

Contenuti contaminanti: odio e violenza nei social media

MARIA ROSARIA VITALE*

DOI: <https://doi.org/10.15162/1827-5133/2236>

ABSTRACT

I contenuti *user-generated* sono la linfa vitale dei social media, e la *content moderation* è necessaria per limitarne varie categorie. Ciò include i contenuti d'odio e violenti, sebbene possano rimanere visibili. Il saggio esplora la relazione tra le strategie di *content moderation* e le concettualizzazioni di odio e violenza. Viene posta una domanda filosofico-politica relativa a come la visibilità e la regolamentazione dei contenuti d'odio e violenti modellano le dinamiche di aggregazione dell'utenza.

User-generated content is the lifeblood of social media platforms, and *content moderation* is required to restrict various categories of it. This includes hateful and violent content, although it may still be visible. The essay explores the relationship between *content moderation* strategies and conceptualizations of hate and violence. It raises the politico-philosophical question of how the visibility and regulation of hateful and violent content shapes the dynamics of user aggregation.

* Maria Rosaria Vitale è assegnista di ricerca in Filosofia Politica presso l'Università degli Studi di Urbino Carlo Bo.

Il quindici marzo del 2019, un utente di Facebook riprende, tramite la video-camera che indossa, la prima delle due sparatorie di cui è il responsabile a Christchurch, in Nuova Zelanda¹. I diciassette minuti del video vengono pubblicati in tempo reale, usufruendo della funzione *Facebook Live*. Lo stesso video rimane visibile nel profilo dell'utente, per poi essere rimosso dal team di moderazione, il cui lavoro consiste anche nella valutazione dei cosiddetti contenuti *user-generated* non ammessi. Il profilo social viene disattivato, ma copie del video originale vengono create da altri utenti, e commenti a sostegno della sparatoria sono pubblicati. Diciassette minuti di violenza, ventiquattro ore di eliminazione di più di un milione di contenuti e di decontaminazione tramite la cosiddetta *content moderation*². È quest'ultima a definire ciò che può essere e non può essere pubblicato in qualsiasi social media, ponendo le condizioni di espressione e interconnessione dell'utenza tramite i contenuti, appunto, *user-generated*³.

I contenuti si dividono in categorie di contenuti pubblicabili e non pubblicabili. Il video dell'atto violento del suprematista bianco rientra nella categoria dei contenuti non ammessi da Facebook, le cui vittime sono state soggetti musulmani. Al riguardo, la *content moderation* va assunta quale mera forma di regolamentazione interna dei social media⁴, sebbene alcune categorie di contenuti non pubblicabili possano coincidere con le categorie illecite di espressioni e atti anche per un qualche ordinamento giuridico statale. Il caso di Christchurch, difatti, rende conto di un atto violento sanzionabile, perché previsto dalla normativa statale neozelandese quale *hate crime*. I commenti a sostegno della sparatoria costituiscono possibili casi di *hate speech*, previsti da

¹ La seconda sparatoria non viene pubblicata né su Facebook né su nessun'altra piattaforma. Si veda, per esempio, G. Macklin, "The Christchurch Attacks: Livestream Terror in the Viral Video Age", in «CTC Sentinel», vol. XII, 2019, n. 6; M. Harwood, "Living Death: Imagined History and the Tarrant Manifesto", in «Emotions: History, Culture, Society», vol. V, 2021, n.1.

² Vedi <<https://about.fb.com/news/2019/03/update-on-new-zealand/>> (consultato il 25/03/25).

³ Si veda, per esempio, T. Gillespie *et al.*, "Expanding the debate about content moderation: scholarly research agendas for the coming policy debates", in «Internet Policy Review», vol. IX, 2020, n. 4; L. Edelson, "Content Moderation in Practice", in *Media and Society After Technological Disruption*, K. Langvardt e J. Hurwitz (a cura di), Cambridge University Press, Cambridge 2024.

⁴ Si veda, per esempio, E. Douek, "Content Moderation as Systems Thinking", in «Harvard Law Review», vol. CXXXVI, 2022, n. 2.

diverse normative statali, considerando che l'utenza, per quanto globale, rinvia a più giurisdizioni⁵.

Le condizioni di espressione e interconnessione poste dai social media strutturano, quindi, opportunità che divengono problematiche anche al di là del caso neozelandese: i contenuti d'odio e i contenuti violenti sono visibili, nonostante non dovrebbero esserlo, e questa anomalia sarà oggetto di analisi. Qualche riga sopra, è stato utilizzato il termine *decontaminazione*. La mia scelta rinvia alla considerazione, nel dibattito scientifico e non-scientifico, della presunta tossicità dell'odio e della violenza nei social media⁶. Che l'odio e la violenza online siano il *nuovo male* delle società liberal-democratiche è una plausibile ipotesi di ricerca. In questo lavoro si intende problematizzare, in termini filosofico-politici, la possibilità che specifici contenuti *user-generated* aggregino i soggetti-utenti di spazi digitali. In considerazione della rilevanza politica della questione, vengono presentate due concettualizzazioni di odio e di violenza, con lo scopo di comprendere in quale misura esse rendano conto dei contenuti indagati, in un'intersezione con la *content moderation* dei social media.

Il riferimento ad una delle sparatorie di Christchurch è stato illustrato solo a mo' di esempio. Difatti, dei contenuti definiti non ammissibili da certi social media sono stati pubblicati prima di quel quindici marzo, così come successivamente. I contenuti, inoltre, sono multiformi (trattandosi di testi, video, audio, immagini) e la percezione di una loro ubiquità resistente può essere assunta. Al riguardo, i social media rinnovano costantemente il proprio impegno nei confronti di una moderazione attenta: la visibilità dei contenuti d'odio e dei contenuti violenti permette, tuttavia, di analizzare come quello stesso impegno possa essere considerato mancante.

⁵ Si veda per esempio, S. Leitch e P. Pickering, *Rethinking Social Media and Extremism*, ANU Press, Canberra 2022; R. Gorwa, "After Christchurch. Diverging Regulatory Responses in New Zealand and Australia", in Id., Gorwa, *The Politics of Platform Regulation: How Governments Shape Online Content Moderation*, Oxford University Press, Oxford and New York 2024.

⁶ Si veda, per esempio, R. Cohen-Almagor, *Confronting the Internet Dark Side. Moral and Social Responsibilities on the Free Highway*, Cambridge University Press, Cambridge 2015; J. Waldron, *The Harm in Hate Speech*, Harvard University Press, Cambridge (MA) 2012; K. Lumsden e E. Harmer (a cura di), *Online Othering. Exploring Digital Violence and Discrimination on the Web*, Palgrave Macmillan, Cham 2019.

Una concettualizzazione che si intende analizzare è quella di odio ascrivito. Definisco l'odio *ascrittivo* nella misura in cui il soggetto è odiato alla luce delle caratteristiche ascrivite (effettivamente o presumibilmente) possedute. Razzismo, xenofobia, sessismo, transfobia, antisemitismo, islamofobia, omofobia, abilismo saranno, quindi, indagabili come possibili forme di odio tramite le caratteristiche ascrivite di razza, origine etnica, origine nazionale, genere, religione, orientamento sessuale, disabilità⁷.

A mio avviso, tali caratteristiche ascrivite costituiscono le condizioni analitiche minime in quanto indicative di un'attribuzione di comportamenti, credenze, attitudini al soggetto odiato che sarà, per esempio, assunto quale non-umano, sub-umano, moralmente deplorabile, socialmente dannoso. È possibile così individuare delle qualificazioni che, pur essendo multiformi, si pongono come elementi negativizzanti e generalizzanti. Difatti, come sostengono Brudholm e Schepern Johansen, "non si odia mai semplicemente un altro 'perché è nero'. Un soggetto odia questo altro perché si associa l'essere nero a una minaccia"⁸. Ciò vale anche quando una sua inferiorità o subordinazione, variamente interpretabile, venga rivendicata e qualunque sia la specifica caratteristica ascrivita da cui l'odio muove.

L'odio crea delle associazioni tra le caratteristiche ascrivite e il soggetto odiato, in una fossilizzazione che è resistente. Più nello specifico, il destinatario o target di odio viene de-individualizzato: non è odiato in quanto singolo e specifico soggetto, ma meramente perché membro del gruppo al quale appartiene su base ascrivita. Le associazioni proprie dell'odio investono chiunque possiede le medesime caratteristiche ascrivite. La fossilizzazione prima citata definisce, quindi, una distribuzione uniforme delle associazioni create,

⁷ Si veda, per esempio, A. Brown, "The 'who?' question in the hate speech debate: Part 1: Consistency, practical and formal approaches", in «Canadian Journal of Law & Jurisprudence», vol. XXIX, 2016, n. 2; T. Brudholm e B. Schepern Johansen (a cura di), *Hate, Politics, Law. Critical Perspectives on Combating Hate*, Oxford University Press, Oxford and New York 2018.

⁸ T. Brudholm e B. Schepern Johansen, "Pondering Hatred", in *Emotions and Mass Atrocity. Philosophical and Theoretical Explorations*, T. Brudholm e J. Lang (a cura di), Cambridge University Press, Cambridge and New York 2018, pp. 91-92.

tramite cui ogni differenziazione individuale viene annullata⁹. L'altro odiato diventa un *loro*. Ed è questo elemento che sembra poter permettere un'analisi per contrapposizione, seguendo una logica per appartenenza: se un *loro* è discernibile tramite l'odio, qual è il *noi*?

Da una prospettiva filosofico-politica, le società liberal-democratiche sono tali anche in termini delle condizioni necessarie per il riconoscimento dello status di soggetti eguali. Qui la questione è specificatamente normativa. Con Alfieri, "qualsiasi aggregato sociale [...] si realizza soltanto attraverso la costruzione di confini: i confini non sarebbero tali se non ci fosse qualcuno che ne resta al di fuori [...]". Inclusione ed esclusione si richiamano a vicenda: sono due poli ugualmente necessari di un'unica realtà"¹⁰. L'autorità politica liberal-democratica definisce così un *noi* costituito da soggetti eguali, e tra questi troviamo membri di gruppi paradigmaticamente target di odio¹¹. Tuttavia, tale *noi* non coincide esattamente con quello dell'odio ascrivito. Il primo è oggetto di contestazione, in quanto rivendicato quale eccessivamente inclusivo. Dalla prospettiva di chi odia, potrebbe seguire una domanda: 'perché condividere lo stesso status con criminali, parassiti, devianti, peccatori, con chi non si adegua al proprio ruolo o che costituisce un generale *male*?', trattandosi delle associazioni fossilizzanti e generalizzanti dell'odio ascrivito. Il *noi* va, quindi, ridefinito e ristretto nei suoi confini, con la condivisione del medesimo status di *eguale* che viene posta quale questione problematica: a mio avviso, si tratta della rivendicazione di un *surplus di diritti* di cui il soggetto odiato è titolare e che rende analizzabile uno scenario di contestazione, anche in relazione a contenuti individuabili entro i social media generalisti.

Nell'estate del 2014, l'hashtag *#gamergate* si diffonde su Internet come campagna avente lo scopo di sensibilizzare sull'etica giornalistica inerente ai videogiochi, sebbene a detta di Massanari sia "servita come vessillo per le molestie su larga scala subite da prominenti giornaliste, *developer* di videogiochi e critiche di media", svelando "paure legate alla percezione della diminuzione

⁹ Si veda, per esempio, T. Szanto, "In Hate we Trust: The Collectivization and Habitualization of Hatred", in «Phenomenology and the Cognitive Sciences», vol. XIX, 2020, n. 3.

¹⁰ L. Alfieri, "Il terzo che deve morire", in *Figure e simboli dell'ordine violento. Percorsi fra antropologia e filosofia politica*, L. Alfieri, C.M. Bellei e D.S. Scalzo (a cura di), Giappichelli, Torino 2003, p. 95.

¹¹ Si veda, per esempio, T. Brudholm e B. Schepern Johansen (a cura di), *Hate, Politics, Law*, cit.

dello status di giovani uomini bianchi a livello globale”¹². Per quanto anche prima di quell'estate molestie fossero state perpetrate, *gamergate* mostra una coordinazione nella difesa (sessista) della supposta identità dei soggetti *video-gamer*¹³. Più nello specifico, quello che muove dalla pubblicazione (in un blog di nicchia) di un post, da parte dell'ex compagno di una *developer* su una supposta recensione positiva ricevuta in cambio di sesso, diventa un ammasso di correlati contenuti sessisti su più piattaforme con *doxxing*, intimidazioni, minacce di stupro e morte e stereotipizzazioni.

I contenuti multiformi entro il *gamergate* hanno mostrato una notevole capacità di diffusione. Una questione relativa ad un ambito alquanto di nicchia si è riversata in piattaforme generaliste, definibili tali per la diversificazione della loro utenza e delle condizioni di espressione e interconnessione poste. Tali contenuti possono essere oggetto di riflessione tramite l'analisi di O'Donnell circa l'assunto resistente, nel contesto culturale e accademico, per cui i videogiochi siano un hobby e un ambito esclusivamente maschile¹⁴. *Gamergate* permette, quindi, di problematizzare la considerazione secondo cui le donne siano delle intruse e costituiscano un'anomalia nella *community* e nel settore del *videogaming*.

Anche dieci anni fa, la distribuzione di genere dei soggetti *videogamer* era abbastanza equilibrata, tale per cui gli uomini non erano gli unici membri della *community* del *videogaming*¹⁵. A ciò va aggiunto che i videogiochi, sviluppati e messi in commercio, si caratterizzavano per personaggi femminili quantitativamente minori rispetto a quelli maschili. Si tratta di due questioni dis-

¹² A.L. Massanari, *Gaming Democracy. How Silicon Valley Leveled Up the Far Right*, MIT Press, Cambridge (MA) 2024, pp. 27-28. Cfr. T.E. Mortensen, “Anger, Fear, and Games: The Long Event of #Gamergate”, in «Games and Culture», vol. XIII, 2018, n. 8.

¹³ Si veda, per esempio, J. Fox e W.Y. Tang, “Sexism in Online Video Games: The Role of Conformity to Masculine and Social Dominance Orientation”, in «Computer in Human Behavior», 2014, n. 3; S. Chess e A. Shaw, “A Conspiracy of Fishes, or, How We Learned to Stop Worrying about #Gamergate and Embrace Hegemonic Masculinity” in «Journal of Broadcasting & Electronic Media», vol. LIX, 2015, n. 1.

¹⁴ J. O'Donnell, *Gamergate and Anti-Feminism in the Digital Age*, Palgrave Macmillan, Cham 2022.

¹⁵ Si veda il report per l'anno 2014 dell' Entertainment Software Association: “Gender of Game Players: 52% male, 48% female”, <<https://www.theesa.com/resources/essential-facts-about-the-us-video-game-industry/>> (consultato il 25/03/25).

tinte, ma correlate: una possibile riflessione concerne l'alta frequenza della rappresentazione ipersessualizzata dei personaggi femminili¹⁶, e del suo “operare da attrazione per un pubblico maschile eterosessuale, crea[ndo] anche delle aspettative sulle donne nel mondo del *videogaming*”¹⁷, come *developer*, giornaliste, critiche dei media, giocatrici. Se specifiche donne vengono considerate quale gruppo di intruse, anomalia non prevedibile, contenuti d'odio fossilizzante e generalizzante hanno aggregato specifici uomini? Pongo tale domanda in quanto, nella letteratura scientifica sull'odio (nella forma di *hate speech* e *hate crime*), le donne costituiscono un paradigmatico target di odio ascrivito-sessista¹⁸.

A mio avviso, è interessante quanto è apparso nel sito hackerato di un critico dei contenuti in questione: “Che questo sia un avvertimento a tutte le sviluppatrici SJW [*Social Justice Warrior*] lì fuori, stiamo venendo a prendervi. [...] Noi non permetteremo che distruggiate il settore del *gaming* con le vostre tendenze femministe SJW”¹⁹. Questa rivendicazione ricorda il messaggio che Waldron considera proprio dell'odio espresso: “Il tempo della vostra degradazione e della vostra esclusione, da parte della società che attualmente vi ospita, si sta avvicinando rapidamente”²⁰. Da qui, la diffusione di *gamergate* rende conto delle modalità secondo cui una certa utenza abbia avuto nei propri schermi contenuti indicatori delle conseguenze di essere donna in un ambito percepito quale esclusivamente maschile, con le associazioni proprie dell'odio. Sono quest'ultime a strutturare una fossilizzazione tramite cui le donne sono viste, rappresentate, incolpate. Con Waldron, la questione può ri-

¹⁶ Cfr. T. Lynch *et al.*, “Sexy, Strong and Secondary: a Content Analysis of Female Characters in Video Games across 31 Years”, in «Journal of Communication», vol. LXVI, 2016, n. 4. L'ipersessualizzazione si ritrova, per esempio, nel caso di *Tomb Rider* e della sua protagonista Lara Croft, procace e intelligente, ma anche in quei personaggi femminili secondari rispetto a quelli maschili e nel *topos* della *damsel in distress*.

¹⁷ J. O'Donnell, *Gamergate and Anti-Feminism in the Digital Age*, cit. p. 71.

¹⁸ Si veda, per esempio, I. Zempi e J. Smith, *Misogyny as Hate Crime* (a cura di), Routledge, Abingdon and New York 2022; K. Barker e O. Jurasz, *Online Misogyny as a Hate Crime. A challenge for Legal Regulation?*, Routledge, Abingdon and New York 2019; L. Richardson-Self, “Woman-Hating: on Misogyny, Sexism, and Hate Speech”, in «Hypatia», vol. XXXIII, 2018, n. 2; Ead., *Hate Speech Against Women Online*, Rowman & Littlefield, London 2021.

¹⁹ J. O'Donnell, *Gamergate and Anti-Feminism in the Digital Age*, cit. p. 92.

²⁰ J. Waldron, *The Harm in Hate Speech*, cit., p. 96.

guardare non solo la creazione nei target di odio di aspettative e di trattamenti degradanti, ma anche il veicolare condizioni per un'aggregazione tra soggetti che condividono la medesima prospettiva di senso esclusivista, nell'utilizzo strumentale della propria identità ascrivibile²¹. Se poi si considera che le rivendicazioni emerse nel *gamergate* sono state appropriate da attivisti del *Men's Rights Movement*, sostenitori, tra le altre cose, dello svantaggio di essere uomini (in particolare bianchi) in società liberal-democratiche, vittime di un femminismo imperante, l'identità ascrivibile maschile acquisisce una maggiore trasparenza. Una possibile problematizzazione non concerne tanto uomini *videogamer* scettici sulla correttezza del giornalismo del settore, ai quali possono essersi aggiunti *troller* disinteressati. Si tratta, piuttosto, di una rivendicazione su base ascrivibile in funzione eliminatoria.

A questo punto, il terzo elemento costitutivo dell'odio si aggiunge alla fossilizzazione e alla generalizzazione: l'eliminazione propria dell'odio nei termini dell'eliminazione del suo oggetto. Come già detto, le associazioni costruite rendono conto di una qualche minaccia. Minaccia ed eliminazione si intersecano: la prima viene imputata al soggetto target e, affinché non venga subito passivamente, deve realizzarsi nella seconda. Se la questione riguarda un *loro* e un *noi*, occorrono dei confini sui quali una violazione va individuata, e la nozione di *straniero interno* di Escobar è qui utile.

Lo *straniero interno* è il soggetto che migra da un altrove e che "oggi viene e domani rimane"²². Per quanto quell'*altrove* possa essere fisico, in termini simbolico-politici l'attenzione si pone esattamente sui confini tra un fuori e un dentro che definiscono un *loro* e un *noi*. Lo *straniero interno* viene analizzato in quanto si fa interno, e non perché lo è: esso si situa in uno spazio simbolico che non è il proprio, portando con sé elementi estranei; "accogliamo la [sua] richiesta [di entrare] moralmente, socialmente, giuridicamente, politicamente, o almeno non la neghiamo con un'espulsione fisica, di nuovo al di là dei confini. [...] E tuttavia [...], è e resta definibile solo in negativo: non-Noi"²³. È esattamente il *noi* che subisce una destabilizzazione, un'apertura al di là di sé stesso.

²¹ J. Waldron, *The Harm in Hate Speech*, cit.

²² R. Escobar, "Rivalità e mimesi. Lo straniero interno", in «Filosofia», vol. VI, 1992, n. 1, p. 95.

²³ R. Escobar, *Metamorfosi della paura*, Il Mulino, Bologna 1997, pp. 162-163.

Da qui, ritengo che non sia considerabile straniero solo chi ha una cittadinanza altra, ma che possa esserlo anche chi può vantare quella cittadinanza dello spazio propriamente statale in cui si trova. La sua non è (solo) una migrazione fisica, nell'addentrarsi lì dove prima non poteva e dove gli erano negate le condizioni di accesso. L'*altrove* da cui lo *straniero interno* si allontana è, dunque, lo status di non-eguale: si tratta di uno sconfinamento dal mancato riconoscimento al riconoscimento dell'autorità politica.

Il caso di *gamergate* può essere letto, a mio avviso, entro una prospettiva à la Escobar. È come se il *videogaming* fosse stato immune alle donne fino al momento in cui esse si sono addentrate nel suo spazio. Supponendo che quest'ultimo venga considerato parte di uno più ampio, come lo *spazio politico* liberal-democratico e quindi di soggetti eguali, le rivendicazioni dei *gamergater* si collegano alla contestazione della titolarità allo status di *eguale* che ha mostrato una notevole capacità pervasiva. L'aspettativa di alcune donne di porsi quali membri di una *community* e di un settore percepiti come esclusivamente maschili viene costruita in termini di elemento estraneo: il *noi* viene destabilizzato in una privazione di sé stesso. Quella che si credeva fosse un'apolitica roccaforte maschile è stata violata. Da qui, movimenti esterni e interni del *noi*, proposti da Escobar, possono innervare l'analisi dell'odio ascrivito e non semplicemente quello individuabile nel *gamergate*. La minaccia percepita fa sì che:

verso l'esterno, il noi sarà sempre più chiuso in confini di pietra: non certo come un darsi e un farsi, ma proprio come un dato e un fatto. La sua identità sarà resa fissa, sottratta a ogni mutamento, in un certo senso naturalizzata. [...] Verso l'interno, poi, la mobilità del noi verrà bloccata, bloccando quella dei singoli. Le loro specifiche differenze, incerte e indebolite, verranno bloccate.²⁴

“Fuori ci sarà il *diverso*, dentro ci sarà il *medesimo*”²⁵. I due movimenti pongono così condizioni di mancato accesso al *noi* e di mancata uscita dal *noi*. Sono questi gli elementi che, a mio avviso, permettono di riflettere sull'aggregazione dell'odio tramite l'eliminazione dei soggetti odiati.

L'eliminazione può essere concettualizzata come una risposta alla viola-

²⁴ R. Escobar, *Paura e libertà*, Morlacchi, Perugia 2009, p. 172.

²⁵ Ivi, p. 173.

zione dei confini che è privazione. Con Ahmed, l'odio "genera il suo oggetto come difesa dal danno", con "la prossimità [dell'altro che] minaccia non solo di togliere qualcosa al soggetto [che odia], ma di prendere il posto del soggetto"²⁶. Da qui, mi sembra che ciò di cui (presumibilmente) si viene privati renda possibile un'aggregazione che è *riaggregazione*, serrando quel confine esterno del *noi* preesistente all'intrusione dell'altro. Che la privazione riguardi l'identità di *videogamer*, posti di lavoro, scuole, quartieri, retaggio culturale, nazione o altro, si tratta sempre di un *noi* che viene meno nella sua certa stabilizzazione. Un qualche disorientamento può coinvolgere i soggetti che odiano. Il confine era, forse, troppo debole, o è l'altro che si caratterizza per essere una forza che non si era stati in grado di prevedere? Cosa certa è che alla privazione occorre rispondere, e sebbene l'eliminazione possa essere fisica, come nel caso delle sparatorie di Christchurch, essa non è risolutoria – il responsabile non ha eliminato l'intero gruppo target, e la presunta minaccia continua a resistere. Solo l'autorità politica dispone dei mezzi per un'espulsione che sia fisica. Da qui, l'eliminazione può essere rivendicata tramite contenuti *user-generated*, in una negazione della giustificabilità del riconoscimento definito dall'autorità politica. È quest'ultima a porre le condizioni della titolarità dei suoi soggetti, strutturando da sé il dilemma della condivisione dello status di eguali. Interpreto ciò anche in relazione a *gamergate*. Pur essendo un caso di nicchia, presenta delle implicazioni politiche ampie: l'intrusione ha portato con sé il germe egualitario; l'intrusione non era solo quella di donne tra uomini.

L'aggregazione realizzabile tra coloro che odiano dalla comune prospettiva di senso necessita così di porsi in termini persuasivi, affinché attraverso la chiusura del *noi* si possa rispondere *insieme* alla mancanza di cui si è vittima. Tuttavia, dalla prospettiva di chi odia, le società liberal-democratiche sembrano veicolare qualcos'altro. Lo status di vittima è esclusivamente proprio di alcuni soggetti dalla specifica identità ascrivibile, e si può notare ciò alla luce della regolamentazione dell'odio nello *hate speech* e negli *hate crime*²⁷. Se tale

²⁶ S. Ahmed, *The Cultural Politics of Emotion*, Edinburgh University Press, Edinburgh 2014, pp. 42-43.

²⁷ Si veda, per esempio, R. Post, "Concluding Thoughts. The legality and Politics of Hatred", in *Hate, Politics, Law*, T. Brudholm e B. Scheepelers Johansen (a cura di), Oxford University Press, Oxford and New York 2018.

regolamentazione non è solo statale ed è anche riscontrabile nei social media, quali aziende private, l'esclusività dello status di vittima si ripresenta. Non solo siamo le vittime di uno sconfinamento e di una privazione che sono causate dall'altro, ma subiamo anche un'indifferenza che è onnipervasiva.

L'eliminazione costitutiva dell'odio deve essere quindi realizzata, in una piena considerazione secondo cui, in società liberal-democratiche, una parallela eliminazione *nei confronti* dell'odio è in corso. Tuttavia, *gamergate* e la sparatoria di Christchurch mostrano come tale seconda eliminazione non sia esattamente riuscita nel suo intento. Contenuti sessisti pubblicati in piattaforme di nicchia si sono riversati in social media generalisti, dove la moderazione è stata mancante e l'aggregazione dei soggetti-utenti è resistita per mesi. Ventiquattro ore, di contro, sono state necessarie per l'eliminazione di contenuti razzisti e islamofobici per un unico social media, rispondendo all'aggregazione realizzata. Si tratta di casi ben specifici, ma che rendono conto delle modalità tramite cui le condizioni di espressione e interconnessione, poste dai social media, possano divenire problematiche, e in termini di aggregazione tramite i contenuti.

Violenza e disorientamento

Non succede quotidianamente che vengano pubblicate coordinate minacce di morte e di stupro, informazioni personali come l'indirizzo di casa, o che un uomo riprenda in contemporanea la sparatoria che sta mettendo in atto, ponendo nelle condizioni di guardarla come se si avessero i suoi occhi. Cosa dire, allora, di quei contenuti che utenti ordinari pubblicano con associazioni ascrivibili e in funzione eliminatoria? Faccio riferimento a quei contenuti diffusi su più social media oppure, se su uno stesso social, in più spazi – in una sezione commenti, su un profilo.

Forse, si tratta di una contaminazione dei social media alla luce di una *content moderation* che viene meno alla sua promessa di invisibilità dei contenuti oggetto di questo lavoro. Qui la moderazione va intesa sia in termini reattivi, ossia di eliminazione dei contenuti perché pubblicati, sia in termini proattivi, come controllo pre-pubblicazione dei contenuti²⁸. Il caso di una

²⁸ Si veda, per esempio, Gorwa R., Binns R. e Katzenbach C., "Algorithmic Content Moderation:

delle sparatorie di Christchurch rende conto di entrambe le forme, come esempio di contenuto violento²⁹: il video era di per sé non pubblicabile in quanto non conforme agli standard della *community*, ma è stato visibile per poi essere rimosso; copie della sparatoria sono state pubblicate, alcune sono state eliminate post-pubblicazione, altre non sono state rese pubblicabili e l'algoritmo utilizzato è stato modellato in modo da sopperire all'errore circa la mancata azione sul video originale. La *content moderation* non si esaurisce, tuttavia, nella valutazione dei contenuti (qualsiasi essi siano), ma nella loro cura, in termini di amplificazione, de-amplificazione, suggerimenti tramite sistemi algoritmici³⁰. Ciò che appare negli schermi dell'utenza, quindi, non è esattamente randomico, con le esperienze digitali di ogni utente che sono sì diversificabili e personalizzabili, ma che possono essere anche ricondotte a comuni interessi e preferenze – alcuni contenuti saranno visibili in misura maggiore o minore ad alcuni soggetti rispetto ad altri³¹.

Intendo restringere l'ipotesi relativa alla contaminazione tramite la considerazione per cui specifici contenuti muovano dalla percezione di una crisi o di un disorientamento. Alla luce di un'incertezza (anche se ci si può chiedere per chi lo sia e come), la contrapposizione *noi-loro* rende individuabili dinamiche conflittuali diffuse? Tale domanda viene posta in quanto i social media definiscono i contenuti d'odio e i contenuti violenti in termini di attacchi. In relazione all'incertezza indagabile, trovare un colpevole come

Technical and Political Challenges in the Automation of Platform Governance”, in «Big Data», vol. VII, 2020, n. 1.

²⁹ La pubblicazione del video era stata preannunciata su *8chan*, piattaforma non generalista: “Well lads, it’s time to stop shit posting and time to make a real-life effort post”; “The Facebook link is below – by the time you read this, I should be going live”; “If I don’t survive the attack, goodbye, god bless, and I will see you all in Valhalla”. Cfr. S. Leitch, “The Making of a ‘made for social media’ massacre”, in S. Leitch e P. Pickering (a cura di), *Rethinking Social Media and Extremism*, cit.

³⁰ Si veda, per esempio, D. Keller, “Amplification and Its Discontents. Why Regulating of Online Content is Hard”, in «Journal of Free Speech», vol. CCXXVII, 2021, n. 1; T. Gillespie, “Do Not Recommend? Reduction as a Form of Content Moderation”, in «Social Media + Society», vol. VIII, 2022, n. 3.

³¹ Si veda, per esempio, R. DiResta, “Algorithms, Affordances, and Agency”, in *Social Media, Freedom of Speech and the Future of Our Democracy*, L.C. Bollinger e G. Stone (a cura di), Oxford University Press, Oxford and New York 2022; C.R. Sunstein, *#Republic: Divided Democracy in the Age of Social Media*, Princeton University Press, Princeton 2017.

soluzione ad essa è un meccanismo che si è ripetuto più volte nelle nostre società e che è ripetibile, trattandosi del capro espiatorio teorizzato da Girard, quale gestione della violenza. Da qui, un'analisi della lettura girardiana della violenza viene proposta, ipotizzando che i social media possano essere assunti come (specifici) aggregati sociali in cui conflitti si danno. Segue così una concettualizzazione di violenza, tramite cui ci si chiede come essa renda conto di contenuti *user-generated*.

Affinché un adattamento alla teoria girardiana della violenza sia oggetto di riflessione, occorre che vi sia una crisi, quale condizione necessaria. Tale crisi va intesa in termini specifici, trattandosi per Girard di una crisi di indifferenziazione, in cui il desiderio è l'elemento da cui muove tutto. Il desiderio viene concettualizzato come mimetico in quanto costitutivamente imitativo: il desiderio umano non né innato né autonomo, desiderando ciò che gli altri desiderano; avremo così (almeno) un soggetto desiderante o discepolo e (almeno) un soggetto modello il cui desiderio viene imitato³². I soggetti girardiani si caratterizzano per relazioni mimetiche innervate da un *doppio imperativo*, quale messaggio che il modello veicola, implicitamente, al discepolo: 'Imitami e non imitarmi'. Il modello deve, quindi, essere imitato per essere e rimanere tale, ma allo stesso tempo l'imitazione del discepolo non deve realizzarsi, in quanto il modello non sarebbe più modello. "Due desideri che convergono sullo stesso oggetto si fanno scambievolmente ostacolo. Qualsiasi *mimesis* che verta sul desiderio va automaticamente a sfociare nel conflitto"³³. Si tratta, così, del desiderio di un soggetto che diventa identico a quello dell'altro nel cosiddetto *conflitto fra doppi*, tale per cui le differenze tra i soggetti vengono meno perché ognuno, nel desiderare, si pone come specchio dell'altro. Stessi desideri, stessa spinta alla soddisfazione del desiderio, ma diversa soddisfazione realizzata. O meglio, affinché un desiderio possa essere soddisfatto, occorre che l'altro soggetto che desidera venga eliminato: l'unica relazione possibile è quella mimetica che è reciprocamente violenta. La soddisfazione del desiderio non può, comunque, essere considerata stabilizzata nella misura in cui alla soddisfazione di oggi può seguire quella di domani. Lo

³² Si veda, R. Girard, *Menzogna romantica e verità romanzesca. Le mediazioni del desiderio nella letteratura e nella vita*, trad. it. di L. Verdi-Vighetti, Bompiani, Milano 2020.

³³ R. Girard, *La violenza e il sacro*, trad. it. di O. Fatiga e E. Czerkl, Adelphi, Milano 1980, p. 205.

scenario mimetico, quindi, si caratterizza per una logica auto-distruttiva.

In una tale erosione delle differenze, i conflitti inter-individuali si accumulano e si diffondono in un'escalation mimetica che contamina l'intera società. La crisi d'indifferenziazione è spazialmente pervasiva, sebbene temporalmente limitata e il meccanismo del capro espiatorio si inserisce esattamente in essa, affinché possa venire meno. Per Alfieri,

Girard ha prodotto una sorta di teoria 'idraulica' della violenza. È come se questa fosse un fluido che scorre in un contenitore chiamato società. Se il livello della violenza sale troppo, tutto ne viene travolto, le stesse pareti del contenitore saltano e c'è solo il disastro, l'impossibilità di sopravvivere. Quindi occorre di tanto in tanto lasciare traboccare la violenza in un punto determinato e in un tempo prefissato comunque breve.³⁴

La violenza viene trasformata da diffusa a polarizzata *solo* su un soggetto o alcuni soggetti, in uno specifico atto violento collettivo. La violenza aggrega, cioè, coloro che prima erano divisi dalla reciprocità dei conflitti inter-individuali.

Qui sembra possibile utilizzare una terminologia presentata nelle pagine precedenti. La violenza collettiva è diretta alle girardiane categorie vittimarie, le quali nel marasma della violenza diffusa (l'accumularsi di violenze interindividuali) hanno mantenuto caratteristiche percepite come diverse, con l'indifferenziazione che è quindi finita. "I segni della differenza diventano segni di colpa e, in quanto orientano la scelta dei persecutori, *segni vittimari*"³⁵: la colpa è del disordine, della crisi sociale, in un'imputazione che è funzionale alla polarizzazione della collettività. Quest'ultima si ritrova unita e si nutre delle associazioni che muovono dalle caratteristiche possedute dalle vittime. Al riguardo, due condizioni vanno realizzate. I soggetti che rientrano nelle categorie vittimarie devono essere marginali alla società, ponendosi *tra* i suoi confini (interni ed esterni). L'atto violento deve essere unanime. La soddisfa-

³⁴ L. Alfieri, "La violenza di massa in Elias Canetti e René Girard", in «Teoria politica», 2017, n. 7, p. 302.

³⁵ L. Alfieri, "Dal conflitto dei doppi alla giudiziaria", in *Figure e simboli dell'ordine violento. Percorsi fra antropologia e filosofia politica*, L. Alfieri, C.M. Bellei e D.S. Scalzo (a cura di), Giappichelli, Torino 2003, p. 21.

zione delle due condizioni può realizzarsi solo tramite una loro intersezione: una collettività si muove come un tutt'uno, mancante di defezioni, esattamente contro chi non è pienamente esterno e interno; l'unanimità protegge la collettività da una violenza che potrebbe sorgere dalla vendicabilità³⁶ delle categorie vittimarie (come continuazione della violenza diffusa), ed è tale rischio ad essere eluso grazie alla posizione ambigua dei soggetti che vanno eliminati perché sono facilmente eliminabili. Solo se il colpevole del disordine viene individuato e immolato tramite un unico atto violento che aggrega ciò che era prima disgregato, l'ordine può essere ristabilito. La vittima è sì colpevole, ma anche salvifica. La violenza polarizzata su di essa ferma la violenza diffusa: la violenza elimina sé stessa.

Ogni crisi è tale perché finita, ma ciò non implica che non si ripresenti. Il meccanismo del capro espiatorio permette, per Girard, una lettura dell'umanità e della fondazione della società fino ad arrivare a noi. Ad ogni crisi segue un ordine, e ogni ordine può degenerare in crisi. Per quanto possa trattarsi di un circolo vizioso, forse ci si può affidare al meccanismo del capro espiatorio, con la “vittima espiatoria” [che] finirà necessariamente per apparire e riconciliare la comunità, poiché la stessa esasperazione della crisi, legata a un mimetismo sempre accresciuto, deve necessariamente suscitarsela”³⁷.

Relativamente ai social media, nelle regole e negli standard della *community* è individuabile un impegno di regolamentazione di contenuti che, come prima accennato, vengono assunti in termini di attacchi contro soggetti in base alle loro caratteristiche ascrivibili, rispetto a qualunque altra caratteristica. Se si guarda a Facebook sono presenti degli esempi che possono orientare l'utenza circa la comprensione della inammissibilità di contenuti dalle associazioni ascrivibili: “persone di colore e scimmie o creature simili alle scimmie, ebrei e ratti, musulmani e maiali, messicani e vermi”; “persone di credo islamico che fanno sesso con capre o maiali”; “le dichiarazioni secondo cui gli ebrei controllano le istituzioni finanziarie, politiche o mediatiche; i riferimenti ai dalit come lavoratori di bassa estrazione e il paragone delle persone nere alle at-

³⁶ Cfr. F. Sciacca, *Il potere della vendetta. Quattro lezioni*, AlboVersorio, Milano 2018.

³⁷ R. Girard, *Delle cose nascoste sin dalla fondazione del mondo. Ricerche con Jean-Michel Oughourlian e Guy Lefort*, trad. it. di R. Damiani, Adelphi, Milano 1983, p. 43.

trezzature agricole”³⁸. Questi sono gli esempi che si leggono nella categoria “Comportamenti di incitamento all’odio” in cui vengono menzionate specifiche identità ascrivibili, indicando anche la loro natura non esaustiva. Nella categoria “Violenza e istigazione alla violenza” è anche presente il riferimento ad un’intersezione tra minacce di violenza, probabili lesioni (di gravità variabile) e la protezione di soggetti individuabili su base ascrivibile (“persone o gruppi di persone prese di mira in base alle loro caratteristiche protette”³⁹). In questo caso e rispetto al primo, gli esempi non vengono forniti. Certamente si tratta di paradigmatici target d’odio, e alcuni di essi coincidono con le categorie vittimarie girardiane. Come già detto, occorre una crisi e se quelli che vengono considerati attacchi (“in forma scritta o visiva”) sono visibili nei social media, segue una problematizzazione circa la loro capacità di ristabilire l’ordine. Come suggerisce Bellei, ci si può chiedere se la crisi sia strutturale, non finita e quindi qualcosa che non viene meno a seguito dell’immolazione di capri espiatori⁴⁰.

“Ogni frase, ogni ingiuria diventa ferita visibile, segno tangibile di una possibile sconfitta, privazione del proprio sé che implica una rivalsa in grado di riequilibrare ciò che è andato perso. La prospettiva temporale svanisce, assorbita da urgenze che si impossessano delle persone trasformandole in macchine odianti”⁴¹: questo è quanto Bellei afferma su alcuni contenuti che sono visibili nei social media. Se si tratta di contenuti d’odio, per come sono stati analizzati in questo lavoro, si presentano quali fossilizzanti, generalizzanti ed eliminatori. Si può anche parlare di una dinamica violenta *à la* Girard? In una prospettiva girardiana che innerva la citazione appena presentata, una mancata soddisfazione stabilizzata del desiderio mimetico coinvolge più utenti, con la vittima da immolare che li unisce in collettività. Tuttavia, all’immolazione di oggi segue quella di domani, con costanti capri espiatori: la dimensione

³⁸ <<https://transparency.meta.com/it-it/policies/community-standards/hate-speech/>> (consultato il 25/03/25).

³⁹ <<https://transparency.meta.com/it-it/policies/community-standards/violence-incitement/>> (consultato il 25/03/25).

⁴⁰ Cfr. C.M. Bellei, “Il sacrificio impossibile. I social network come luoghi di paura”, in *Buoni e cattivi. Etica, politica e potere al tempo di Internet*, F. Sciacca (a cura di), Mimesis, Milano-Udine 2022.

⁴¹ Ivi, p. 140.

temporalmente finita della crisi e un'unica collettività violenta non si presentano; ripetute eliminazioni sono visibili come “una necessità che produce dipendenza”⁴², con più collettività che si formano e si riformano non unanimemente. Al riguardo, la possibilità di defezioni interne fa sì che una certa utenza non sia coinvolta nell'eliminazione, tuttavia ciò non implica il venir meno della crisi. Dinamiche conflittuali (quindi violente) resistono, non ponendo le condizioni per quella rassicurazione girardiana secondo cui la violenza si esaurisce da sé.

In un tale disorientamento, individuo un'ulteriore dinamica che non ha base ascrivibile, ma da essa muove. Faccio riferimento a quelle circostanze in cui una certa utenza contesta in funzione eliminatoria sia i contenuti di odio e violenti (entrambi ascrivibili) altrui, sia chi li pubblica. Entro un social media ognuno vanta l'*eguale* status di utente, ma è tale condivisione a divenire problematica per alcuni soggetti – contestazione che si interseca, ma non è coincidente, con quella dello *spazio politico* affrontata precedentemente. Ci si può, quindi, chiedere se la contestazione esclusivista sia un elemento positivo nonostante i social media vengano presentati come spazi digitali inclusivi: se lo sguardo è posto su di essi, e in ogni aggregato sociale inclusione ed esclusione si necessitano vicendevolmente, qual è il *noi* e il *loro*?

Generalissimi *noi* e *loro* sono quelli che sottoscrivono i servizi erogati da un certo social media e coloro che non lo fanno. Una volta che si diventa utenti, la *content moderation* segna un confine tra ciò che è ammissibile e ciò che non è ammissibile pubblicare, con l'utenza che può avere i propri contenuti visibili, non visibili o eliminati, tale per cui le condizioni di espressione e interconnessione sono rese fruibili o negate. Tuttavia, la regolamentazione è mutevole in quanto gli standard e le regole della *community* sono oggetto di frequenti aggiornamenti. Con Gillespie, “le piattaforme si trovano di fronte a un'arma a doppio taglio: troppa poca moderazione e l'utenza potrebbe andarsene per evitare l'ambiente tossico che ha preso piede; troppa moderazione e l'utenza potrebbe comunque andarsene, rifiutando le piattaforme perché troppo invadenti o troppo antisettiche”⁴³. Nel primo caso, l'utenza è posta

⁴² Ivi, p. 143.

⁴³ T. Gillespie, “Regulation of and by Platforms”, in *The SAGE Handbook of Social Media*, J. Burgess, A. Marwick e T. Poell (a cura di), SAGE, Los Angeles 2018, pp. 262-263.

nelle condizioni di migrare verso social media più attenti, e nel secondo caso, verso altri più permissivisti.

Propongo, a questo punto, un duplice scenario in cui l'utenza si divide nella percezione che ha della *content moderation*. Se contenuti d'odio e contenuti violenti sono potenzialmente visibili a chiunque, ma solo alcuni li ritengono tali, i soggetti sono posti nelle condizioni (opzionali) di segnalarli, affinché su di essi sia attivato un controllo da parte dei team di moderazione (tramite una valutazione algoritmica e/o umana). La moderazione non viene, così, considerata esattamente attenta. Che la segnalazione riesca o meno (che il contenuto venga rimosso o mantenuto), quell'utenza può comunque pubblicare contenuti tramite i quali una contestazione si realizza: dei contenuti d'odio e violenti e dell'altrui *eguale* status di utente. Se, di contro, quegli stessi contenuti non vengono percepiti come di odio e violenti da una certa utenza, ma come mero esercizio di libertà di espressione, i social media hanno pienamente rispettato la promessa assunta come uno dei loro impegni fondamentali. I contenuti in questione sono considerati accettabili e normalizzati. Anche in questo caso e data la loro visibilità, è possibile che vi sia una contestazione.

In una lettura incrociata, si tratta di un *noi* e un *loro*, entrambi interni ai social media, con la diversa percezione della *content moderation* che disgrega l'utenza per poi aggregarla su quelli che mi sembrano essere fronti antagonistici. L'opposizione muove dalle aspettative di ciò che dovrebbe essere pubblicabile e non pubblicabile. Visibilità e invisibilità si pongono così in termini di soddisfazione e insoddisfazione dalle proprie esperienze digitali di espressione e interconnessione. Che utenti si aggregino deliberatamente o meno, sia i contenuti d'odio e contenuti violenti, sia i contenuti di critica e di indignazione nei loro confronti generano flussi di dati e quindi profitti, con le dinamiche conflittuali che sembrano nutrire gli interessi economici dei social media.

Forse l'aggregazione tramite i primi contenuti può essere problematizzata come più pressante rispetto alla seconda, e rispetto a qualsiasi altra aggregazione che non ha uno scopo eliminatorio su base ascrivibile. Un'ipotesi è stata proposta, ed è quella secondo cui l'odio ascrivibile e la violenza girardiana aggregano i soggetti-utenti. A questo punto, mi chiedo se tali due concettualizzazioni vanno assunte sul medesimo piano analitico.

Nell'odio ascrivito e nella violenza girardiana, target e categorie vittimarie coincidono, in quanto soggetti che vanno eliminati. La logica di individuazione del soggetto è la stessa: le caratteristiche ascrivite sono indicatori della giustificabilità dell'eliminazione tramite le (multiformi) associazioni che si pongono come promemoria della minaccia a cui occorre rispondere. Tuttavia, ritengo che le due eliminazioni si diramino nell'aggregazione che realizzano. L'odio ascrivito aggrega per disgregare, la violenza girardiana (se riesce) aggrega per fare venire meno la disgregazione. O meglio, l'odio aggrega un *ridefinito* e *ristretto noi*, affinché il *noi* inclusivo dei target venga frammentato. Di contro, la violenza aggrega per *ricomporre* un *noi* frammentato e porlo come un tutt'uno contro la vittima da immolare. Le due eliminazioni non coincidono.

Ciò che rimane nella sua solidità è il gruppo a cui il target e la vittima vengono ricondotti, trattandosi di una fossilizzazione che ipotizzo debba rimanere tale affinché un orientamento di senso accompagni chi si dà all'odio e alla violenza. Se la titolarità allo status di *eguale* del soggetto target-vittima si pone da sé quale elemento perturbatore, le associazioni si strutturano in termini di tentativo di rispondere all'incertezza che preme.

Ritrovo la rilevanza dello status di *eguale* anche in relazione ai social media, in quanto l'utenza non è solo utenza che sottoscrive i termini del servizio erogato. Essa è innanzitutto soggetto politico, titolare di un diritto di libertà di espressione definito dall'autorità politica. Ed è questa libertà che permette all'utenza di esprimersi tramite contenuti, altrimenti un *geo-blocking* graverebbe sull'accesso ai social media – cosa che accade, anche se limitatamente. La questione della contestazione di quel riconoscimento eccessivamente inclusivo dello status di *eguale* da parte dell'autorità politica si insinua e si esacerba, in quanto la minaccia del *loro* diventa sempre più patente: loro sono *anche* utenti, i loro contenuti sono *anche* a me visibili, loro possono *anche* connettersi con me. È come se la distanza non fisica si ponesse come vicinanza invadente.

Per quanto le eliminazioni dell'odio ascrivito e della violenza girardiana non coincidano, o meglio questa è la mia ipotesi, esse si ricongiungono tramite *content moderation*. Alla luce di quanto detto, ritengo che i social media intendano impegnarsi in un'eliminazione delle eliminazioni proprie

dell'odio e della violenza indagate. La prima eliminazione mi sembra essere specificatamente analizzabile quale eliminazione delle aggregazioni, in un tentativo di neutralizzare la diffusione dei contenuti in termini di contaminazione. Da qui, è la dimensione cumulativa dei contenuti che si presenta come un elemento di preoccupazione: per i social media, per l'utenza e per l'autorità politica (con i suoi tentativi di regolamentare la relazione tra piattaforme social e soggetti politici-utenti⁴⁴).

Se le associazioni su base ascriviva sono un affidabile promemoria analitico dell'attribuzione di una minaccia variamente definibile e giustificabile, contenuti *user-generated* indicano cosa viene detto ai e dei soggetti de-individualizzati, e come la realtà di tali soggetti può essere costruita tramite essi. Ipotizzare una mancanza della *content moderation* permette di problematizzare la visibilità di contenuti d'odio e contenuti violenti come condizione di provvisorietà. Questi potranno essere eliminati, emergere, o ancora resistere perché sfuggiti all'occhio non esattamente onnipervasivo della moderazione, implicando, dunque, un disorientamento di senso dell'utenza.

⁴⁴ Si veda, per esempio, J. Bayer *et al.*, *Perspectives on Platform Regulation. Concepts and Models of Social Media Governance Across the Globe*, Nomos, Baden-Baden 2021; R. Gorwa, *The Politics of Platforms Regulation. How Governments Shape Online Content Moderation*, Oxford University Press, Oxford and New York 2024.

BIBLIOGRAFIA

- AHMED S., *The Cultural Politics of Emotion*, Edinburgh University Press, Edinburgh 2014.
- ALFIERI L., «La violenza di massa in Elias Canetti e René Girard», in «Teoria politica», 2017, n.7.
- , «Il terzo che deve morire», in *Figure e simboli dell'ordine violento. Percorsi fra antropologia e filosofia politica*, Luigi Alfieri, Cristiano Maria Bellei e Domenico Sergio Scalzo (a cura di), Giappichelli, Torino 2003.
- , «Dal conflitto dei doppi alla giudiziaria», in *Figure e simboli dell'ordine violento. Percorsi fra antropologia e filosofia politica*, Luigi Alfieri, Cristiano Maria Bellei e Domenico Sergio Scalzo (a cura di), Giappichelli, Torino 2003.
- BARKER K. e JURASZ O., *Online Misogyny as a Hate Crime. A Challenge for Legal regulation?*, Routledge, Abingdon and New York 2019.
- BAYER J. et al., *Perspectives on Platform Regulation. Concepts and Models of Social Media Governance Across the Globe*, Nomos, Baden-Baden 2021.
- BELLEI C.M., «Il sacrificio impossibile. I social network come luoghi di paura», in *Buoni e cattivi. Etica, politica e potere al tempo di Internet*, F. Sciacca (a cura di), Mimesis, Milano-Udine 2022.
- BROWN A., «The 'who?' question in the hate speech debate: Part 1: Consistency, practical and formal approaches», in «Canadian Journal of Law & Jurisprudence», vol. XXIX, 2016, n. 2.
- BRUDHOLM T. e SCHEPELERN JOHANSEN B. (a cura di), *Hate, Politics, Law*, Oxford University Press, Oxford and New York 2018.
- , «Pondering Hatred», in *Emotions and Mass Atrocity. Philosophical and Theoretical Explorations*, T. Brudholm e J. Lang (a cura di), Cambridge University Press, Cambridge 2018.
- CHESS S. e SHAW A., «A Conspiracy of Fishes, or, How We Learned to Stop Worrying about #Gamergate and Embrace Hegemonic Maculinity», in «Journal of Broadcasting & Electronic Media», vol. LIX, 2015, n. 1.
- COHEN-ALMAGOR R. *Confronting the Internet Dark Side. Moral and Social Responsibilities on the Free Highway*, Cambridge University Press,

Cambridge 2015.

- DIRESTA R., "Algorithms, Affordances, and Agency", in *Social Media, Freedom of Speech and the Future of Our Democracy*, L.C. Bollinger e G. Stone (a cura di), Oxford University Press, Oxford and New York 2022.
- DOUEK E., "Content Moderation as Systems Thinking", in «Harvard Law Review», vol. CXXXVI, 2022, n. 2.
- EDELSON L., "Content Moderation in Practice", in *Media and Society After Technological Disruption*, K. Langvardt e J. Hurwitz (a cura di), Cambridge University Press, Cambridge 2024.
- ESCOBAR R., "Rivalità e mimesi. Lo straniero interno", in «Filosofia», vol. VI, 1992, n. 1.
- , *Metamorfosi della paura*, Il Mulino, Bologna 1997.
- , *Paura e libertà*, Morlacchi, Perugia 2009.
- FOX J. e TANG W.Y. "Sexism in Online Video Games: The Role of Conformity to Masculine and Social Dominance Orientation", in «Computer in Human Behavior», 2014, n. 33.
- GILLESPIE T. *et al.*, "Expanding the debate about content moderation: scholarly research agendas for the coming policy debates", in «Internet Policy Review», vol. IX, 2020, n. 4.
- GILLESPIE T., "Regulation of and by Platforms", in *The SAGE Handbook of Social Media*, J. Burgess, A. Marwick e T. Poell (a cura di), SAGE, Los Angeles 2018.
- , *Custodians of the Internet. Platforms, Content Moderation and the Hidden Decisions that Shape Social Media*, Yale University Press, New Haven and London 2018.
- , "Do Not Recommend? Reduction as a Form of Content Moderation", in «Social Media + Society», vol. VIII, 2022, n. 3.
- GIRARD R., *Delle cose nascoste sin dalla fondazione del mondo. Ricerche con Jean-Michel Oughourlian e Guy Lefort*, trad. it. di R. Damiani, Adelphi, Milano 1983.
- , *La violenza e il sacro*, R. Girard, trad. it. di O. Fatica e E. Czerkl, Adelphi, Milano 2003.
- , *Menzogna romantica e verità romanzesca. Le mediazioni del desiderio nella letteratura e nella vita*, trad. it. L. Verdi-Vighetti, Bompiani, Milano 2020.
- GORWA R., BINNS R. e KATZENBACH C. "Algorithmic Content Moderation: Tech-

- nical and Political Challenges in the Automation of Platform Governance”, in «Big Data», vol. VII, 2020, n. 1.
- GORWA R., *The Politics of Platforms Regulation. How governments Shape Online Content Moderation*, Oxford University Press, Oxford and New York 2024.
- HARWOOD M., “Living Death: Imagined History and the Tarrant Manifesto”, in «Emotions: History, Culture, Society», vol. V, 2021, n.1.
- KELLER D., “Amplification and Its Discontents. Why Regulating of Online Content is Hard”, in «Journal of Free Speech», vol. CCXXVII, 2021, n. 1.
- LEITCH S. e PICKERING P. (a cura di), *Rethinking Social Media and Extremism*, ANU Press, Canberra 2022.
- LEITCH S., “The Making of a ‘made for social media’ massacre”, in *Rethinking Social Media and Extremism*, S. Leitch e P. Pickering (a cura di), ANU Press, Canberra 2022.
- LYNCH T. *et al.*, “Sexy, Strong and Secondary: a Content Analysis of Female Characters in Video Games across 31 Years”, in «Journal of Communication», vol. LXVI, 2016, n. 4.
- LUMSDEN K. e HARMER E. (a cura di), *Online Othering. Exploring Digital Violence and Discrimination on the Web*, Palgrave Macmillan, Cham 2019.
- MACKLIN G., “The Christchurch Attacks: Livestream Terror in the Viral Video Age”, in «CTC Sentinel», vol. XII, 2019, n. 6.
- MASSANARI A.L., *Gaming Democracy. How Silicon Valley Leveled Up the Far Right*, MIT Press, Cambridge (MA) 2024.
- MORTENSEN T.E., “Anger, Fear, and Games: The Long Event of Gamergate”, in «Games and Culture», vol. XIII, 2018, n. 8.
- POST R., “Concluding Thoughts. The legality and Politics of Hatred”, in *Hate, Politics, Law*, T. Brudholm e B. Scheepelern Johansen (a cura di), Oxford University Press, Oxford and New York 2018.
- RICHARDSON-SELF L., “Woman-Hating: on Misogyny, Sexism, and Hate Speech”, in «Hypatia», vol. XXXIII, 2018, n. 2.
- , *Hate Speech Against Women Online*, Rowman & Littlefield, London 2021.
- SCIACCA F., *Il potere della vendetta. Quattro lezioni*, AlboVersorio, Milano 2018.
- SUNSTEIN C.R., *#Republic: Divided Democracy in the Age of Social Media*, Princeton University Press, Princeton 2017.

- SZANTO T., "In Hate We Trust: The Collectivization and Habitualization of Hatred", in «Phenomenology and the Cognitive Sciences», vol. XIX, 2020, n. 3.
- WALDRON J., *The Harm in Hate Speech*, Harvard University Press, Cambridge (MA) 2012.
- ZEMPI I. e SMITH J. (a cura di), *Misogyny as hate crime*, Routledge, Abingdon and New York 2022.

SITOGRAFIA

<<https://transparency.meta.com/it-it/policies/community-standards/hate-speech/>> (consultato il 25/03/25).

<<https://transparency.meta.com/it-it/policies/community-standards/violence-incitement/>> (consultato il 25/03/25).

<<https://about.fb.com/news/2019/03/update-on-new-zealand/>> (consultato il 25/03/25).

<<https://www.theesa.com/resources/essential-facts-about-the-us-video-game-industry/>> (consultato il 25/03/25).

