

Il contrasto dell'odio online: possibili rimedi

Giovanni Ziccardi¹

1. Alcune considerazioni introduttive

Disegnare una strategia di contrasto per uno dei fenomeni, l'odio online, che sta caratterizzando larga parte del dibattito attorno alla diffusione di contenuti sulle piattaforme di social network e che si presenta mutevole, liquido e complesso, non è affatto semplice.

Le uniche tre armi che gli Stati hanno – l'educazione degli utenti/cittadini, un uso efficace dello strumento giuridico e l'utilizzo della tecnologia stessa per mappare e bloccare contenuti d'odio in maniera automatizzata – per cercare di contenere un fenomeno così complesso e dilagante, non sono di facile gestione.

La prima, l'educazione alla legalità, alla controparola, alla discussione pacata, sta attraversando un momento di grandissima crisi dovuto all'utilizzo massiccio di odio da due degli ambiti della società più influenti e che dovrebbero dare un buon esempio: i grandi media e il mondo della politica. In entrambi i settori si è scoperto che il far circolare odio porta lettori, consensi, elettori, e in tutti gli ambiti non si disdegna l'utilizzo dell'odio come *valuta* causando anche un aumento sensibile nella tolleranza alle espressioni d'odio in capo ai cittadini, che si sono ormai abituati a determinati toni.

¹ Professore di Informatica Giuridica, Università degli Studi di Milano. Coordinatore Scientifico del Centro di Ricerca Coordinato in "Information Society Law" (ISLC) dell'Università degli Studi di Milano.

La seconda, un uso del diritto che aiuti a contenere l'odio, richiede un approccio chirurgico per non toccare il diritto di libertà di manifestazione del pensiero e per non intaccare lo stesso ecosistema tecnologico.

La terza arma, un'automazione, tramite algoritmi e analisi semantica, che aiuterebbe le tecnologie stesse a individuare e bloccare l'odio, si scontra con la difficoltà di analisi del linguaggio umano e la sua complessità e con l'impossibilità di prevedere metodi che censurino espressioni legittime.

Un costante equilibrio di queste tre forze è il metodo più efficace per contrastare l'odio online. Ma, prima, è indispensabile comprendere il fenomeno.

2. La natura delle espressioni d'odio online e la loro evoluzione

Se volessimo disegnare il quadro delle espressioni d'odio che ora circolano online, al vertice di un'immaginaria piramide si potrebbero collocare le espressioni e le azioni d'odio così come sono disciplinate a livello internazionale, politico e normativo, spesso con accordi tra Stati.

Di solito, la necessità di regolamentare tali fenomeni si basa sul *ricordo* e sulla *memoria* o, al contrario, è dettata dalla più stringente attualità. Il negazionismo², l'odio politico, l'apologia di regimi, la discriminazione etnica o basata sulle abitudini sessuali, l'odio religioso e razziale, l'attacco al *diverso* in ogni sua accezione, la propaganda terroristica³ sono i temi che vengono affrontati e che variano a seconda della tradizione del Paese che legifera.

A seguire, vi sono manifestazioni considerate d'odio perché offensive in base al diritto locale. Questo è un aspetto altrettanto interessante: ogni singolo Stato può criminalizzare esternazioni che invece, nella tradizione giurisprudenziale sovranazionale, si sostiene debbano essere protette.

A un terzo livello vi sono le discussioni legittime ma *accese*: ciò che è dialogo infuocato o inappropriato, ma che il diritto ritiene ammissibile. Anche in questo caso, il confine è sempre stato, nel corso della storia, molto labile, spesso

² Per un'introduzione al tema del negazionismo vedi: F.R. Recchia Luciani, L. Patruno (a cura di), *Opporsi al negazionismo – Un dibattito necessario tra filosofi, giuristi e storici*, Il melangolo, Genova 2013; D. Di Cesare, *Se Auschwitz è nulla – Contro il negazionismo*, Il melangolo, Genova 2012; C. Vercelli, *Il negazionismo – Storia di una menzogna*, Laterza, Roma-Bari 2013; V. Pisanty, *L'irritante questione delle camere a gas – Logica del negazionismo*, Bompiani, Milano 2014. Da un punto di vista giuridico vedi P. Lobba, *Il negazionismo come abuso della libertà di espressione: la giurisprudenza della Corte di Strasburgo*, in «Rivista Italiana di Diritto e Procedura Penale», 4, 2014, pp. 1815-1853.

³ Con riferimento alla presenza del terrorismo in rete, sia nei discorsi d'odio sia nelle attività di propaganda, radicalizzazione e organizzazione di attentati, si vedano: D. Tosini, *Internet e violenza politica nel XXI secolo*, in «Equilibri», XII (2008), 2, pp. 193-206; Id., *La lunga guerra contro Al-Qaeda*, in «Nuova Informazione Bibliografica», III (2006), 4, pp. 667-682; Id., *Rischio politico e rischio giuridico – Terrorismo e antiterrorismo nelle democrazie contemporanee*, in «Equilibri», IX (2005), 3, pp. 601-622; F. Fabbrini, *Lotta al terrorismo: da Bush a Obama, passando per la Corte Suprema*, «Quaderni costituzionali», XXXI (2011), 1, pp. 89-103.

condizionato da scriminanti o attenuanti concesse da decisioni giurisprudenziali (correlate, ad esempio, ai dibattiti politici o al diritto di cronaca e al giornalismo d'inchiesta).

Infine vi sono tutte le espressioni che sono ammesse perché non sono né d'odio né comunemente considerate, una volta effettuata l'analisi del contesto, offensive.

Non esiste, a oggi, una nozione condivisa di cosa si debba intendere per "hate speech"; molti Stati stabiliscono i confini di loro iniziativa, e non sulla base di principi uniformi⁴. Le differenze possono essere sensibili spostandosi da un ordinamento giuridico all'altro. Si può notare, tuttavia, un'evoluzione circa i contenuti di questa definizione, evoluzione che è bene illustrare per tracciare, sin dall'inizio, un quadro più preciso del contesto che si andrà ad analizzare.

Una nozione che può essere presa a riferimento è contenuta nella Convenzione (o Patto) Internazionale sui Diritti Civili e Politici, un trattato che nasce dall'esperienza della Dichiarazione Universale dei Diritti dell'Uomo, adottato nel 1966 ed entrato in vigore nel 1976. L'articolo 20 del Patto recita: «1. Qualsiasi propaganda a favore della guerra deve essere vietata dalla legge. 2. Qualsiasi appello all'odio nazionale, razziale o religioso che costituisca incitamento alla discriminazione, all'ostilità o alla violenza deve essere vietato dalla legge».

Una prima definizione di *hate speech*, quindi, sembra contenere i seguenti sei elementi centrali, divisibili in due gruppi: tre "ambiti" (i nazionalismi, il razzismo⁵ e la religione quali strumenti di discriminazione e di odio) e l'incitamento a tre "comportamenti" (discriminazione, ostilità e violenza). Si noti il termine *incitamento* che solleverà ampi dibattiti, soprattutto con riferimento alla reale idoneità a portare imminente violenza nel caso concreto.

⁴ Con riferimento al "tipico" approccio nordamericano e alla sua origine su questi temi si vedano: G. Ziccardi, *La Corte Suprema americana e la libertà di espressione in Internet*, in «Quaderni Costituzionali», 1 (1988), pp. 123-134; J. Mailland, *The Blues Brothers and the american constitutional protection of hate speech: teaching the meaning of the First Amendment to foreign audiences*, in «Michigan State International Law Review», 21 (2013), 2, pp. 451-468; R. Kiska, *Hate speech: a comparison between the European Court of Human Rights and the United States Supreme Court jurisprudence*, in «Regent University Law Review», 25 (2012), pp. 107-151; Robert A. Kahn, *Why do Europeans ban hate speech? A debate between Karl Loewenstein and Robert Post*, in «Hofstra Law Review», 41 (2012), pp. 545-585; L. Goisis, *Libertà d'espressione e odio omofobico. La Corte Europea dei Diritti dell'Uomo equipara la discriminazione in base all'orientamento sessuale alla discriminazione razziale*, in «Rivista Italiana di Diritto e Procedura Penale», I (2013), pp. 418-441; C.E. Haupt, *Regulating hate speech – Damned if you do and damned if you don't: lessons learned from comparing the German and U.S. approaches*, in «Boston University International Law Journal», 23 (2005), pp. 299-335.

⁵ Con riferimento a un'analisi accurata del rapporto tra odio politico e razzismo, si veda G. Pino, *Discorso razzista e libertà di manifestazione del pensiero*, in «Politica del Diritto», XXXIX (2008), 2, pp. 287-305. Con riferimento, invece, alla natura della parola d'odio e la sua evoluzione sociologica e storica, si veda R. Abel, *La parola e il rispetto*, Giuffrè, Milano 1996.

Una seconda fonte di grande interesse definitorio è la Raccomandazione n. 20 del 1997 del Consiglio d'Europa sull'*hate speech*⁶. Il termine, secondo gli estensori, deve essere interpretato come idoneo a comprendere tutte quelle forme espressive⁷ che diffondono, incitano, promuovono o giustificano odio razziale, xenofobia, antisemitismo o altre forme di odio basate sull'intolleranza, compresa quella espressa da nazionalismo aggressivo ed etnocentrismo, la discriminazione e l'ostilità contro le minoranze, i migranti e le persone di origine straniera. La definizione è molto più specifica e, si consenta, più *moderna*: si parla di *giustificazione* dell'odio (per tenere sotto controllo il pericolo di una tolleranza diffusa di tali espressioni) e si cita il problema quanto mai attuale dei migranti, della xenofobia, dell'antisemitismo e delle minoranze.

Anche la risalente Convenzione internazionale sull'eliminazione di ogni forma di discriminazione razziale (ICERD) del 1965 obbliga i Governi a condannare ed eliminare atti di discriminazione razziale, sia nell'ambito delle istituzioni pubbliche, sia in ambito privato.

Le definizioni contenute in questo documento sono molte, e assai interessanti. Si citano le dottrine di superiorità fondate sulla distinzione tra le razze (dichiarandole false scientificamente, condannabili moralmente, ingiuste e pericolose socialmente), la discriminazione tra esseri umani fondata sul colore, sulla razza, sull'origine etnica, le politiche di apartheid e di segregazione o di separazione.

Volendo fare un primo sunto, però, appare abbastanza chiaro come, in quasi tutte le situazioni interpretative, siano necessari almeno *tre requisiti* affinché un'espressione si possa considerare *hate speech*:

- i) una chiara volontà e intenzione di incitare odio (*hate*) con la parola o ogni altro mezzo di comunicazione;
- ii) oltre alla volontà, che vi sia l'incitamento vero e proprio, che sia idoneo a causare atti d'odio e di violenza nei confronti dei soggetti presi di mira;
- iii) che gli atti di violenza o di discriminazione si verifichino, o che il rischio che ciò avvenga sia imminente.

Ciò comporta, in linea di principio, che se le affermazioni, anche le più crude, *non* sono portate con l'intento di incitare terzi all'odio, non rientrano nell'ambito dell'*hate speech*.

Perché ci sia *hate speech* è fondamentale che l'incitamento all'odio o alla violenza sia connesso in maniera molto stretta all'idea di danno, discriminazione o

⁶ Circa le origini del termine anglosassone "hate speech" e i suoi rapporti con la tradizione di odio politico europea si veda J. Waldron, *The harm in hate speech*, Harvard University Press, Cambridge 2012. Circa il rapporto tra *hate speech* e dignità di tutti i cittadini si veda J. Waldron, *Dignity and defamation: the visibility of hate*, in «Harvard Law Review», 123 (2010), pp. 1597-1657. Con riferimento ai primi tentativi di normare l'*hate speech* si veda J. Mchangama, *The problem with hate speech laws*, in «The Review of Faith & International Affairs», 13 (2015), 1, pp. 75-82.

⁷ Con riferimento ai limiti della libertà di manifestazione del pensiero, si veda V. Zeno-Zencovich, *La libertà d'espressione*, Il Mulino, Bologna 2004.

violenza (anche se poi, in concreto, l'atto violento non viene portato a compimento).

Premesso che, come si notava poco sopra, la percezione di cosa sia *hate speech* è, a volte, soggettiva e gli effetti delle espressioni d'odio sono diversi per ogni persona, i medici e gli psicologi hanno cercato, nel corso degli anni, di identificare le *patologie* che conseguono a tali azioni, dal momento che viene sovente sottostimato l'impatto dell'*hate speech* sulla vittima.

Tra queste, le più comuni sono la perdita di autostima, un senso di rabbia, un isolamento forzato, un costante immotivato atteggiamento sulla difensiva, uno stato di shock, di confusione o di disgusto, sino a configurare una vera e propria esperienza traumatica e patologica sul breve e lungo periodo.

Gordon Allport, noto psicologo sociale statunitense, elaborò nel 1954 una scala del pregiudizio e della discriminazione, nel tentativo di "misurare" l'odio nella società (*Allport's Scale of Prejudice and Discrimination*). La scala va da uno a cinque, con aumento della gravità dei comportamenti.

1) *Antilocution*. Si tratta del comportamento più lieve, nel quale i componenti di un gruppo più forte, di maggioranza, dileggiano, burlano, tratteggiano con stereotipi negativi o immagini non veritiere, spesso in maniera colorita o enfatizzandone i difetti, un gruppo di minoranza. È già considerato frasario ingiurioso, o discriminatorio, ma non preoccupa per i toni in sé ma per il potenziale di odio che potrebbe, in una escalation successiva, generare.

2) *Avoidance*. In questa fase, e come secondo livello di gravità, le persone del gruppo di maggioranza *evitano* volontariamente i componenti di un gruppo di minoranza. Il danno arrecato, in questo caso, prende la forma dell'*isolamento* delle persone ed è manifestazione chiara, ad esempio, di xenofobia o di odio nei confronti del diverso o dello straniero che si finalizza con l'*esclusione*.

3) *Discrimination*. Nel terzo livello un gruppo (solitamente di maggioranza) discrimina un altro (solitamente di minoranza) negando ai componenti di quel gruppo l'accesso in condizioni di eguaglianza alle opportunità, ai beni e ai servizi della società in cui tutti e due i gruppi vivono. Di solito s'impedisce al gruppo preso di mira di potersi garantire un'istruzione come tutti, o di poter ottenere posizioni lavorative di prestigio o, comunque, senza discriminazioni. Il pregiudizio, in questa fase, è azionato anche con leggi o regolamenti a ciò finalizzati.

4) *Physical attack*. L'aggressione fisica è il quarto stadio. Tali comportamenti sono definiti, in molti ordinamenti, quali *hate crimes*. I componenti dei gruppi di minoranza sono, in questo caso, presi di mira con atti vandalici, distruzione delle loro proprietà o attacchi fisici violenti. Il linciaggio delle persone di colore, i ghetti ebraici, le violenze contro la comunità LGBT sono alcuni esempi.

5) *Extermination*. Infine, il gruppo di maggioranza può voler *sterminare* i componenti del gruppo di minoranza; in pratica, vuole eliminare *l'esistenza stessa* degli avversari. I quattro esempi più noti di genocidio, crimini di guerra e pulizia etnica sono quelli contro i nativi americani, la soluzione finale nazista nei confronti degli ebrei, il

genocidio in Ruanda e gli episodi di pulizia etnica in Bosnia e durante le guerre nella ex-Jugoslavia.

3. Le differenze tra l'odio offline e l'odio online e la necessità di nuove modalità di contrasto

In un report di Unesco molto dettagliato⁸, sono individuate quattro differenze sostanziali che connotano l'*online speech* rispetto all'*hate speech* "tradizionale".

La prima è la *permanenza* dell'odio, ossia la possibilità dell'odio online di rimanere attivo per lunghi periodi di tempo e in diversi formati, di spostarsi attraverso varie piattaforme con la possibilità di essere ripetutamente collegato ad altri contenuti. L'architettura della piattaforma influenza direttamente il tempo di vita delle discussioni. In ambienti simili a quello di Twitter⁹, sono i *trending topics* a facilitare la disseminazione di messaggi d'odio o la loro più o meno ampia visibilità.

La seconda prende la forma di un ritorno imprevedibile dell'odio: anche se il contenuto è stato rimosso, può riapparire e vivere di nuovo in un altro luogo, in un altro tempo, o sulla stessa piattaforma con un altro nome o un'altra area del sistema.

La terza differenza è l'importanza che assume nel mondo online l'anonimato¹⁰, insieme a pseudonimi e nomi falsi. Con la possibilità che la rete offre alle persone di (credere) di essere anonimi, in molti si sentono più a loro agio nell'esprimere odio: pensano di non essere scoperti, o di non subire conseguenze.

La quarta differenza è la transnazionalità: aumenta l'effetto dell'*hate speech*, e pone complicazioni circa l'individuazione dei meccanismi legali per combatterlo.

Per reagire a una situazione simile, che pone tali aspetti innovativi, sono suggerite almeno cinque strategie differenti.

Il primo punto su cui operare sarebbe quello dell'educazione e di un conseguente aumento di consapevolezza, e maggiore attenzione, nelle conversazioni online.

Fondamentale sarebbe poi chiarire la precisa responsabilità degli *opinion makers*, dei politici e dei mass media professionali in questo ambito dato che, molto spesso, sono i primi, in determinati contesti, a veicolare espressioni d'odio a fini

⁸ I. Gagliardone, D. Gal, T. Alves, G. Martinez, *Countering online hate speech*, consultabile in Internet all'indirizzo <http://unesdoc.unesco.org/images/0023/002332/233231e.pdf>.

⁹ Con riferimento alle (limitate) possibilità di automatizzare un'analisi semantica dell'odio online e di intervenire rimuovendo determinate informazioni su Twitter, e a un'analisi collegata, appunto, ai citati *trending topics*, si vedano: N.D. Gitari, Z. Zuping, H. Damien, J. Long, *A lexicon-based approach for hate speech detection*, in «International Journal of Multimedia and Ubiquitous Engineering», 10 (2015), 4, pp. 215-230; P.B. Gerstenfeld, Diana R. Grant, Chau-Pu Hiang, *Hate online: a content analysis of extremist Internet sites*, in «Analyses of Social Issues and Public Policy», 3 (2003), 1, pp. 29-44; K. Al-Hasan Haldar, M. Al-Rubaian, *BiSAL. A bilingual sentiment analysis lexicon to analyse dark web forums for cyber security*, in «Digital Investigation», 14 (2015), pp. 53-62.

¹⁰ Con riferimento al delicato aspetto dell'anonimato si vedano G. Resta, *Anonimato, responsabilità, identificazione: prospettive di diritto comparato*, in «Diritto dell'Informazione e dell'Informatica», 2 (2014), pp. 171-205; G. Finocchiaro (a cura di), *Diritto all'anonimato*, CEDAM, Padova 2008.

elettorali e di audience. Opportuno sarebbe anche ripensare all'intervento della legge e della sanzione penale e di quelle misure giuridiche che si potrebbero coordinare al meglio con le azioni investigative, e valutare l'opportunità di formalizzare il fatto che l'*hate speech* si debba considerare un vero e proprio crimine in tutti gli Stati.

Utile sarebbe, poi, aprire una discussione pubblica sul punto delle responsabilità per i contenuti odiosi e della necessità di una maggiore trasparenza nella gestione delle informazioni da parte delle community online più frequentate e delle piattaforme di social media¹¹. In molti dibattiti si sostiene infatti che i provider non stiano affrontando con sufficiente serietà la questione.

Un aspetto interessante potrebbe essere, infine, quello di stimolare lo sviluppo di misure tecniche utili anche nell'ambito del traffico di odio in rete: si pensi a sistemi di filtraggio dei contenuti individuati come estremi (con tutti i pericoli conseguenti di "soffocare" forme di parlato legittime) o, addirittura, alla possibilità di influenzare i discorsi in rete, e la loro visibilità o meno, tramite algoritmi.

Con riferimento agli strumenti di reazione, infine, il rapporto Unesco ne elabora cinque.

Il primo consiste in un processo di *monitoring* e di analisi dell'odio da parte della società civile, e l'idea è quella di disegnare una mappa e di tenere sotto controllo gli episodi d'odio con riferimento a una determinata zona territoriale o a parti della società.

Il secondo punto prevede un'attività di promozione, presso i singoli individui, di contro-parlato che vada a individuare gli specifici episodi e le precise espressioni, e cerchi di mitigarli.

Una terza azione efficace potrebbe essere quella di coordinare, all'interno delle organizzazioni non governative, il processo di denuncia alle autorità dei casi più violenti di odio online.

La quarta reazione utile sarebbe certamente una campagna di sensibilizzazione, congiuntamente a una strategia di intervento, di tutte quelle società informatiche che ospitano, o fanno transitare sulle loro piattaforme, determinati contenuti.

La quinta risposta potrebbe essere strutturale, ossia il dare potere agli utenti tramite un percorso educativo e di training conoscitivo, affinché sviluppino un'etica, un galateo e capacità necessarie per gestire la corretta, libera espressione su Internet. Il termine in lingua inglese per questa azione, molto suggestivo, è *Media and Information Literacy*.

¹¹ Con riferimento agli equilibri politici (con conseguenze giuridiche) che si possono generare (e altrare) su una piattaforma di social network, si vedano: R. Sandoval-Almazan, J. Gil-Garcia, *Towards cyberactivism 2.0? Understanding the use of social media and other information technologies for political activism and social movements*, in «Government Information Quarterly», 31 (2014), pp. 365-378; S. Scalzini, *I servizi di online social network tra privacy, regole di utilizzo e violazione dei diritti di terzi*, in «Giurisprudenza di Merito», 2012, 12, pp. 2569-2591.

Unire tutte le precedenti azioni di risposta e di reazione all'odio online, anche in maniera creativa, può produrre risultati concreti, contribuendo a far rimanere Internet un luogo potenzialmente positivo e aiutando a costruire e mantenere la società della conoscenza sulla base della pace, dei diritti umani¹² e di uno sviluppo sostenibile.

Un ulteriore fattore importante di differenza tra odio online e odio offline è che, nonostante la maggior parte delle espressioni d'odio siano, anche in rete, portate nei confronti di etnia e nazionalità (con una crescita sensibile di offese su religione¹³ e classe sociale), l'odio online si distingue per tre fattori molto importanti: i) il mondo online rende inefficaci o inappropriate alcune misure pensate per gestire l'odio sui media tradizionali; ii) si registra una sensibile differenza tra un messaggio online che desta nessuna o poca attenzione e uno che, invece, diventa virale, per cui i messaggi possono essere diversi tra loro anche in un'ottica di impatto sociale; iii) vi è un problema di applicazione della legge su piattaforme che hanno sedi in diversi Paesi del mondo.

Quindi, se l'*hate speech* online, per quanto riguarda i "temi", non è intrinsecamente diverso da simili espressioni trovate offline, ci sono sfide uniche che sono portate dal contenuto online e dalla sua regolamentazione.

A nostro avviso, i due aspetti più problematici che lo studio Unesco individua, e che sono caratteristici dell'odio online, sono la *permanence* e la *itinerancy*.

L'idea di *permanence* l'abbiamo già esposta: più il contenuto rimane online, più il danno potenziale da infliggere alla vittima aumenta, e più questa caratteristica dà un senso di potere a chi porta avanti le espressioni d'odio.

Al contempo, se si riesce a rimuovere il contenuto in una fase avanzata, si può in qualche modo limitare l'esposizione al danno.

L'architettura che caratterizza le differenti piattaforme può, poi, permettere all'argomento dell'espressione d'odio di rimanere vivo per periodi più lunghi o più brevi.

Le conversazioni su Twitter organizzate attorno ai *trending topics*, ad esempio, possono facilitare l'ampia e veloce diffusione di messaggi di odio, ma permettere anche a tali messaggi di essere più evidenti e, quindi, contrastabili o eliminabili.

¹² Con riferimento al ruolo dei diritti civili nell'ambiente tecnologico si vedano G. Ziccardi, *Resistance, liberation technology and human rights in the digital age*, Springer, Dordrecht-Heidelberg-New York-London 2012; M.L. Siegel, *Hate speech, civil rights, and the Internet: the jurisdictional and human rights nightmare*, in «Albany Law Journal of Science & Technology», 9 (1999), pp. 375-398.

¹³ Anche l'odio politico motivato dalla religione ha, negli ultimi anni, assunto un'importanza sempre maggiore nel dibattito pubblico. Si veda, sul punto, S. Ferrari, *Libertà religiosa e sicurezza nazionale in Europa dopo l'11 settembre*, in «Quaderni di Diritto e Politica Ecclesiastica», 1, pp. 161-184. Sullo stesso punto si veda anche C. Salazar, *I «destini incrociati» della libertà di espressione e della libertà di religione: conflitti e sinergie attraverso il prisma del principio di laicità*, in «Quaderni di Diritto e Politica Ecclesiastica», 1 (2008), pp. 67-118. Su una delle prime regolamentazioni in Europa in questo ambito si veda A. Gianfreda, *Il «Racial and Religious Hatred Act 2006»: il dibattito parlamentare e il testo approvato*, in «Quaderni di Diritto e Politica Ecclesiastica», 2 (2006), pp. 407-424.

Facebook, al contrario, permette delle conversazioni multiple e in parallelo, con catene di commenti, che possono non essere notati (a meno che non siano segnalati), creando in questo caso degli spazi che permettono agli aggressori di offendere, dileggiare o discriminare i soggetti che hanno preso di mira.

Il secondo aspetto fa notare invece come l'*hate speech* online possa essere *itinerante*. Il significato è semplice da comprendere: anche quando il contenuto è stato rimosso, può trovare espressione altrove, anche sulla stessa piattaforma con un nome differente o in diversi spazi online.

Se un sito web è oscurato, può essere riaperto presso un servizio di web hosting che abbia politiche meno stringenti del precedente, o migrando in un paese dove le leggi sull'*hate speech* siano meno rigorose.

La durata dei materiali di *hate speech* online è unica, a causa dei bassi costi di produzione e di diffusione, e al potenziale costante che possano essere ripescati o fatti rivivere immediatamente, tanto che possono riapparire quando il discorso prende una certa direzione.

Si è in presenza, in definitiva, di un fenomeno in evoluzione e che richiede sforzi collettivi.

Il focalizzarsi solo sulle misure *repressive* può far perdere di vista la complessità di un fenomeno che è ancora poco compreso e che domanda interventi personalizzati e risposte coordinate da una serie di differenti attori nella società.

Interessante può essere l'idea di rendere *social* anche il processo di moderazione e di controllo all'interno delle grandi piattaforme, permettendo agli utenti di risolvere le controversie tra loro o di segnalare, anche insieme, certi comportamenti.

Tale idea di *social reporting* può essere anche più ampia dei limiti restrittivi dell'odio online, ossia dare anche la possibilità a un utente di segnalare espressioni che comunque non violano la policy e i termini di servizio ma che sono offensivi.

4. Contrastare in maniera efficace l'*hate speech* e la propaganda

In uno studio recente si è riflettuto su come sia possibile fronteggiare l'odio in rete muovendo da un approccio originale¹⁴.

L'autore cita, in esordio, una suggestiva frase di Edmund Burke sul pericolo del non agire: «tutto ciò che è necessario per il trionfo del male è che gli uomini buoni non facciano nulla» per proporre una tesi basata sulla centralità della *propaganda*, ossia sul fatto che il motivo di base della diffusione dell'*hate speech*, essendo pensato per minacciare pubblicamente, sia tendenzialmente un'attività di propaganda, soprattutto a beneficio delle organizzazioni già esistenti offline.

¹⁴ Vedi R. Cohen-Almagor, *Countering hate on the Internet*, in «Annual Review of Law and Ethics», 22, 2014, pp. 431-443.

Analizzando il panorama online statunitense, l'Autore nota come un gruppo molto odiato in America sul web, e oggetto di questa attività di propaganda negativa, siano innanzitutto gli ebrei, descritti sui siti estremisti quali cospiratori, uniti in un patto segreto per governare il mondo, abili a mentire per tenere nascosto il loro piano e per effettuare il lavaggio del cervello alle menti dei cristiani americani, mentre controllano le università, i mass media e le banche.

L'Autore cita frasi estremiste che, ad esempio, sostengono che i mass media (ovviamente controllati dagli ebrei) stanno convincendo intera generazione che l'omosessualità sia un normale e accettabile modo di vita, o che non vi sia nulla di sbagliato se le donne bianche si sposano maschi di colore o se i bianchi si sposano le asiatiche, ed arrivano persino ad affermare che tutte le razze sono uguali.

Nei siti d'odio lo studioso evidenzia poi, come obiettivo di propaganda negativa, gli afro-americani: anche loro, sul web estremo, sono obiettivi comuni, descritti sui forum come brutali, primitivi e biologicamente inferiori nonché causa della decadenza culturale in America.

Oggetto d'odio sono poi, inevitabilmente, le comunità omosessuali, viste online, nei siti estremisti, come gruppi di peccatori che minacciano i valori religiosi della comunità e la sopravvivenza della razza. Più di recente, si sono diffuse azioni di propaganda d'odio contro gli immigrati e i musulmani.

In un simile quadro di propaganda, diventato così potente grazie al mezzo, l'idea, pur nobile, di combattere lo *speech* con *more speech*, ossia cercando di dialogare e di proporre tesi alternative, può non rivelarsi una risposta sufficiente.

Bisognerebbe, secondo l'Autore, promuovere al contempo la diffusione di standard di responsabilità sociale e morale, un'azione combinata che superi la strategia *speech v. speech* che è la prima, tipica risposta americana.

Una maggiore comunicazione, tramite apertura ed esposizione del problema, può arrivare a mostrare che tutte le razze umane meritano rispetto e attenzione, che hanno la stessa dignità e che una società basata sulle distinzioni razziali nega quei valori liberal-democratici che sono cari a tutti i cittadini quali il pluralismo, la diversità, l'individualità, la libertà, l'eguaglianza, la tolleranza, la giustizia.

Il mondo della scuola è, ovviamente, cruciale e, secondo lo studioso, già nelle scuole elementari occorrerebbe avvertire circa la presenza dell'odio su Internet, spiegare e far capire correttamente cosa è, chi sono le vittime e la sua evoluzione storica.

I percorsi educativi contro l'odio online possono essere anche più articolati (e vincolanti) e prendere la forma di *policy* nelle scuole primarie e secondarie, nelle università e sul posto di lavoro, così come all'interno delle istituzioni. Si tratta di semplici regolamenti, accettati dagli utenti e con sanzioni in caso di violazione, che vietano forme di odio e aggressione online e offline e che annunciano anche controlli affinché i computer non siano utilizzati per tali scopi.

Lo studioso richiama un termine a suo dire interessante che può significativamente illustrare questo comportamento del cittadino contrario all'odio online che potrebbe contrastare la propaganda e mutare il quadro: *netcitizenship*.

È da intendersi come il comportamento del “buon cittadino di Internet”, ossia lo sviluppare delle condotte responsabili quando si naviga, portando sempre contributi utili alla discussione e ai dibattiti e aumentando la cautela quando ci si trova coinvolti in dibattiti ad alto rischio di generare odio.

Una buona “cittadinanza digitale” può anche prevedere esperimenti di contro-parola collettiva: più cittadini che creano una comunità virtuale dedicata a diffondere contro-affermazioni su temi critici. Un esempio è *Wipeout Homophobia*, un progetto di risposta evoluta alle minacce ai gay su Internet, che vanta milioni di visitatori su Facebook.

Circa il delicato tema della responsabilità dei provider, l'autore separa due punti, a suo avviso complementari ma distinti.

Un primo passaggio dovrebbe essere quello di sviluppare standard, all'interno delle società, per far sì che si diffondano prassi accettabili e responsabili da parte degli utenti, e ciò si può ottenere con la redazione di policy chiare e trasparenti sull'*hate speech* e con il fornire meccanismi semplici e amichevoli per la segnalazione dell'occorrenza di episodi d'odio.

Un secondo passaggio consisterebbe nell'influenzare i risultati dei motori di ricerca più importanti. Lo studioso fa l'esempio della ricerca di Martin Luther King su Google: uno dei primi risultati è www.martinlutherking.org, un sito d'odio che è mascherato da fonte storica obiettiva con notizie sul leader dei diritti umani americano. Si possono manipolare gli algoritmi di ricerca per evitare una simile confusione che può ingannare persone disattente o poco esperte?

La cooperazione internazionale è, comunque, un elemento essenziale in ogni strategia di contrasto, soprattutto in Europa¹⁵ dove l'attenzione è molto alta grazie anche a network quali l'*International Network against Cyber Hate*, finalizzato alla condivisione di conoscenza, allo scambio di *best practices* ma anche al monitoraggio dell'uso della rete in vari Paesi.

Iniziative simili sono la LICRA in Francia e la CIDI (focalizzata su Israele) in Olanda. Un'azione interessante è la pubblicazione di elenchi di siti web che veicolano odio su basi regolari, evidenziando sia i contenuti, sia quali siano i provider che li ospitano senza verifica.

La cooperazione tra individui contro l'odio online prende la forma, anche, del contatto sociale tra le persone, che mira a ridurre il pregiudizio razziale fornendo opportunità ai membri di razze differenti di interagire, spesso in gruppi (si pensi alle scuole o ai centri sportivi).

Questo approccio, dai nobili intenti, muove dalla teoria che, spesso, il pregiudizio razziale e le frizioni etniche siano frutto di ignoranza. Si assimila l'idea, proveniente da altri ma mai verificata di persona, che soggetti di altre razze non

¹⁵ Con riferimento al contrasto tra interpretazione europea e approccio nordamericano sul tema delle espressioni d'odio si veda C. D. Van Blarcum, *Internet hate speech: the European framework and the emerging American haven*, in «Washington & Lee Law Review», 62 (2005), pp. 781-830.

siano, ad esempio, affidabili; se la società, però, offre opportunità per confrontarsi e conoscersi, ci si renderà conto che si è tutti uguali¹⁶.

Non è semplice, per l'interprete, fondere tutti questi punti, astrattamente corretti, nel caotico "mondo" dell'online e delle piattaforme di social network ed elaborare una strategia di contrasto efficace valida per ogni cultura e ogni "ambiente" digitale.

Il lato dell'educazione digitale, o della cittadinanza elettronica, ha la sua vulnerabilità nell'abbassamento drastico dell'età media di utilizzo degli smartphone da parte dei bambini e adolescenti (in Italia, secondo i dati ISTAT, attorno agli 8 anni). Non è quindi facile spiegare con cura a menti così giovani i principi cardine della legalità, dell'empatia, dell'uso corretto del linguaggio e dell'affettività.

Il diritto può essere, dal canto suo, uno strumento pericoloso quando si opera al confine con l'opinione e l'espressione dell'essere umano: può diventare strumento di censura, di repressione, di governo autoritario e, a sua volta, di odio. Al contempo, il diritto si trova come sempre a "rincorrere" la tecnologia e fatica sensibilmente a trovare rimedi che siano applicabili a un quadro così fluido.

Il lasciare, infine, la scelta (e l'azione) alle piattaforme (aumentando il controllo automatizzato da parte dei provider o potenziando le possibilità di segnalazione dall'interno, da parte degli utenti stessi) rischia di vanificare l'azione a causa dell'enorme mole, per di più in aumento esponenziale, dei contenuti d'odio che circolano (soprattutto in occasione di eventi di crisi della politica o elettorali).

Sarà quindi compito dell'interprete, in ogni, singolo momento storico, cercare di trovare un equilibrio tra diritti di manifestazione del pensiero e rispetto (e protezione) della dignità della vittima cercando, da un lato, di individuare principi solidi su cui costruire valide fondamenta teoriche ma, dall'altro, mantenendo quella flessibilità che è indispensabile quando si tratta di questi temi e che consente di adattarsi, ed adeguarsi, a cambiamenti che sono, solitamente, repentini.

¹⁶ Vedi R. Delgado, J. Stefancic, *Hate speech in cyberspace*, in «Wake Forest Law Review», 49, 2, 2014, pp. 319-343.