-Theorie als Teil des Systems der Content-Moderation gesehen werden kann und dass angenommen werden kann, dass Voting eine community-erhaltenden Funktion erfüllt, indem die Sichtbarkeit von unerwünschten Postings indirekt beeinflusst werden kann. Squirrell beschreibt in Hinblick auf das Verhältnis zwischen Moderator*innen und dem Einsatz der

Verwendung der Voting-Affordanz, dass diese nur sehr eingeschränkte Möglichkeiten der Einflussnahmen haben, wenn diese feststellen, dass Voting seitens der Nutzer*innen angewendet wird, um bestimmte Perspektiven oder User*innen-Gruppen aus der Community herauszudrängen (Squirrell, 2019, S. 1922). So gibt es bereits Reddit-Foren, die spezifische Anweisungen in Hinblick auf die Verwendung der Voting-Affordanz in ihre Regelsysteme aufgenommen haben, solche die ihre Foren durch die Anpassung der CSS überarbeiten, indem sie den Downvote-Button als unsichtbar einstellen, und solche, bei denen ein Pop-up-Fenster die Nutzer*innen über den gewünschten Umgang mit dem Voting-System informiert (Squirell, 2019, S. 1922). Auf diese Weise versuchen Moderator*innen Einfluss auf den Umgang mit dem Voting-System seitens der Nutzer*innen zu nehmen. Wird berücksichtigt, dass es kein eindeutig hierarchisches Verhältnis zwischen Nutzer*in und Moderator*in gibt, stellt sich in Hinblick auf den hier nachgewiesenen Zusammenhang zwischen Moderationsstatus und Voting-Score die Frage, ob das Voting-System eine Möglichkeit darstellt, seitens der Nutzer*innen Einfluss auf die Entscheidungen der Community-Moderator*innen zu nehmen und den Themenfokus sowie die Identität der Gruppe mitzuformen. Diese Überlegung liegt insofern nahe, als dass die Anzahl der Votes den Moderator*innen eine Orientierung in Hinblick auf die Bedürfnisse größerer Gruppen von Nutzenden geben kann. Weist ein Posting oder Kommentar besonders viele Downvotes auf, können die Moderator*innen davon ausgehen, dass dieser und ähnliche Inhalte entfernt werden können, ohne mit einem großen Risiko potenzieller Gegenrede seitens der Nutzenden rechnen zu müssen. Dieser Umstand wird zudem noch verstärkt, da die Sichtbarkeit des Inhaltes durch den geringen Voting-Score bereits durch die automatisierte Hierarchisierung von Inhalten innerhalb der Plattform verringert wurde, sodass User*innen, die möglicherweise eine befürwortende Einstellung bezüglich des negativ bewerteten Inhaltes aufweisen, nur noch sehr erschwert Zugang zu diesem Posting erhalten. Auf Grund der Erkenntnisse aus dem Forschungsstand und dem hier nachgewiesenen Zusammenhang zwischen Voting-Score und Moderationsstatus kann davon ausgegangen werden, dass die Voting-Affordanz als Teil des Systems der Content-Moderation verstanden werden kann und zudem eine relevante Rolle für den Erhalt der maskulinistischen Communities erfüllt.

(3) Innerhalb der maskulinistischen Community, für die der Zusammenhang zwischen Moderationsstatus, toxischem Verhalten und gruppenidentitätsstörenden Inhalten untersucht wurde, konnte kein Zusammenhang zwischen der Löschung von Inhalten und deren Gehalt toxischen Verhaltens nachgewiesen werden. Gleichzeitig gibt es eine Assoziation zwischen dem

Vorhandensein gruppenidentitätsstörender Inhalte in einem Kommentar und dessen Löschung. In Verbindung mit den Ergebnissen für die Analyse der Regelsysteme der Foren scheint diese Erkenntnis zunächst widersprüchlich, da ein auffallend großer Anteil der Regeln die Unerwünschtheit toxischen Verhaltens thematisiert. Diese Widersprüchlichkeit bestärkt allerdings die Feststellung, dass die Regelsysteme möglicherweise eine wesentlich weniger zentrale Rolle als Akteur der Content-Moderation einnehmen, als zunächst angenommen wurde, da die Moderationspraxis in *r/shortguys* stark von den Regeln der Community abzuweichen scheint.

Grundsätzlich weisen diese Ergebnisse darauf hin, dass die Moderatoren dieser Community keine Arbeit im normativen Sinne der Inhaltsmoderation erfüllen, die ihren Fokus auf die Verminderung beleidigender und inziviler Inhalte legt. Die Community-Moderatoren in *r/shortguys* tragen demnach nicht zu einem gesünderen Diskurs rund um Geschlechterverhältnisse und Gleichstellung bei, da toxische Inhalte in den Gruppen der nicht-moderierten und moderierten Posting einen ähnlichen Anteil ausmachen. Der Moderationsstatus eines Postings kann demnach nicht durch seinen Gehalt an toxischem Verhalten erklärt werden.

In der Analyse des Moderationsstatus nach gruppenidentitätsstörenden Inhalten zeigte sich nicht nur ein signifikanter Zusammenhang zwischen beiden Variablen, sondern auch, dass unter den potenziell identitätsstörenden Inhalten am häufigsten solche Kommentare der Entfernung von Moderatoren ausgesetzt waren, die der Ausprägung „Störung der Homogenität“ zugeordnet wurden. Kommentare wurden dieser Ausprägung nach kategorisiert, wenn sie einen alternativen Zugang zu Problemstellungen bezüglich Geschlechterungleichheiten einhielten oder eine Form der Gegenrede in Bezug auf die innerhalb des Forums dominierenden Position zu Geschlechterverhältnissen darstellten. Dieses Ergebnis könnte darauf hindeuten, dass Störungen der Identitätsdimensionen der Kategorisierung und der Abgrenzung zu Out-Groups als weniger bedrohlich für die Stabilität der Community gesehen werden, während die Übereinstimmung der Haltung zu Geschlechterverhältnissen innerhalb der Gruppe eine relevante Voraussetzung für das Fortbestehen des Kollektivs darstellt. Da das Forum *r/shortguys*, anhand dessen Forschungsfrage 3 untersucht wurde, eine sehr spezialisierte Community in Hinblick auf die thematische Ausrichtung rund um die Intersektion zwischen Geschlecht und Körpergröße darstellt, wäre es für zukünftige Forschung daher interessant zu überprüfen, ob Foren, die explizit für männliche Nutzer betrieben werden, gleichzeitig aber einen wesentlich breiteren Themenfokus aufweisen, einen ähnlichen Schwerpunkt auf die

Löschung homogenitätsstörender Inhalte legen oder ob diese möglicherweise stärker anhand der Dimensionen der Kategorisierung moderieren, da die Community weniger stark auf spezifischen Inhalten aufbaut, sondern sich darüber definiert, einen exklusiv männlichen Raum anzubieten.

Eine der wesentlichen Stärken der Akteur-Netzwerk-Theorie als zentrale Grundlage dieser Arbeit war es dabei, soziale Gruppen als konstruiert zu verstehen und die Arbeitsleistungen ihrer Akteure als bedeutende Voraussetzung für das Fortbestehen ebendieser zu sehen. Dies ermöglicht eine Auseinandersetzung mit Content-Moderation jenseits einer Perspektive, die den normativen Nutzen digitaler Inhaltsfilterung in den Mittelpunkt rückt, da dies weder in der klassischen Top-down-Moderation noch in der Community-Moderation als einzig zentrales Motiv der Moderationspraxis gesehen werden muss. Dies scheint für maskulinistische Reddit-Foren, deren Moderatoren selbst auch als Mitglieder und Nutzer auftreten, naheliegend. Die erfolgreiche Anwendung einer nicht-normativen theoretischen Perspektive wie der Akteur-Netzwerk-Theorie auf Systeme der Content-Moderation könnte allerdings auch einen Anstoß dazu geben, die netzwerk-erhaltende Rolle dieser in zukünftiger Forschung stärker zu berücksichtigen. Gleichzeitig darf die Relevanz der gruppenidentitätsstützenden Funktion der Moderation in der hier untersuchten Community nicht überschätzt werden, da insgesamt nur 1,60% der Kommentare der Löschung von Moderatoren ausgesetzt waren.

Um eine alternative Perspektive zu den normativen Standpunkten rund um die Funktionen von Content-Moderation auf populären Plattformen einzunehmen, zeigte sich auch der Fokus auf die Dynamik von Online-Communities als äußerst sinnvoll. Der hier erfolgte Nachweis eines Zusammenhangs zwischen der Entfernung von Postings und deren Aufweisen von gruppenidentitätsstörenden Inhalten, macht die Verwobenheit zwischen den sozialen Voraussetzungen digitaler Gruppen und den technischen Rahmenbedingung in Form von Plattform-Affordanzen sichtbar.

Gegenwärtige feministische Forschung zu sozialen Interaktionen im digitalen Raum verweist auf das Netz als vergeschlechtlichten Raum, in dem Geschlechterungleichheiten, Sexismus und Frauenfeindlichkeit fortbestehen. Wie viele weitere Untersuchungen der Affordanzen populärer Plattformen, zeigt auch die im Rahmen dieser Arbeit unternommene Analyse auf, dass die Hierarchisierung der Geschlechter im Netz nicht als digitales Äquivalent zu analoger Geschlechterungleichheit gesehen werden sollte, sondern durch die Verwobenheit technischer

Affordanzen und menschlicher Akteure in sozialen Interaktionen eine eigene Beschaffenheit aufweist. In der feministischen Auseinandersetzung mit digitalen Affordanzen populärer Plattformen zeigt sich zunehmend, dass diese oftmals in einem stützenden bis hin zu förderndem Verhältnis im Hinblick auf Sexismus und Misogynie stehen. Auch im Rahmen dieser Arbeit wurde eine derartige Tendenz sichtbar. Durch die Verknüpfung der These zur Fragmentierung der Öffentlichkeit in digitalen Räumen durch Online-Communities, die ihre Stabilität aus Homogenisierung durch eine geteilte Identität beziehen, mit einem Fokus auf der Verwendung von Plattform-Affordanzen durch Nutzer und Moderatoren konnte sichtbar gemacht werden, dass diese Handlungsoptionen auch dafür eingesetzt werden, Gruppenidentitäten (insbesondere durch die Aufrechterhaltung von Homogenität innerhalb der Community) durch Moderationsmaßnahmen zu schützen.

Aus feministischer Sicht sind dezentrale Moderationssysteme, wie die Auseinandersetzung mit community-basierter Content-Moderation auf Reddit zeigt, ambivalent zu bewerten. Einerseits können die bestehenden Affordanzen zur Moderation durch marginalisierte Gruppen genutzt werden, um eigene Gegenöffentlichkeiten zu schaffen und diese selbstbestimmt zu gestalten, ohne dass diese maßgeblich durch Formen der Top-down-Moderation durch die Plattformbetreibenden reguliert werden. Auf diese Weise können sichere Räume geschaffen werden, in denen es möglich ist, subversive Perspektiven zu entwickeln, durch die hegemoniale (Geschlechter-)Debatten kritisiert werden können. Marginalisierte Gruppen können auf diese Weise eigene Argumente formulieren, ohne sich den Bedingungen dominanter Öffentlichkeiten anpassen zu müssen (Fraser, 1990, S. 68). Gleichzeitig bieten Reddits Affordanzen zur dezentralen Moderation auch für solche Communities Möglichkeiten zur Entwicklung von Gegenöffentlichkeiten, die auf Grund ihres inhaltlichen Fokus auf die (Re-)Souveränisierung hegemonialer Männlichkeit und ihrer Ablehnung von geschlechtlicher Egalität, toxisches Verhalten in Form von Sexismus und Misogynie aufweisen. Auf diese Weise wird die Dezentralität der Moderation, die für die Entwicklung emanzipatorischer Argumente und Strategien förderlich sein kann, gleichzeitig ein stützender Akteur für die Kulturvierung antiegalitärer Einstellungen.

8.1 Limitationen

In Hinblick auf die im Rahmen des Forschungsprojekts entstandenen Ergebnisse müssen einige Einschränkungen berücksichtigt werden.

Limitationen des theoretischen Rahmens

Das Konzept des Akteur-Netzwerks sowie die Berücksichtigung nicht-menschlicher Entitäten in der Entstehung von Sozialität innerhalb des Netzwerks bieten eine sinnvolle Grundlage zur Erforschung von sozialen Phänomenen im digitalen Raum. Zentral hierbei ist vor allem, dass die ANT-Perspektive dabei helfen kann zu verhindern, dass soziale Phänomene im Netz auf die Rolle der Nutzer*innen oder die Relevanz der technischen Rahmenbedingungen reduziert werden. Die Akteur-Netzwerk-Theorie ermöglicht es, diese Aspekte gleichermaßen in die Analyse zu integrieren. Gleichzeitig sollten auch die Schwächen des Ansatzes berücksichtigt werden.

Da aus dieser theoretischen Perspektive immer schon alle Akteure eines Netzwerks an der Entstehung spezifischer Formen von Sozialität beteiligt sind, ist die ANT-Perspektive nicht dazu geeignet, nach spezifischen Akteuren zu fragen, von denen Handlungen ausgehen oder ausgelöst wurden. Die Theorie lenkt damit einen stärkeren Fokus auf die Beschreibung von Netzwerken, indem Entitäten als Akteure identifiziert werden, bietet aber keine geeignete Grundlage, um kausale Verhältnisse zwischen konkreten Akteuren zu untersuchen, da diese in der Akteur-Netzwerk-Theorie nicht vorgesehen sind.

In Hinblick auf die Frage der Machtverhältnisse und ethischer Verantwortung in der Analyse von Plattform-Affodanzen, muss zudem berücksichtigt werden, dass die Akteur-Netzwerk-Theorie davon ausgeht, dass sich die Verteilung von Macht in einem Netzwerk zirkulär verhält. Dieser Grundsatz ermöglicht die Analyse der Aneignung von Affordanzen durch Moderator*innen und anderen Nutzer*innen entgegen ihrer durch die Plattformbetreibenden intendierten Nutzungsweise. Hieraus ergibt sich allerdings auch die Einschränkung, dass Machtmissbrauch und Ausbeutungsverhältnisse durch die ANT-Perspektive ausgeblendet werden, da alle Akteure mit der gleichen Relevanz an der Entstehung von Sozialität beteiligt sind (Mladenović, 2025, S. 255-256). Dies verhindert, dass Machtungleichgewichte zwischen den Eigentümer*innen von Reddit und seinen Nutzer*innen in die Untersuchung mit einbezogen werden können. In Bezug auf die wissenschaftliche Auseinandersetzung mit community-basierter Content-Moderation schränkt dies die Möglichkeiten ein, das Verhältnis zwischen den ökonomischen Interessen der Eigentümer*innen und der Profitgenerierung durch die unbezahlte Arbeit der Community-Moderator*innen einer machtkritischen Analyse zu unterziehen.

Limitationen im sozialwissenschaftlichen Nutzen von Big Data

Große Datenmengen, die über digitale Plattformen gesammelt und unter anderem auch für sozialwissenschaftliche Forschungsanliegen verwendet werden können, sollten immer auch unter Berücksichtigung ihres Entstehungskontexts und Zwecks behandelt werden. Die Datenerhebung populärer Plattformen geschieht oftmals nicht nach wissenschaftlichen Gütekriterien. In den meisten Fällen liegt keine Transparenz seitens der Plattformbetreibenden in Bezug auf die Gewinnung ebensolcher Daten vor. Zudem werden oftmals Variablen in der Erhebung ausgespart, die für sozialwissenschaftliche Untersuchungen von Interesse sind, da sie für die wirtschaftlichen Interessen der Erhebenden nicht relevant sind (Salganik, 2018, S. 24).

Für diese Arbeit gilt diese Einschränkung insofern, als dass Reddit die Mitgliederzahlen von Foren nur in gerundeter Form öffentlich angibt. Da die Mitglieder nicht einzeln eingesehen werden können, musste für die Erhebung die auf Reddit verfügbare ungenaue Zahl für die Erhebung der Variable der Größe des Forums verwendet werden. Zudem lässt sich über Reddit nicht nachverfolgen, wann einzelne Moderator*innen eines Forums ihre Rolle als solche angetreten haben oder wie aktiv diese sind. So kann es sein, dass in der Erhebung Moderatoren berücksichtigt wurden, die ihre Rolle als solche nicht weiter ausüben, deren Accounts von den anderen Moderatoren aber nie entfernt worden sind.

Zudem ist es bei Untersuchungen von sozialen Phänomenen auf digitalen Plattformen nur sehr schwer möglich, einen zuverlässigen Überblick über die Grundgesamtheit zu erlangen, da ohne transparente Erhebungen seitens der Plattformbetreibenden niemals gesichert alle Communities zu einem spezifischen Thema gefunden werden können. Somit kann die Repräsentativität der Stichprobe im Rahmen von Social-Media-Forschung nicht gesichert angenommen werden.

Einschränkungen in der Erhebungsmethode

Grundsätzlich muss für die Untersuchung von Prozessen der Content-Moderation mittels quantitativer Inhaltsanalyse berücksichtigt werden, dass keine Aussagen über die Motive der Moderatoren bezüglich ihrer Entscheidungen zur Löschung von Inhalten gemacht werden können. Das Motiv des Gruppenerhalts kann durch die hier produzierten Erkenntnisse nur verstärkt vermutet werden. Zudem ergibt sich eine weitere erhebliche Einschränkung der Analyse der Community-Moderation, da dieses System sich im Gegensatz zur klassischen Top-down-Moderation hierdurch auszeichnet, dass es nicht nur auf der Entfernung von Postings und Kommentaren basiert. Community-Moderator*innen verfügen über weitaus mehr

Möglichkeiten, den Diskurs innerhalb des Forums zu beeinflussen, indem sie beispielsweise Verwarnungen an einzelne Nutzende aussprechen, deren Zugriff auf das Forum in Form eines time-outs für eine bestimmte Zeit blockieren oder diese nach mehrfachen Regelverstößen gänzlich aus dem Forum entfernen. All diese weiteren Affordanzen der Community-Moderation auf Reddit konnten durch die hier gewählte Erhebungsmethode nicht mit in die Untersuchung einbezogen werden, stellen allerdings einen Teil des Systems der Content-Moderation dar.

Zudem konnten im Rahmen der Erhebung der Kommentare aus einer ausgewählten maskulinistischen Reddit-Community durch die Methode zur Datengenerierung nicht alle relevanten Postings erhoben werden. Da zwischen den Downloads der Daten immer mehrere Stunden lagen, konnten solche gelöschten und entfernten Kommentare nicht regeneriert werden, die vor dem Download des Kommentar-Texts durch die Löschung nicht mehr über das Forum abgerufen werden konnten. Im Datensatz konnte für diese Postings daher nur ihr Status als gelöschter oder entfernter Post registriert werden, der Text hierzu fehlte allerdings. Dies betrifft insbesondere Kommentare, die sehr schnell entfernt wurden. Der größte Anteil von Texten, die nicht rekonstruiert werden konnten, entfällt auf solche Kommentare, die automatisiert durch Reddit entfernt wurden. Da Reddit nur wenige Regeln in Bezug auf verbotene Inhalte vorgibt, die sich hauptsächlich auf illegale Inhalte (insbesondere Gewalt, Copyright und Inhalte, die sexuellen Missbrauch zeigen) beziehen, kann angenommen werden, dass es sich bei den nicht erfassten Inhalten um derartige Postings handelt. Bei dem einzigen von Reddit entfernten Posting, dass rekonstruiert werden konnte, handelt es sich um die Suizidankündigung eines Nutzers. Da die durch die automatisierte Moderation entfernten Postings nicht rekonstruiert werden konnten, mussten sie aus der Untersuchung der Kommentarinhalte ausgeschlossen werden, obwohl die Affordanzen der Auto-Moderation die transformierende Rolle eines Akteurs im System der Content-Moderation aufweisen könnten.
Zudem muss bei der Bewertung der Ergebnisse berücksichtigt werden, dass für die statistische Analyse zu FF3 nur eine geschichtete Stichprobe verwendet wurde, da in einer Zufallsstichprobe ohne Schichtung zu wenig gelöschte und entfernte Kommentare enthalten gewesen wären.

Limitationen aus forschungspragmatischen Gründen

Auf Grund des hohen Aufwands der Datengewinnung bezüglich der gelöschten Postings musste die tiefergehende inhaltsanalytische Untersuchung der betreffenden Kommentare auf ein

spezifisches Reddit-Forum begrenzt werden. Dies schränkt die Aussagekraft der Ergebnisse erheblich ein. Da die Analyse signifikante Ergebnisse zeigt, weist dies darauf hin, dass es für zukünftige Untersuchungen lohnend sein kann, die Datenerhebung für weitere Communities fortzuführen, um die hier gewonnenen Erkenntnisse über das community-basierte Moderationssystem zu vertiefen und gegebenenfalls zu korrigieren. Zudem darf die Analyse in Hinblick auf die möglichen Akteure, die mit dem System der Content-Moderation assoziiert sind, nicht als erschöpfend verstanden werden. Aus pragmatischen Gründen wurde ein Fokus auf die Affordanzen der community-spezifischen Regelsysteme, das Voting und die Möglichkeit der Umsetzung von Moderationsentscheidungen durch Löschung von Inhalten gesetzt. Es sind allerdings viele weitere Assoziationen zwischen Reddit-Affordanzen als Akteuren denkbar, sodass zukünftige Analysen die Affordanzen zur Hierarchisierung von Inhalten anhand deren Aktualität, die Affordanzen zur Festlegung und Verwendung community-eigener *flairs* (farblich hervorgehobene Markierungen, die genutzt werden können, um Inhalte von Postings oder Eigenschaften von Nutzer*innen hervorzuheben) sowie viele weitere Handlungsoptionen fokussieren könnten, um sie in ihrer Beziehung zum System der Community-Moderation zu verstehen.

8.2 Ausblick

In Anbetracht der im Rahmen des vorliegenden Forschungsprojekts produzierten Erkenntnisse wäre es für weiterführende Forschung sicherlich gewinnbringend, verschiedene systematische Vergleiche anzustellen, um verallgemeinerbare Aussagen über die Verwendung von Plattform-Affordanzen zum Zweck der community-basierten Content-Moderation über maskulinistische Subreddits hinaus treffen zu können. Die Analyse weiterer politischer Communities – insbesondere solcher mit emanzipatorischen Anliegen – könnte klären, inwiefern der in dieser Arbeit festgestellte geringe Fokus auf toxisches Verhalten bei gleichzeitiger Priorisierung von Identitätskohäsion in den Moderationsentscheidungen spezifisch für maskulinistische Räume gilt oder ob sich ähnliche Moderationslogiken auch für Communities beobachten lassen, die sich in ihrem Selbstverständnis auf eine Sensibilität für Diskriminierung und Marginalisierung berufen. Hierdurch ließe sich untersuchen, inwiefern Community-Moderation eine grundsätzliche Tendenz zu identitätsschützenden Maßnahmen aufweist, die als unabhängig von der politischen Haltung des Kollektivs verstanden werden können.

Zudem erscheint es als sinnvoll, die Forschung zu dezentraler Moderation auf weitere populäre Plattformen auszuweiten. Hierfür würden sich insbesondere Discord und Telegram als

Vergleichsfälle anbieten, da beide Anbieter Community-Moderation ermöglichen, gleichzeitig aber über eine eigene Plattform-Architektur verfügen, die sich deutlich von Reddit unterscheidet. Eine vergleichende Analyse würde es ermöglichen festzustellen, inwiefern die Erkenntnisse bezüglich des Verhältnisses zwischen Plattform-Affordanzen, dezentraler Moderation und dem Schutz der kollektiven Identität auf der spezifischen Struktur von Reddit basieren oder aber auch für andere Plattformen identifiziert werden können.

Die Akteur-Netzwerk-Theorie erweist sich als ein sehr produktives theoretisches Framework zur Analyse von Plattform-Moderation. Vor allem ihr konsequenter Einbezug nicht-menschlicher Akteure in soziale Prozesse ermöglicht es, die Affordanzen der Content-Moderation nicht nur nebensächlich in die Analyse miteinzubeziehen, sondern als einen wirkmächtigen Bestandteil der hier untersuchten Communities aufzugreifen. Die Zentralität aller Akteure kann hierbei die Gefahr mindern, in einen sozialen oder technologischen Determinismus in der Beschäftigung mit digitalen Plattformen abzugleiten. Gleichzeitig ist es innerhalb der ANT nicht vorgesehen, Akteure nach ihrer Relevanz innerhalb des Netzwerks zu unterscheiden. Die Ergebnisse bezüglich der Rolle der Regeln für die Content-Moderation auf Reddit zeigen allerdings auf, dass eine solche Unterscheidung einen tieferen Einblick in das Netzwerk ermöglichen könnte. Darüber hinaus erschwert der eher deskriptive und explorative Charakter der Akteur-Netzwerk-Theorie die kritische Analyse möglicher Machtverhältnisse zwischen verschiedenen Akteuren. Diese Einschränkung sollte bei zukünftigen Untersuchungen unbedingt berücksichtigt werden, damit machtgeladene Aushandlungsprozesse zwischen Nutzer*innen und Moderator*innen in Hinblick auf den community-erhaltenden Umgang mit den Plattform-Affordanzen innerhalb verschiedener digitaler Gruppen nicht ausgeblendet werden.

Der Fokus auf die Community-Dimension hat sich in dieser Arbeit als äußerst fruchtbar für die Analyse von Moderationsprozessen gezeigt. Da im Rahmen des vorliegenden Forschungsprojekts die Verwobenheit von technischen und sozialen Rahmenbedingungen auf digitalen Plattformen gezeigt werden konnte, erweist sich der starke Fokus auf analoge Communities in der bisherigen Theoriebildung als Forschungslücke, die bearbeitet werden sollte. Das alleinige Zurückgreifen auf theoretische Konzepte zu Offline-Communities könnte langfristig dazu führen, dass digitale Gruppen als analog zu ebendiesen verstanden werden, wobei ihre spezifische Beschaffenheit und insbesondere ihre enge Verknüpfung mit technischen Akteuren ausgeblendet wird.

Abschließend kann konstatiert werden, dass eine intensivere kommunikationswissenschaftliche Auseinandersetzung mit community-basierten Moderationssystemen sehr zu befürworten ist, obwohl dieses Modell nur von wenigen populären Plattformen angeboten wird, da Plattformen keine undynamischen Entitäten darstellen, sondern in einem ständigen Prozess des Werdens begriffen sind und Betreibende und Nutzende den digitalen Raum unaufhörlich neu formen (Van Dijck, 2013, S. 6). Wird dieser Umstand berücksichtigt, ist eine zukünftige Tendenz in der Inhaltsregulierung hin zu dezentralisierter Moderation nicht auszuschließen, insbesondere da diese Form der Content-Filterung wesentlich weniger Personalkosten für die Plattformbetreibenden verursacht als professionelle Formen der Top-down-Moderation. Aus feministischer Perspektive gilt es daher eine gut ausgebauchte Erkenntnisgrundlage in Hinblick auf den ambivalenten Charakter von community-basierter Moderation zu erarbeiten, um mögliche Umgangsweisen mit dieser Zweischneidigkeit entwickeln zu können, denn „*the same structures that allow minority groups to create supportive spaces allow e.g. white nationalist groups to create safe spaces of their own"* (Seering, 2020, S. 8).

Literaturverzeichnis

Aigner, I., Lenz, I. (2023). Antifeminismus und Antigenderismus in medialen und digitalen Öffentlichkeiten. In: Dorer, J., Geiger, B., Hipfl, B., Ratković, V. (Hg.): *Handbuch Medien und Geschlecht*. Springer VS. 721-730

Are, C. (2025). Flagging as a silencing tool: Exploring the relationship between de-platforming of sex and online abuse on Instagram and TikTok. In: *new media & society*. 27(6). 3577-3595. https://doi.org/10.1177/14614448241228544

Beauvais, T. (2023). Hybrid representative sampling of social media. In: *Bulletin de M´ethodologie Sociologique*. 160. 57-70. https://doi.org/10.1177/07591063231196162

Bivens, R. (2017). The gender binary will not be deprogrammed: Ten years of coding gender on Facebook. In: *new media & society*. 19(6). 880-898. https://doi.org/10.1177/1461444815621527

Blais, M., Dupuis-Déri, F. (2012). Masculinism and the Antifeminist Countermovement. In: *Social Movement Studies*. 11(1). 21-39. http://doi.org/10.1080/14742837.2012.640532

Bucher, T. (2018). *If...Then. Algorithmic Power and Politics*. Oxford University Press. https://doi.org/10.1093/oso/9780190493028.001.0001

Cameron-McKee, S. (2020). The affordances of Reddit and framing of extremism on r/the_donald. In*: Journal of Applied Linguistics and Professional Practice*. 17(3). 231-251. https://doi.org/10.1558/jalpp.21907

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., Starnini, M. (2021). The echo chamber effect on social media. In: *Proceedings of the National Academy of Sciences*. 118(9). 1-8. https://doi.org/10.1073/pnas.2023301118

Copland, S. (2020). Reddit quarantined: Can changing platform affordances reduce hateful material online?. In: *Internet Policy Review*. 9 (4). 1-26. https://doi.org/10.14763/2020.4.1516%0A

Crawford, K., Gillespie, T. (2016). What is a flag for? Social media reporting tools and the vocabulary of complaint. In: *new media & society*. 18(3). 410–428. https://doi.org/10.1177/1461444814543163

Davis, J. L., Chouinard, J. B. (2017). Theorizing affordances: from request to refuse. In: *Bulletin of Science, Technology & Society*. 36(2). 1-8. https://doi.org/10.1177/0270467617714944

Deligianni, A., & Horne, Z. (2023*). Analysing hate speech towards women on Reddit*. https://doi.org/10.31234/osf.io/fsrq7

Duguay, S. (2016). Lesbian, Gay, Bisexual, Trans, and Queer Visibility Through Selfies: Comparing Platform Mediators Across Ruby Rose's Instagram and Vine Presence. In: *Social Media + Society*. 2(2). 1-12. https://doi.org/10.1177/2056305116641975

Engelmann, I., Marzinkowski, H., Langmann, K. (2024). Salient deliberative norm types in comment sections on news sites. In: *new media & society*. 26(2). 921-940. https://doi.org/10.1177/14614448211068104

Evans, R. (2018, 11. Oktober). From Memes to Infowars: How 75 Fascist Activists Were “Red-Pilled”. *Bellingcat*. https://www.bellingcat.com/news/americas/2018/10/11/memes-infowars-75-fascist-activists-red-pilled/

Farrell, T., Fernandez, M., Novotny, J., Alani, H. (2019). Exploring Misogyny across the Manosphere in Reddit. In: *Proceedings of the 10th ACM Conference on Web Science*. 87-96. https://doi.org/10.1145/3292522.3326045

Fiesler, C., Jiang, J., McCann, J., Frye, K., Brubaker, J. (2018). Reddit Rules! Characterizing an Ecosystem of Governance. In: *Proceedings of the International AAAI Conference on Web and Social Media*. 12(1). 72-81. https://doi.org/10.1609/icwsm.v12i1.15033

Forster, E. (2006). Männliche Resouveränisierungen. In: *Feministische Studien*. 24 (2). 193-207

Fox, J., Cruz, C., Lee, J. Y. (2015). Perpetuating online sexism offline: Anonymity, interactivity, and the effects of sexist hashtags on social media. In: *Computers in Human Behavior*. 52. 436-442. http://doi.org/10.1016/j.chb.2015.06.024

Fraser, N. (1990). Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. In: *Social Text*. 25/26. 56-80

Fraser, N. (2013, Dezember). Neoliberalismus und Feminismus: Eine gefährliche Liaison. *Blätter für deutsche und internationale Politik*. https://www.blaetter.de/ausgabe/2013/dezember/neoliberalismus-und-feminismus-eine-gefaehrliche-liaison

Ganz, K., Meßmer, A. (2015). Anti-Genderismus im Internet. Digitale Öffentlichkeiten als Labor eines neuen Kulturkampfes. In: Hark, S., Villa, P. (Hg.): *Anti-Genderismus. Sexualität und Ge¬schlecht als Schauplätze aktueller politischer Auseinandersetzungen*. Transcript. 59-78

Gergorić, M., Ghosh, K., Kuteleva, A. (2025). The Global Rise of Anti-Gender Movements and Feminist Resistance Strategies: A Critical Analysis. In: *POLITIKON: The IAPSS Journal of Political Science*. 59. 2-9. https://doi.org/10.22151/politikon.12025.0

Gerrard, Y. (2020). Social media content moderation: six opportunities for feminist intervention. In: *Feminist Media Studies*. 20(5). 748-751. https://doi.org/10.1080/14680777.2020.1783807

Gerrard, Y., Thornham, H. (2020). Content moderation: Social media’s sexist assemblages. In: *new media & society*. 22 (7). 1266-1286. https://doi.org/10.1177/1461444820912540

Gillespie, T. (2018). *Custodians of the Internet. Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press

Ging, D., Siapera, E. (2018). Special issue on online misogyny. In: *Feminist Media Studies*. 18(4). 515-524. https://doi.org/10.1080/14680777.2018.1447345

Ging, D. (2019). Alphas, Betas, and Incels: Theorizing the Masculinities of the Manosphere. In: *Men and Masculinities.* 22(4). 638-657. https://doi.org/10.1177/1097184X17706401

Goanta, C., Ortolani, P. (2021). *Unpacking Content Moderation: The Rise of Social Media Platforms as Online Civil Courts*. http://dx.doi.org/10.2139/ssrn.3969360

Habermas, J. (2022). *Ein neuer Strukturwandel der Öffentlichkeit und die deliberative Politik*. Suhrkamp

Han, X., Yin, C. (2023). Mapping the manosphere. Categorization of reactionary masculinity discourses in digital enivronment. In: *Feminist Media Studies*. 23 (5). 1923-1940. https://doi.org/10.1080/14680777.2021.1998185

Haraway, D. (1995). Ein Manifest für Cyborgs. Feminismus im Streit mit den Technowissenschaften. In: Hammer, C., Stieß, I. (Hg.): *Die Neuerfindung der Natur. Primaten, Cyborgs und Frauen*. Campus Verlag. 33-72

Harvey, A. (2020). *Feminist Media Studies*. Polity Press

Hintz, E. A., Betts, T. (2022). Reddit in communication research: current status, future directions and best practices. In: *Annals of the International Communication Association*. 46 (2). 116-133. https://doi.org/10.1080/23808985.2022.2064325

Holm, M. (2023). Beyond Antifeminist Discourses: Analyzing How Material and Social Factors Shape Online Resistance to Feminist Politics. In: *Social Politics*. 30(2). 422-443. https://doi.org/10.1093/sp/jxac022

Jhaver, S., Bruckman, A., Gilbert, E. (2019). Does Transparency in Moderation Really Matter?: User Behavoir After Content Removal Explanations on Reddit. In: *Proceedings of the ACM on Human-Computer Interaction.* 3. 1-27. https://doi.org/10.1145/3359252

Kaiser, S. (2021). *Politische Männlichkeit. Wie Incels, Fundamentalisten und Autoritäre für das Patriarchat mobilmachen*. Suhrkamp

Kracher, V. (2020). *Incels. Geschichte, Sprache und Ideologie eines Online-Kults*. Ventil Verlag

Latour, B. (1990). Technology is society made durable. In: *The Sociological Review*. 38 (1). 103-131. https://doi.org/10.1111/j.1467-954X.1990.tb03350.x

Latour, B. (2015). *Eine neue Soziologie für eine neue Gesellschaft. Einführung in die Akteur-Netzwerk-Theorie*. Suhrkamp

Leidig, E. (2023). *The Women of the Far Right: Social Media Influencers and Online Radicalization*. Columbia University Press

Luckman, S. (1999). (En)Gendering the digital body: Feminism and the internet. In: *Hecate*. 25 (2). 36-47

Marwick, A. E., Caplan, R. (2018). Drinking male tears: language, the manosphere, and networked harassment. In: *Feminist Media Studies*. 18(4). 543-559. https://doi.org/10.1080/14680777.2018.1450568

Massanari, A. (2017). #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. In: *new media & society*. 19 (3). 329–346. https://doi.org/10.1177/1461444815608807

McPherson, M., Smith-Lovin, L., Cook, J. M. (2001). Birds of a feather: Homophily in social networks. In: *Annual Review of Sociology*. 27. 415-444. https://doi.org/10.1146/annurev.soc.27.1.415

Mladenović, I. (2025). Beyond the Network. A Critical Inquiry into the Limits of the Actor-Network Theory. In: *Bulletin de l'Institut etnographique*. 73(1). 247-268. https://doi.org/10.2298/GEI2501247M

Nagle, A. (2018). *Die digitale Gegenrevolution Online-Kulturkämpfe der Neuen Rechten von 4chan und Tumblr bis zur Alt-Right und Trump*. Transcript

Nakamura, L. (2002). *Cybertypes. Race, Ethnicity, and Identity on the Internet*. Routledge

Nakamura, L. (2013, 10. Dezember). Glitch Racism: Networks as Actors within Vernacular Internet Theory. *Culture Digitally*. https://culturedigitally.org/2013/12/glitch-racism-networks-as-actors-within-vernacular-internet-theory/

Ortiz, S. M. (2024). "If Something Ever Happened, I'd Have No One to Tell:" how online sexism perpetuates young women's silence. In: *Feminist Media Studies*. 24(1). 119-134. https://doi.org/10.1080/14680777.2023.2185565

Papacharissi, Z. (2002). The virtual sphere. The internet as a public sphere. In: *new media & society.* 4(1). 9-27. https://doi.org/10.1177/14614440222226244

Papacharissi, Z. (2004). Democracy online: civility, politeness, and the democratic potential of online political discussion groups. In: *new media & society*. 6(2). 259-283. https://doi.org/10.1177/1461444804041444

Petö, A. (2015). Epilogue: „Anti-Gender" Mobilisational Discourse of Conservative and Far Right Parties as a challenge for Progressive Politics. In: Kováts, E., Põim, M. (Hg.): *Gender as symbolic glue*. Friedrich Ebert Stiftung. 126-131

Pilkington, H. (2025). Why study Non-Radicalisation? In: *Crest Security Review*. Herbst 2025. 27-29

Preece, J. (2000). *Online Communities. Designing Usability, Supporting Sociability*. Wiley

Reich, S., Bachl, M. (2025). When Sexism Becomes the Norm: The Effect of Sexism on Women's Participation in Political Online Discussions. In: *Communication Research*. 1-13. https://doi.org/10.1177/00936502251343988

Ren, Y., Harper, M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., Kraut, R. E. (2012). Building Member Attachment in Online Communities: Applying Theories of Group Identity and Interpersonal Bonds. In: *MIS Quarterly*. 36(3). 841-864.

Rivera, I. (2022, 3. März). *RedditExtractor. A minimalistic R wrapper for the Reddit API*. https://github.com/ivan-rivera/RedditExtractor

Roose, K. (2024, 21. März). Reddit's I.P.O. Is a Content Moderation Success Story. *New York Times*. https://www.nytimes.com/2024/03/21/technology/reddit-ipo-public-content-moderation.html

r/shortguys (2013). *Short Guys. r/ShortGuys is a space for short men to discuss height, heightism, and how it affects us*. https://www.reddit.com/r/shortguys/

r/shortguys (2025). *R/Shortguys Lorebook (Vol 1)*. https://www.reddit.com/r/shortguys/comments/1olonsw/rshortguys_lorebook_vol_1/

Salganik, M. J. (2018). *Bit by bit: social research in the digital age.* Princeton University Press

Sanders, R., Jenkins, L. D. (2022). Special issue introduction: Contemporary international anti-feminism. In: *Global Constitutionalism*. 11(3). 369-378. https://doi.org/10.1017/S2045381722000144

Sauer, B. (2021). Maskulinismus der Mitte? Zum Erfolg autoritär-rechtspopulistische Mobilisierung. In: Verwiebe, R., Wiesböck, L. (Hg.): *Mittelschicht unter Druck*. Springer VS. 59-78

Schmincke, I. (2018). Frauenfeindlich, Sexistisch, Antifeministsch? Begriffe und Phänomene bis zum aktuellen Antigenderismus. In: *Aus Politik und Zeitgeschichte*. 68. 28-33

Seering, J. (2020). Reconsidering Self-Moderation: The Role of Research in Supporting Community-Based Models for Online Content Moderation. In: *Proceedings of the ACM on Human-Computer Interaction*. 4. 1-28. https://doi.org/10.1145/3415178

Shaw, A. (2014). The Internet Is Full of Jerks, Because the World Is Full of Jerks: What Feminist Theory Teaches Us About the Internet. In: *Communication and Critical/Cultural Studies*. 11(3). 273-277. https://doi.org/10.1080/14791420.2014.926245

Shaw, A. (2017). Encoding and decoding affordances: Stuart Hall and interactive media technologies. In: *Media, Culture & Society*. 39(4). 592-602. https://doi.org/10.1177/0163443717692741

Solska, D., Sawicki, J. (2024). Decoding gender bias through a textual exploration of Reddit /r/MensRights community. In: *Beyond Philology*. 21 (1). 167-202. https://doi.org/10.26881/bp.2024.1.06

Squirrell, T. (2019): Platform dialectics: The relationships between volunteer moderators and end users on reddit. In: *new media & society*. 21 (9). 1910-1927. https://doi.org/10.1177/1461444819834317

Träbert, A. (2017). At the Mercy of Femocracy? Networks and Ideological Links Between Far-Right Movements and the Antifeminist Men's Rights Movement. In: Köttig, M., Bitzan, R., Petö, A. (Hg.): *Gender and Far Right Politics in Europe.* Springer VS. 273-288. https://doi.org/10.1007/978-3-319-43533-6_18

Turkle, S. (1998). *Leben im Netz: Identität in Zeiten des Internet*. Rowohlt

Van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media*. Oxford University Press

Van Zoonen, L. (2001). Feminist Internet Studies. In: *Feminist Media Studies*. 1(1). 67-72. https://doi.org/10.1080/14680770120042864

Wilhelm, A. G. (1998). Virtual sounding boards: How deliberative is online political discussion?. In: *Information, Communication & Society*. 1(3). 313-338. https://doi.org/10.1080/13691189809358972

Wimbauer, C., Motakef, M., Teschlade, J. (2015). Prekäre Selbstverständlichkeiten. Neun prekarisierungstheoretische Thesen zu Diskursen gegen Gleichstellungspolitik und Geschlechterforschung. In: Hark, S., Villa, P. (Hg.): *Anti-Genderismus. Sexualität und Geschlecht als Schauplätze aktueller politischer Auseinandersetzungen.* Transcript. 41-57

Anhang

Abstract (Deutsch)

Im Rahmen dieses Forschungsprojekts wurden die Assoziationen zwischen verschiedenen Affordanzen der community-basierten Content-Moderation auf der populären Plattform Reddit untersucht. Hierbei sollte herausgefunden werden, inwiefern die Affordanzen des Votings, der Festlegung community-spezifischer Regeln sowie die Anwendung von Löschungen als Moderationsmaßnahme mit dem System der Content-Moderation verbunden sind und in welchem Verhältnis diese zur Aufrechterhaltung antifeministischer (spez. maskulinistischer) Online-Communities auf der Plattform Reddit stehen. Als theoretische Grundlage wurde die Perspektive der Akteur-Netzwerk-Theorie eingenommen, um die Relevanz von menschlichen und nicht-menschlichen Akteuren gleichberechtigt in die Analyse mit einzubeziehen. Bezüglich der thematischen Ausrichtung der Communities und den von ihnen verwendeten Regelsystemen konnte kein signifikanter Zusammenhang entdeckt werden. In der Untersuchung der Voting-Affordanz innerhalb eines ausgewählten maskulinistischen Subreddits konnte ein Zusammenhang zwischen der Löschung von Postings durch den automatisierten Reddit-Filter oder der Community-Moderation und dem Voting-Score der Kommentare nachgewiesen werden. Ein leichter signifikanter Effekt zeigte sich außerdem zwischen dem Vorhandensein von gruppenidentitätsstörenden Inhalten in Kommentaren und deren Moderationsstatus. Für die Variablen des toxischen Verhaltens in Kommentaren und deren Löschung durch die Community-Moderation konnte kein signifikanter Zusammenhang gefunden werden. Die Ergebnisse verweisen auf eine erhaltende Rolle der community-basierten Moderation in Hinblick auf die Identität der Gruppe. Diese übernimmt hierbei demnach weniger die Funktion von Content-Moderation im normativen Sinne. Insgesamt konnte im Rahmen dieser Arbeit aufgezeigt werden, dass Reddits Affordanzen der Content-Moderation genutzt werden können, um den Erhalt maskulinistischer Communities zu stützen.

Abstract (Englisch)

This research project is focused on analysing the association between different affordances of Reddit's community-based mode of content moderation. The research questions were aimed at discovering whether certain affordances are used by moderators and users in order to regulate content within masculinist online communities on Reddit. The actor-network theory was used as theoretical framework in order to not only focus the analysis on human actors but also consider technical affordances as part of forming digital sociality and agency. The results of this project show no significant correlation between the community rules and specific community topics. There are however significant associations between the voting system that is used on Reddit and the moderation status of the analysed user-generated content. Comments that are removed by the community moderators or the automatic content filter tend to have significantly more downvotes than comments with a different availability status. There is also a small significant effect between content that contains language that could put the collective identity of the community at risk and the moderation status of those comments. Comments that are endangering for the group identity are removed significantly more often than contents that do not contain such language. There was no significant relation between comments containing toxic behavior and their moderation status. The results show that Reddit's community moderation in the studied masculinist online community has a tendency to work as a stabiliser for the identity of the group rather than fulfil the role of a content moderation system from a normative perspective, which is more focused on reducing toxicity and harm in digital spaces. Therefore Reddit's affordances of content moderation can be used to maintain masculinist online communities

Suchstring zur Identifizierung von maskulinistischen Reddit-Foren

Der folgende Suchstring wurde verwendet, um über die Suchfunktion der Plattform Reddit verschiedene maskulinistische Communities zu finden. Hierdurch sollte eine möglichst genaue Annäherung an die Grundgesamtheit erreicht werden.

men OR *man* OR *guy* OR *women* OR *woman* OR *feminis* OR *misandry* OR *masculin* OR *feminin* OR *female* OR *male* OR *incel* OR *pill* OR *gender* OR *traditional* OR *divorce* OR *custody* OR *gyno* OR *andro* OR *pick up* OR *alpha* OR *beta* OR *cuck* OR *foid*

Codebuch

Variable: **Thema des maskulinistischen Forums**

CODE	Ausprägungen	Indikatoren
1	**Antifeministische Gegenbewegung, Organisiert** Diskurs orientiert sich an antifeministischen Bewegungen oder Assoziationen	○ Forum ist nach einer antifeministischen Organisation, Bewegung oder Assoziation benannt ○ Forum nutzt Veröffentlichungen der antifeministischen Organisation, Bewegung oder Assoziation in der Erklärung ihres Selbstverständnisses ○ Forum bezieht sich auf die Geschichte und Entwicklung der Organisation, Bewegung oder Assoziation
2	**Antifeministische Gegenbewegung, Personenzentriert** Diskurs orientiert sich an den maskulinistischen Positionen einer Person des öffentlichen Lebens	○ Forum ist nach einer Person des öffentlichen Lebens benannt, die für maskulinistische Politiken steht ○ Forum bezieht sich in der Erklärung ihres Selbstverständnisses auf Veröffentlichungen der Person des öffentlichen Lebens ○ Postings des Forums beziehen sich auf öffentliche Auftritte oder Veröffentlichungen der Person des öffentlichen Lebens
3	**Antifeministische Gegenbewegung, Unabhängig** Diskurs entspricht der antifeministischen Gegenbewegung, orientiert sich aber weder an einer Person des öffentlichen Lebens noch an einer antifeministischen Organisation oder Assoziation	○ Forum thematisiert gesellschaftliche Dimension des Geschlechterverhältnisses, allerdings ohne sich auf eine konkrete Bewegung, Organisation oder Assoziation oder eine Person des öffentlichen Lebens zu beziehen
4	**Individualistisch, maskulinistische Diskurse, Philosophisch**	○ Forum ruft Männer zur Distanzierung von Frauen auf ○ Mitglieder werden dazu animiert, negative Erfahrungen mit Frauen zu teilen ○ Forum bewirbt einen spezifischen Lebensstil
5	**Individualistisch, maskulinistische Diskurse, Pragmatisch**	○ Forum thematisiert Strategien im Umgang mit Frauen ○ In Hinblick auf das Geschlechterverhältnis stehen sexuelle Beziehungen im Mittelpunkt der Debatte ○ Forum thematisiert Männlichkeit in Hinblick auf Selbstoptimierung und Disziplin
6	**Gemischte Foren**	○ Forum besteht zu ähnlichen Anteilen sowohl aus individualistischen Debatten als auch aus Debatten rund um gesellschaftspolitische Themen

Variable: **Regeltypen**

CODE	Ausprägung	Indikatoren
NA	Keine Zuordnung möglich	Regel enthält keine der hier angeführten Indikatoren
1	Qualität der Inhalte	Low Quality, Low Value (Falls dies in Verbindung mit absichtsvollem Trolling thematisiert wird, ist CODE 4 zu vergeben) Verwendung von AI Spam Reposting, Crossposting Angabe von Quellen Werbung, Self-Promo, Crowdfunding Fake News Rage-Bait
2	Format der Inhalte	Vorschriften zu Textformat, Bildformat, Videoformat Vorschriften zum Titel von Postings Labeln von Inhalten (z.B. NSFW, Satire) Verwendung von Community-Flairs
3	Orientierung an Reddit	Verweis auf die Reddiquette bzw. Reddits Content-Policies Illegale Inhalte (Copyright, Gewalt, Kinderpornographie …)
4	Toxisches Verhalten, allgemein	Trolling Belästigung Beleidigung Doxxing, Angabe von persönlichen Informationen Dritter Brigading Gatekeeping
5	Toxisches Verhalten gegenüber eine spezifische Personengruppe	Hate Speech (Gruppenbezogene Menschenfeindlichkeit, Sexismus, Rassismus, Homophobie, Antisemitismus …) Generalisierungen in Bezug auf eine Personengruppe
6	Thema das Subreddits	Debatten, kontroverse Inhalte & politische Inhalte Off-Topic Postings Unzulässige Ratschläge Rating-Posts (z.B. „Rate my …“) Meta-Debatten Posten von Umfragen oder Teilnahmeaufrufen zu wissenschaftlichen Studien
7	Abgrenzung nach Außen	Gegenrede, ablehnende Haltung gegenüber dem Thema des Forums Abgrenzung zu anderen antifeministischen Gruppen (durch explizite Nennung der anderen Community oder durch Regulierung der Verwendung von Neologismen anderer maskulinistischer Gruppen (z.B. verbieten von „Incel-Sprache“) Angabe, welche Personen teilnehmen dürfen und welche nicht Verweis auf andere Subreddits Simping, White Knighting (Verteidigung von Frauen, Gegenrede im Sinne von Frauen) Crosslinking
8	Moderation	Hinweis auf Konsequenzen bei Regelverstößen Hinweis auf Aufgaben der Community-Moderation Hinweis darauf, in welchen Fällen Community-Moderatoren zu kontaktieren sind

Variable: **Gruppenidentitätsstörende Inhalte**

CODE	Ausprägungen	Indikatoren	Beispiel
NA	Bedeutung des Kommentars konnte auch unter Einbeziehung des dazugehörigen Postings und weiterer Kommentare nicht rekonstruiert werden		
4	**Keine Gefährdung der Gruppenidentität**	Kommentar entspricht keinem der hier angegebenen Indikatoren	
1	**Störung der KATEGORISIERUNG** Gruppenidentifikation baut auf der Zugehörigkeit zu einer Kategorie auf (Im Anwendungsfall: Männer mit niedriger Körpergröße)	o Person gibt preis, dass sie eine Frau ist o Person gibt preis, dass sie eine große Körpergröße hat	*„I'm 5'7". That's tall-ish for a woman. Yes, ideally I would like my partner to be 4+ inches taller than me."* Timestamp: 1768655484
2	**Störung der HOMOGENITÄT** Statt Gemeinsamkeiten werden durch den Kommentar Unterschiede zwischen den Gruppenmitgliedern deutlich gemacht	o Kommentar weist eine alternative Deutungsweise in Bezug auf die Konfliktstellung der Community auf o Kommentar weist einen oppositionellen Standpunkt zur dominanten Deutungsweise auf (Gegenrede)	*„Dudes im 5'8 dating a 6'3 volleyball player that looks like a victorias secret model. Im not tall by any means. Be nice, kind and genuine and project your masculinity no matter what your height."* Timestamp: 1767601461
3	**Störung OUT-GROUP vs. IN-GROUP** Das negative Verhältnis zu Out-Groups oder die Grenzen der In-Group werden in Frage gestellt	o Positive Erwähnung großer Personen, Frauen, Feminist*innen oder anderer maskulinistischer Gruppen o Kommentar weist Inzivilität (Beleidigung, Hassrede) gegenüber Community-Mitgliedern auf	„*I kind of disagree man. My best friend is 65 and hes a typical tatted up fit guy but hes legit fought people for disrespecting me at clubs and parties. It might not be common, but good tall dudes are out there*." Timestamp: 1767322873

Variable: **Toxisches Verhalten**

CODE	Ausprägung	Indikatoren
1	Kommentar weist toxisches Verhalten auf	Unhöflichkeit o Persönliche Beleidigung o Übertreibung o Bezichtigung des Lügens o Vulgäre Sprache Inzivilität o Ablehnung der Demokratie o Gefährdung der Rechte anderer o Verwendung von Stereotypen in der Beschreibung einer Personengruppe
2	Kommentar weist kein toxisches Verhalten auf	Keine der Indikatoren aus der für Code 1 aufgestellten Ausprägung trifft auf den Inhalt zu

R Code Scraping der Regeln maskulinistischer Communities

```
library(jsonlite)
library(dplyr)
library(writexl)


subreddits <- c("MensRights", "WhereAreAllTheGoodMen", "WomenAreViolentToo",
"datingadviceformen", "BlackPillScience",
                "LeftWingMaleAdcovates", "shortguys", "PickUpArtist", "Egalitarianism",
"everydaymisandry", "DBDR",
                "menslives", "WomenAreNotIntoMen", "PurplePillDebate",
"TheAntiMisandry", "marriedredpill", "seduction",
                "pickup", "antifeminist", "whatmendon'tsay", "SikeOrPsyche",
"MensVenting", "AverageHeightDudes",
                "BasedCampPod", "CircumcisionGrief", "CityCrusherYT", "FeMRADebates",
"IncelSolutions", "intactivism",
                "JordanPeterson", "JustMemesForUs", "LengfOrGirf", "Male_Studies",
"masculinity_rocks", "moreplatesmoredates",
                "people_drawing_women", "ProMaleMemes", "SupportForTheAccused",
"thepassportbros", "TheRedPillStories", "ToxicFeminismIsToxis")


get_subreddit_rules <- function(subreddit) {
  url <- paste0("https://www.reddit.com/r/", subreddit, "/about/rules.json")


  tryCatch({
    rules_data <- fromJSON(url, flatten = TRUE)
    rules_list <- rules_data$rules


    data.frame(
      Subreddit = subreddit,
      Titel = rules_list$short_name,
      Text = rules_list$description,
      stringsAsFactors = FALSE
    )
  }, error = function(e) {
    message(paste("❌ Fehler bei", subreddit, ":", e$message))
    return(NULL)
  })
}

# Subreddit-Regeln zusammenführen
all_rules <- lapply(subreddits, get_subreddit_rules) %>%
  bind_rows()


write_xlsx(all_rules, path = "subreddit_rules_240126.xlsx")

print(all_rules)
```

R Code Scraping der Kommentare aus r/shortguys

```
library(RedditExtractoR)
library(writexl)

ShortGuysThreads <- find_thread_urls(subreddit="shortguys", sort_by="new",
period="week")

ShortGuysComments <- get_thread_content(ShortGuysThreads$url)

View(ShortGuysComments$comments)

write_xlsx(ShortGuysComments$comments, "shortguys_comments_220126_2130.xlsx")
```

R Code zur Rekonstruktion von Texten: Abgleich der Datensätze auf entfernte Kommentare

```
library(readxl)
library(writexl)

# 1. Datensätze laden

df030126_1730 <- read_excel("shortguys_comments_030126_1730.xlsx")
df020126_1230 <- read_excel("shortguys_comments_020126_1230.xlsx")
df020126_1300 <- read_excel("shortguys_comments_020126_1300.xlsx")
df020126_1430 <- read_excel("shortguys_comments_020126_1430.xlsx")
df020126_1800 <- read_excel("shortguys_comments_020126_1800.xlsx")
df020126_1830 <- read_excel("shortguys_comments_020126_1830.xlsx")
df020126_2200 <- read_excel("shortguys_comments_020126_2200.xlsx")
df020126_2300 <- read_excel("shortguys_comments_020126_2300.xlsx")
df030126_0930 <- read_excel("shortguys_comments_030126_0930.xlsx")
df030126_1300 <- read_excel("shortguys_comments_030126_1300.xlsx")
df050126_1100 <- read_excel("shortguys_comments_050126_1100.xlsx")
df050126_1130 <- read_excel("shortguys_comments_050126_1130.xlsx")
df050126_1400 <- read_excel("shortguys_comments_050126_1400.xlsx")
df050126_1500 <- read_excel("shortguys_comments_050126_1500.xlsx")
df050126_1630 <- read_excel("shortguys_comments_050126_1630.xlsx")
df050126_1730 <- read_excel("shortguys_comments_050126_1730.xlsx")
df050126_1830 <- read_excel("shortguys_comments_050126_1830.xlsx")
df050126_2230 <- read_excel("shortguys_comments_050126_2230.xlsx")
df050126_2300 <- read_excel("shortguys_comments_050126_2300.xlsx")
df060126_0900 <- read_excel("shortguys_comments_060126_0900.xlsx")
df060126_1100 <- read_excel("shortguys_comments_060126_1100.xlsx")
df060126_1200 <- read_excel("shortguys_comments_060126_1200.xlsx")
df060126_1700 <- read_excel("shortguys_comments_060126_1700.xlsx")
df060126_1900 <- read_excel("shortguys_comments_060126_1900.xlsx")
df070126_0900 <- read_excel("shortguys_comments_070126_0900.xlsx")
df070126_1230 <- read_excel("shortguys_comments_070126_1230.xlsx")
df070126_1400 <- read_excel("shortguys_comments_070126_1400.xlsx")
df070126_1500 <- read_excel("shortguys_comments_070126_1500.xlsx")
df070126_1700 <- read_excel("shortguys_comments_070126_1700.xlsx")
df080126_0800 <- read_excel("shortguys_comments_080126_0800.xlsx")
df080126_1430 <- read_excel("shortguys_comments_080126_1430.xlsx")
df080126_1600 <- read_excel("shortguys_comments_080126_1600.xlsx")
df080126_1730 <- read_excel("shortguys_comments_080126_1730.xlsx")
df080126_2100 <- read_excel("shortguys_comments_080126_2100.xlsx")
df080126_2300 <- read_excel("shortguys_comments_080126_2300.xlsx")
df090126_0800 <- read_excel("shortguys_comments_090126_0800.xlsx")
df090126_1100 <- read_excel("shortguys_comments_090126_1100.xlsx")
df090126_1430 <- read_excel("shortguys_comments_090126_1430.xlsx")
df090126_1900 <- read_excel("shortguys_comments_090126_1900.xlsx")
df090126_2230 <- read_excel("shortguys_comments_090126_2230.xlsx")
df100126_1300 <- read_excel("shortguys_comments_100126_1300.xlsx")
df100126_1800 <- read_excel("shortguys_comments_100126_1800.xlsx")
df100126_2230 <- read_excel("shortguys_comments_100126_2230.xlsx")
df110126_1600 <- read_excel("shortguys_comments_110126_1600.xlsx")
df120126_0800 <- read_excel("shortguys_comments_120126_0800.xlsx")
df120126_1300 <- read_excel("shortguys_comments_120126_1300.xlsx")
df120126_1700 <- read_excel("shortguys_comments_120126_1700.xlsx")
df130126_0830 <- read_excel("shortguys_comments_130126_0830.xlsx")
df130126_1300 <- read_excel("shortguys_comments_130126_1300.xlsx")
df130126_2300 <- read_excel("shortguys_comments_130126_2300.xlsx")
df140126_0900 <- read_excel("shortguys_comments_140126_0900.xlsx")
df140126_1300 <- read_excel("shortguys_comments_140126_1300.xlsx")
df140126_1900 <- read_excel("shortguys_comments_140126_1900.xlsx")
df150126_0800 <- read_excel("shortguys_comments_150126_0800.xlsx")
df150126_1530 <- read_excel("shortguys_comments_150126_1530.xlsx")
df150126_1930 <- read_excel("shortguys_comments_150126_1930.xlsx")
df150126_2130 <- read_excel("shortguys_comments_150126_2130.xlsx")
df160126_1000 <- read_excel("shortguys_comments_160126_1000.xlsx")
df160126_1600 <- read_excel("shortguys_comments_160126_1600.xlsx")
df160126_2300 <- read_excel("shortguys_comments_160126_2300.xlsx")
df170126_1630 <- read_excel("shortguys_comments_170126_1630.xlsx")
df180126_0100 <- read_excel("shortguys_comments_180126_0100.xlsx")
df180126_1700 <- read_excel("shortguys_comments_180126_1700.xlsx")
df180126_2130 <- read_excel("shortguys_comments_180126_2130.xlsx")
```

```
df190126_1200 <- read_excel("shortguys_comments_190126_1200.xlsx")
df190126_1500 <- read_excel("shortguys_comments_190126_1500.xlsx")
df190126_1800 <- read_excel("shortguys_comments_190126_1800.xlsx")
df190126_2200 <- read_excel("shortguys_comments_190126_2200.xlsx")
df200126_0730 <- read_excel("shortguys_comments_200126_0730.xlsx")
df200126_1200 <- read_excel("shortguys_comments_200126_1200.xlsx")
df200126_1500 <- read_excel("shortguys_comments_200126_1500.xlsx")
df200126_2200 <- read_excel("shortguys_comments_200126_2200.xlsx")
df210126_0800 <- read_excel("shortguys_comments_210126_0800.xlsx")
df210126_1730 <- read_excel("shortguys_comments_210126_1730.xlsx")
df220126_0830 <- read_excel("shortguys_comments_220126_0830.xlsx")
df220126_1200 <- read_excel("shortguys_comments_220126_1200.xlsx")
df220126_1530 <- read_excel("shortguys_comments_220126_1530.xlsx")
df220126_2130 <- read_excel("shortguys_comments_220126_2130.xlsx")
df230126_0800 <- read_excel("shortguys_comments_230126_0800.xlsx")
df230126_1330 <- read_excel("shortguys_comments_230126_1330.xlsx")
df230126_1630 <- read_excel("shortguys_comments_230126_1630.xlsx")
df230126_2130 <- read_excel("shortguys_comments_230126_2130.xlsx")
df240126_0800 <- read_excel("shortguys_comments_240126_0800.xlsx")
df240126_1030 <- read_excel("shortguys_comments_240126_1030.xlsx")
df240126_1500 <- read_excel("shortguys_comments_240126_1500.xlsx")


# 2. Ziel-Timestamps
timestamps_df <- read_excel("timestamps_unique_commentID_R_OUTPUT.xlsx")

target_timestamps <- timestamps_df$timestamp
target_comment_ids <- timestamps_df$comment_id


# 3. Datensätze sammeln
dfs <- list(
  df020126_1230 = df020126_1230,
  df020126_1300 = df020126_1300,
  df020126_1430 = df020126_1430,
  df020126_1800 = df020126_1800,
  df020126_1830 = df020126_1830,
  df020126_2200 = df020126_2200,
  df020126_2300 = df020126_2300,
  df030126_0930 = df030126_0930,
  df030126_1300 = df030126_1300,
  df030126_1730 = df030126_1730,
  df050126_1100 = df050126_1100,
  df050126_1130 = df050126_1130,
  df050126_1400 = df050126_1400,
  df050126_1500 = df050126_1500,
  df050126_1630 = df050126_1630,
  df050126_1730 = df050126_1730,
  df050126_1830 = df050126_1830,
  df050126_2230 = df050126_2230,
  df050126_2300 = df050126_2300,
  df060126_0900 = df060126_0900,
  df060126_1100 = df060126_1100,
  df060126_1200 = df060126_1200,
  df060126_1700 = df060126_1700,
  df060126_1900 = df060126_1900,
  df070126_0900 = df070126_0900,
  df070126_1230 = df070126_1230,
  df070126_1400 = df070126_1400,
  df070126_1500 = df070126_1500,
  df070126_1700 = df070126_1700,
  df080126_0800 = df080126_0800,
  df080126_1430 = df080126_1430,
  df080126_1600 = df080126_1600,
  df080126_1730 = df080126_1730,
  df080126_2100 = df080126_2100,
  df080126_2300 = df080126_2300,
  df090126_0800 = df090126_0800,
  df090126_1100 = df090126_1100,
```

```
  df090126_1430 = df090126_1430,
  df090126_1900 = df090126_1900,
  df090126_2230 = df090126_2230,
  df100126_1300 = df100126_1300,
  df100126_1800 = df100126_1800,
  df100126_2230 = df100126_2230,
  df110126_1600 = df110126_1600,
  df120126_0800 = df120126_0800,
  df120126_1300 = df120126_1300,
  df120126_1700 = df120126_1700,
  df130126_0830 = df130126_0830,
  df130126_1300 = df130126_1300,
  df130126_2300 = df130126_2300,
  df140126_0900 = df140126_0900,
  df140126_1300 = df140126_1300,
  df140126_1900 = df140126_1900,
  df150126_0800 = df150126_0800,
  df150126_1530 = df150126_1530,
  df150126_1930 = df150126_1930,
  df150126_2130 = df150126_2130,
  df160126_1000 = df160126_1000,
  df160126_1600 = df160126_1600,
  df160126_2300 = df160126_2300,
  df170126_1630 = df170126_1630,
  df180126_0100 = df180126_0100,
  df180126_1700 = df180126_1700,
  df180126_2130 = df180126_2130,
  df190126_1200 = df190126_1200,
  df190126_1500 = df190126_1500,
  df190126_1800 = df190126_1800,
  df190126_2200 = df190126_2200,
  df200126_0730 = df200126_0730,
  df200126_1200 = df200126_1200,
  df200126_1500 = df200126_1500,
  df200126_2200 = df200126_2200,
  df210126_0800 = df210126_0800,
  df210126_1730 = df210126_1730,
  df220126_0830 = df220126_0830,
  df220126_1200 = df220126_1200,
  df220126_1530 = df220126_1530,
  df220126_2130 = df220126_2130,
  df230126_0800 = df230126_0800,
  df230126_1330 = df230126_1330,
  df230126_1630 = df230126_1630,
  df230126_2130 = df230126_2130,
  df240126_0800 = df240126_0800,
  df240126_1030 = df240126_1030,
  df240126_1500 = df240126_1500)

# 4. Presence-Tabelle mit ALLEN Kommentaren
presence <- data.frame(
  timestamp  = target_timestamps,
  comment_id = target_comment_ids
)

for (name in names(dfs)) {
  df <- dfs[[name]]

  presence[[paste0(name, "_comments")]] <-
    mapply(function(t, cid) {

      rows <- df[df$timestamp == t & df$comment_id == cid, ]

      if (nrow(rows) == 0) {
        NA
      } else {
        paste(unique(rows$comment), collapse = " | ")
      }
```

```
    }, target_timestamps, target_comment_ids)
}


# 5. Ergebnisse anzeigen
View(presence)
write_xlsx(presence, "CheckForRemovedComments_240126_1600.xlsx")
```

R Code Fisher-Test zur Interaktion zwischen dem Themenfokus der Communities und den verwendeten Regeltypen

```
library(readxl)
library(dplyr)
library(ggplot2)
library(janitor)
library(vcd)
library(grid)



df_rules <- read_excel("subreddit_rules_FINAL.xlsx")


str(df_rules)
head(df_rules)



df_rules <- df_rules %>%
  mutate(
    topic_subreddit = factor(
      topic_subreddit,
      levels = c(1, 2, 3, 4, 5, 6),
      labels = c("antifem_organisiert", "antifem_person", "antifem_unabhängig",
"individualistisch_phil", "individualistisch_prag", "gemischt")
    ),
    rule_CODE = factor(
      rule_CODE,
      levels = c(1, 2, 3, 4, 5, 6, 7, 8, 9),
      labels = c("quality of content", "format of content", "reddit content-policies",
"toxic behavior", "hate speech", "topic subreddit", "demarcation", "moderation", "no
rules"
      )
    )
  )




data_complete <- df_rules %>%
  filter(!is.na(topic_subreddit) & !is.na(rule_CODE))

#Deskriptive Statistiken

tab_status <- tabyl(data_complete, topic_subreddit)
tab_status


tab_comment <- tabyl(data_complete, rule_CODE)
tab_comment


#Kreustabelle erstellen
crosstab <- table(data_complete$topic_subreddit, data_complete$rule_CODE)
crosstab


# Prozentuale Verteilung
prop.table(crosstab, margin = 1)

#Vorraussetzungen prüfen
chi_test_pre <- chisq.test(crosstab)

# Erwartete Häufigkeiten
chi_test_pre$expected
```

```
#Bedigungen Chi-Quadrat-Test sind nicht erfüllt, daher Fisher-Test

fisher_test <- fisher.test(crosstab, simulate.p.value = TRUE, B = 10000)

fisher_test


#Effektstärke mittels Cramers V berechnen

cramers_v <- assocstats(crosstab)

cramers_v

#Visualisierung
ggplot(data_complete, aes(x = topic_subreddit, fill = rule_CODE)) +
  geom_bar(position = "stack") +
  labs(
    x = "Thema des Subreddits",
    y = "Anzahl der Regelarten",
    fill = "Regeltypen",
    title = "Verteilung der Regeltypen nach themenfokus der antifeministischen
Subreddits"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 15, hjust = 1)
  )
```

R Code Welch-Test für die Interaktion zwischen dem Voting-Score und der Verfügbarkeit eines Kommentars

```
library(readxl)
library(car)
library(dplyr)
library(effectsize)
library(mosaic)
library (knitr)
library (tidyr)
library(ggpubr)


df <- read_excel("shortguys_upvotes_FINAL.xlsx")


df$uv <- factor(df$status_code,
                levels = c (1,2,3, 4),
                labels = c("deleted", "Removed by Moderator", "Removed by Reddit",
"available"))

df$av <- df$upvotes

glimpse(df)


label_uv <- "Verfügbarkeit Kommentar"
label_av <- "Upvotes"



des_stat1 <- favstats(df$av ~ df$uv)

kable (des_stat1,
       col.names = c(label_uv,"Minimum", "1.Quartil",
                     "Median", "3.Quartil", "Maximum",
                     "M", "SD", "N", "Fehlend" ),
       digits = 2)



#Vorraussetzung prüfen mittels Levene-Test
df %>%
  drop_na(uv, av) %>%
  mutate(uv = as.factor(uv)) %>%
  {leveneTest(as.formula("av ~ uv"), data = .)}




#Varianzanalyse durchführen

df_anova <- df %>%
  drop_na(uv, av) %>%
  mutate(uv = as.factor(uv))

model <- aov(av ~ uv, data = df_anova)
summary(model)


#Effektgröße Varianzanalyse
df_effekt <- df %>%
  drop_na(uv, av) %>%
  mutate(uv = as.factor(uv))

model <- aov(av ~ uv, data = df_effekt)

eta_squared(model, partial = FALSE)
```

```
#Welch-Test durchführen
welch_test <- oneway.test(av ~ uv,
                          data = df,
                          var.equal = FALSE)

welch_test

#Welch-Test Effektgröße Überprüfen
eta_squared(welch_test)




#Visualisierung

Titel <- "Zusammenhang Votingscore und Moderationsstatus"

ggline(df %>% drop_na (av, uv),
       x = "uv",
       y = "av",
       title = Titel,
       add = "mean_ci",
       xlab = label_uv,
       ylab = label_av)
```

R Code log-lineares Model für die Dreifach-Interaktion zwischen gruppenidentitätsgefährenden Inhalten, toxischem Verhalten und dem Moderationsstatus

```
library(readxl)
library(dplyr)
library(MASS)
library(vcd)
library(janitor)

data <- read_excel("REDUCED_shortguys_comments.xlsx")

str(data)
head(data)

data <- data %>%
  mutate(
    comment_code = factor(comment_code,
                          levels = c(1,3,4,5),
                          labels = c("Kategorisierung",
                                     "Homogenität",
                                     "Abgrenzung",
                                     "Keine Störung")),

    status_code = factor(status_code,
                         levels = c(1,2),
                         labels = c("nicht moderiert",
                                    "moderiert")),

    incivility_code = factor(incivility_code,
                             levels = c(1, 3),
                             labels = c("toxisches Verhalten",
                                        "kein toxisches Verhalten"))
  )

#3-Dimensionale Kontingenztabelle
table_3d <- xtabs(~ comment_code + status_code + incivility_code, data = data)
table_3d


#Modell mit Interaktion zwischen allen 3 Variablen
model_full <- loglm(~ comment_code * status_code * incivility_code,
                    data = table_3d)

summary(model_full)


#Modell ohne Interaktion zwischen allen 3 Variablen
model_no3way <- loglm(~ comment_code * status_code +
                        comment_code * incivility_code +
                        status_code * incivility_code,
                      data = table_3d)

summary(model_no3way)


#Modellvergleich
anova(model_no3way, model_full)


#Erwartete Häufigkeiten prüfen
expected <- fitted(model_no3way)
min(expected)
sum(expected < 5) / length(expected)
```

R Code Chi-Quadrat-Tests für die Interaktionen zwischen toxischem Verhalten, gruppenidentitätsstörenden Inhalten und dem Moderationsstatus

```
library(readxl)
library(dplyr)
library(ggplot2)
library(vcd)
library(janitor)
library(survey)
library(grid)


comments <- read_excel("REDUCED_shortguys_comments.xlsx")



comments <- comments %>%
  mutate(
    comment_code = factor(comment_code,
                          levels = c(1,3,4,5),
                          labels = c("Störung\nKategorisierung",
                                     "Störung\nHomogenität",
                                     "Störung\nAbgrenzung",
                                     "Keine\nStörung")),

    status_code = factor(status_code,
                         levels = c(1,2),
                         labels = c("nicht\nmoderiert",
                                    "moderiert")),

    incivility_code = factor(incivility_code,
                             levels = c(1,3),
                             labels = c("toxisches\nVerhalten",
                                        "Kein\ntoxisches\nVerhalten"))
  )



comments_complete <- comments %>%
  filter(!is.na(comment_code) &
           !is.na(status_code) &
           !is.na(incivility_code))


#Gewichtung

comments_complete <- comments_complete %>%
  mutate(
    weight = ifelse(status_code == "nicht\nmoderiert",
                    0.983 / 0.75,
                    0.017 / 0.25)
  )


design <- svydesign(
  ids = ~1,
  weights = ~weight,
  data = comments_complete
)



# Interaktion comment_code × status_code


tab1_w <- svytable(~ comment_code + status_code, design)

chi1 <- svychisq(~ comment_code + status_code, design)
chi1
```

```
# Erwartete Häufigkeiten
row_tot1 <- rowSums(tab1_w)
col_tot1 <- colSums(tab1_w)
n_tot1 <- sum(tab1_w)
expected1 <- outer(row_tot1, col_tot1) / n_tot1

min(expected1)
mean(expected1 >= 5)

# Cramér's V
chi_sq1 <- chi1$statistic
df_min1 <- min(nrow(tab1_w)-1, ncol(tab1_w)-1)
cramers_v1 <- sqrt(chi_sq1 / (n_tot1 * df_min1))
cramers_v1

#Visualisierung

plot1_df <- plot1_df %>%
  group_by(status_code) %>%
  mutate(Percent = Freq / sum(Freq))

ggplot(plot1_df,
       aes(x = status_code,
           y = Percent,
           fill = comment_code)) +
  geom_bar(stat = "identity", position = "fill", color = "black") +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c(
    "#6BAED6",  # helles Blau
    "#66C2A5",  # helles Petrol
    "#F2D27E",  # helles Sand
    "#D0849D"   # helles Rosé/Weinrot
  )) +
  labs(x = "Moderationsstatus",
       y = "Prozent",
       fill = "Art des gruppenidentitäts-\nstörenden Inhalts") +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 15, hjust = 1),
        legend.position = "top")




# Interaktion comment_code × incivility_code


tab2_w <- svytable(~ comment_code + incivility_code, design)

chi2 <- svychisq(~ comment_code + incivility_code, design)
chi2

# Erwartete Häufigkeiten
row_tot2 <- rowSums(tab2_w)
col_tot2 <- colSums(tab2_w)
n_tot2 <- sum(tab2_w)
expected2 <- outer(row_tot2, col_tot2) / n_tot2

min(expected2)
mean(expected2 >= 5)

# Cramér's V
chi_sq2 <- chi2$statistic
df_min2 <- min(nrow(tab2_w)-1, ncol(tab2_w)-1)
cramers_v2 <- sqrt(chi_sq2 / (n_tot2 * df_min2))
cramers_v2
```

```
# Interaktion status_code × incivility_code


tab3_w <- svytable(~ status_code + incivility_code, design)

chi3 <- svychisq(~ status_code + incivility_code, design)
chi3

# Erwartete Häufigkeiten
row_tot3 <- rowSums(tab3_w)
col_tot3 <- colSums(tab3_w)
n_tot3 <- sum(tab3_w)
expected3 <- outer(row_tot3, col_tot3) / n_tot3

min(expected3)
mean(expected3 >= 5)

# Cramér's V
chi_sq3 <- chi3$statistic
df_min3 <- min(nrow(tab3_w)-1, ncol(tab3_w)-1)
cramers_v3 <- sqrt(chi_sq3 / (n_tot3 * df_min3))
cramers_v3

#Visualisierung

plot3_df <- as.data.frame(tab3_w)

# Prozent innerhalb von toxischem Verhalten berechnen
plot3_df <- as.data.frame(tab3_w)

plot3_df <- plot3_df %>%
  group_by(incivility_code) %>%
  mutate(Percent = Freq / sum(Freq))

ggplot(plot3_df,
       aes(x = incivility_code,
           y = Percent,
           fill = status_code)) +
  geom_bar(stat = "identity", position = "fill", color = "black") +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c(
    "#6BAED6",  # helles Blau
    "#F4A582"   # helles Rosé/Orange
  )) +
  labs(x = "Toxisches Verhalten",
       y = "Prozent",
       fill = "Moderationsstatus") +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 15, hjust = 1),
        legend.position = "top")
```