EDITORIALE

Questo numero di *Sistemi Intelligenti* si presenta davvero molto ricco, sia sotto il profilo quantitativo (è ciò che tecnicamente si definisce un "balenottero"), sia per la varietà dei contenuti, che sono raccolti intorno a tre nuclei tematici: un focus monografico sulla comunicazione nei social media, che occupa gran parte del volume; l'inaugurazione di una nuova sezione della rivista, *InterMenti*, dedicata a interviste approfondite a figure chiave delle scienze cognitive; infine una nutrita sezione di *Squib e discussioni*, con ben cinque contributi.

Vediamo innanzitutto il lungo focus monografico su "Social network e comunicazione", a cura di Bianca Cepollaro e Paolo Labinaz, caratterizzato dall'integrazione di diverse prospettive teoriche con approccio multidisciplinare. Si tratta di un tema di grande attualità, alla luce dell'importanza che i social network, come Facebook, Instagram o Twitter, rivestono oggi nella vita di tutti noi; un tema che pone sfide del tutto nuove, sia di tipo teorico sia sperimentale, per chi si occupa di fenomeni comunicativi e discorsivi da un punto di vista filosofico, linguistico, psicologico e anche sociologico. I siti di social network si presentano infatti come spazi virtuali che mettono assieme funzioni e servizi di diverso tipo su un'unica piattaforma e li fanno interagire tra loro: a partire da un profilo personale, che può essere pubblico, semi-pubblico o privato, è possibile interagire con gli altri utenti, gestire e/o allargare la propria rete di amicizie, raccogliere informazioni, produrre e condividere contenuti di vario tipo e formato, acquistare prodotti, e molto altro ancora. In particolare, le comunicazioni pubbliche su profili, pagine e gruppi si configurano come veri e propri dialoghi scritti.

Questi dialoghi, per quanto abbiano punti in comune con le interazioni faccia a faccia della nostra vita quotidiana, presentano specificità e peculiarità che stanno divenendo sempre più oggetto privilegiato di analisi da parte di filosofi e studiosi di discipline riconducibili all'ambito delle scienze cognitive (quali la psicolinguistica, la psicologia cognitiva, la filosofia dell'informazione, gli studi sull'interazione uomo-macchina),

dando luogo anche a ricerche interdisciplinari sempre più complesse e avanzate. Le tematiche affrontate sono delle più varie: si va dal fenomeno delle fake news e della loro proliferazione, passando per questioni quali il discorso d'odio, la costruzione e gestione della propria identità, in particolare con riferimento alla componente reputazionale, fino ad argomenti di stampo prettamente linguistico-comunicativo, come l'uso dell'ironia e le forme di cortesia in rete, o ancora le funzioni comunicative di *like*, *share* ed *emoticon*. Vi è inoltre un grande interesse, in particolare da parte di giuristi e filosofi del diritto, per questioni legate a cyberterrorismo, tutela della privacy e conflitto tra censura e libertà di espressione.

Nel focus monografico, ci si propone di offrire una panoramica, seppur limitata, sulle potenzialità e, allo stesso tempo, complessità di questo campo di ricerca in continua e rapida espansione. È importante ricordare che gli studi sulla comunicazione mediata della rete nascono molto prima della diffusione dei servizi di social network (si veda Roversi, 2004). Essi prendono avvio infatti durante quella che è stata definita l'era del Web 1.0. Stiamo parlando di un web ancora "statico", in cui gli utenti potevano consultare i contenuti presenti sulle pagine dei siti ma non c'era alcuna possibilità di interazione diretta. La comunicazione tra utenti avveniva tramite messaggi di posta elettronica, forum e chat basate sul protocollo IRC. A differenza dei giorni nostri, date le limitazioni tecniche, non tutta la comunicazione avveniva in tempo reale e perciò una distinzione fondamentale era quella tra sistemi di comunicazione sincroni (chat) e asincroni (posta elettronica e forum) (Pistolesi, 2004: 14-17).

La nascita e la diffusione dei servizi di social network sono la manifestazione più evidente di quella che è la rivoluzione del Web 2.0. Questa espressione, introdotta nel 2004 come titolo di una serie di conferenze organizzate dalla O'Reilly Media, vuole mettere in luce un nuovo paradigma del web che, andando oltre la staticità del Web 1.0, mette al centro la possibilità da parte degli utenti di creare e diffondere contenuti attraverso reti che sono rimodellabili continuamente in base alle loro preferenze di interazione (O'Reilly, 2005)[1]. Con la diffusione della connessione costante alla rete Internet attraverso dispositivi mobili (smartphone multifunzione e in seguito tablet), si è avuta la consacrazione definitiva dei servizi di social network come sistemi di comunicazione per eccellenza. La contestuale nascita e sviluppo di una grande varietà di applicativi per questi dispositivi (quali WhatsApp, Facebook Messenger,

[1] Nel frattempo il web è già passato attraverso altre due fasi evolutive e siamo già in attesa della prossima: per ora abbiamo vissuto la fase del Web 3.0, o "web semantico", e siamo immersi attualmente nel Web 4.0, o "Internet delle Cose". La prossima fase, che è stata già denominata "web simbiotico", prevede l'inclusione di una componente emozionale nell'interazione tra utente e sistema.

Telegram, Snapchat e così via) ha reso l'interazione e la condivisione di contenuti tra utenti qualcosa alla portata di tutti e realizzabile ovunque, assottigliando sempre di più la distinzione tra sistemi sincroni e asincroni. Oggigiorno tutta la comunicazione mediata della rete è infatti potenzialmente in tempo reale: i tempi degli scambi non dipendono più da vincoli tecnici quanto dalle scelte e preferenze dei singoli individui.

I lavori che compongono questo focus monografico vanno ad affrontare questioni che hanno, per diversi motivi, grande rilevanza ai fini di una migliore comprensione del fenomeno social network e dei modi di affrontarlo in quanto oggetto di studio. Si va da studi che si occupano dell'attrattività dei social network e dei rischi insiti in essi, passando per analisi relative a problemi teorici e pratici legati all'utilizzo di strumenti e concetti tipicamente adatti a contesti non virtuali nello studio di fenomeni collegati alla comunicazione via social network, per arrivare infine a ricerche di tipo sperimentale relativamente ad aspetti caratteristici di questo tipo di comunicazione e alle sue potenzialità in quanto strumento "educativo".

I primi quattro contributi toccano alcune questioni generali circa il fenomeno dei social network e ciò che ruota attorno a essi: dalle trappole cognitive alle bolle epistemiche, dalla disinformazione digitale alla propagazione di pregiudizi tramite algoritmi.

In "Cacciatori (di informazioni) e prede (di trappole cognitive) nel web 2.0: una lettura cognitivo-evoluzionista dell'attrattività dei social network", Marco Fasoli propone una spiegazione dell'attuale successo e diffusione dei social network secondo una prospettiva cognitivo-evoluzionista. Nell'articolo si sostiene che alla base del loro potere di attrazione ci siano fattori di tipo endogeno ed esogeno. I primi fanno riferimento agli elementi intrinseci alle piattaforme social che in parte dipendono dalla struttura delle tecnologie digitali attraverso cui accediamo a esse. In particolare, il fatto che agli utenti sia possibile consultare continuamente le bacheche online dei membri della loro cerchia sociale soddisfa un fondamentale bisogno informativo ancestrale proprio della specie umana. I fattori di tipo esogeno fanno riferimento invece a espedienti tecnici, che non fanno parte della struttura intrinseca delle piattaforme social, ma sono implementati in esse per motivi economici al fine di catturare l'attenzione degli utenti. Fasoli sostiene che questi espedienti tecnici danno luogo a vere e proprie "trappole cognitive". Dall'esame di questi due tipi di fattori l'autore conclude che la nostra condizione in quanto utenti di social network sia piuttosto paradossale, poiché siamo allo stesso tempo cacciatori di informazioni ma anche prede delle trappole cognitive che vengono implementate in essi.

Il contributo di Francesca Ervas – "Metafore visive, comunità immaginate e razionalità differita" – affronta il tema delle comunità virtuali in rete, esaminando alcuni rischi degenerativi a cui sono continuamente esposte. L'autrice concepisce tali comunità come "comunità immagina-

te" secondo la definizione di Benedict Anderson, ovvero comunità i cui membri, pur non essendo *in presentia*, condividono un certo insieme di pratiche, simboli, metafore ecc., e lo stile con il quale si immaginano come una "traduzione collettiva simultanea", che ha natura multimodale ed è resa possibile da una tecnologia che permette ai membri di condividere contenuti reinterpretandoli continuamente. Per Ervas, è proprio lo stile con cui vengono immaginate a esporre le comunità virtuali al rischio di degenerare in comunità immaginarie o inautentiche. Da un lato, nota l'autrice, la mancanza della relazione faccia a faccia può rendere la comunicazione "solitaria" e il ragionamento collettivo sottoposto agli stessi *bias* di conferma del ragionamento individualistico. Secondo la sua analisi, quest'ultimo aspetto è al cuore del noto fenomeno delle "bolle epistemiche", che portano la razionalità altrui a essere "differita". Dall'altro, la natura multimodale e in particolare visiva della comunicazione via social network può limitare l'immaginazione degli utenti, ancorandola a dettagli percettivi. A supporto di ciò, Ervas analizza due casi di studio di metafore visive tradotte e reinterpretate in un gruppo Facebook al fine di mostrare come il fenomeno della "razionalità differita" porti a preservare le "bolle epistemiche", producendo interpretazioni fuorvianti.

In "Epistemologia delle *fake news*", Tommaso Piazza e Michel Croce analizzano dal punto di vista epistemologico il fenomeno della proliferazione di fake news in rete, focalizzando l'attenzione su tre domande centrali. La prima è la questione definitoria su che cosa siano le fake news e come debbano essere definite. La seconda indaga quali meccanismi facciano sì che le fake news proliferino sui social network. La terza, invece, è la questione della responsabilità che si interroga su chi, nel complesso processo che include la generazione, la pubblicazione e la diffusione di fake news, debba essere ritenuto colpevole. Piazza e Croce, tenendo conto dei principali lavori nella letteratura filosofica sulle fake news, rispondono a queste tre questioni. Offrono innanzitutto una definizione di fake news che non sollevi le obiezioni a cui vanno incontro le precedenti proposte in letteratura. Inoltre, cercano di individuare le principali cause della proliferazione di fake news, facendo appello a nozioni quali i *bias* cognitivi e il tipo di strutture comunitarie in cui gli utenti dei social network si organizzano. Infine, analizzano in modo originale la responsabilità epistemica dei consumatori di fake news.

Il contributo "Social network e algoritmi di *machine learning*: problemi cognitivi e propagazione dei pregiudizi" di Teresa Numerico si occupa, da un punto di vista epistemologico, dei possibili effetti discriminatori e pregiudiziali causati dall'utilizzo di tecniche di interpretazione algoritmica dei comportamenti umani nei social network. Sulla base di alcuni studi, l'autrice mostra come gli algoritmi addestrati su basi dati umane riproducono, anche con maggiore rigidità, credenze e pregiudizi simili a quelli degli umani che hanno prodotto le basi dati linguistiche di addestramento. Numerico ritiene che, visto l'ampio utilizzo di tali

algoritmi nei social network, si debba porre maggiore attenzione alla loro capacità di fare previsioni relativamente ai comportamenti futuri degli utenti. In particolare, osserva l'autrice, è necessaria una migliore comprensione di regole e criteri progettati per implementarli nei social network. Allo stato attuale ciò è possibile solo tramite i loro output, in quanto il funzionamento degli algoritmi non è di pubblico dominio. D'altra parte, nota Numerico, se si vuole valutare in maniera analitica la loro efficacia e affidabilità, e comprendere quindi come evitare certi effetti distorcenti che essi producono, andrebbero resi pubblici.

Il focus monografico prosegue poi con tre saggi che affrontano alcuni problemi di tipo teorico e pratico che sorgono quando si tratta di applicare concetti e strumenti tipicamente adatti a contesti non virtuali nello studio di fenomeni collegati alla comunicazione via social network.

In particolare, due contributi si concentrano sulle difficoltà interpretative che la comunicazione tramite social network pone soprattutto agli usi non letterali del linguaggio. Il primo è il contributo di Francesca Panzeri – "Stai scherzando? (Non) riconoscere l'ironia nei social network" –, dedicato al tema del riconoscimento dell'ironia sui social network: l'autrice illustra gli ostacoli che la comunicazione online pone all'impresa di riconoscere l'ironia e le strategie impiegate per ovviarli. Intorno al fenomeno dell'ironia vi è un acceso dibattito in filosofia e in linguistica; tra le questioni più discusse vi è quella dell'individuazione dei cosiddetti marcatori dell'ironia, cioè indizi metacomunicativi volti a scongiurare fraintendimenti e a facilitare l'ascoltatore nell'interpretazione corretta del significato inteso da chi parla. A partire da questi studi, Panzeri si chiede quali possano essere i marcatori dell'ironia una volta che si passi dalla comunicazione orale a quella scritta e da un'interazione faccia a faccia a quella virtuale su social network, per indagare se sia effettivamente più difficile riconoscere l'ironia quando avviene in forma scritta e nel contesto di un social network. La prima parte del saggio consiste in una disamina delle principali teorie che hanno cercato di analizzare il funzionamento dell'ironia, con un'attenzione particolare alla questione dei marcatori dell'ironia nella comunicazione orale e in quella scritta; la seconda parte del lavoro invece è dedicata alla presentazione e discussione di due casi di studio, caratterizzati dal mancato riconoscimento da parte degli utenti dell'ironia intesa in alcuni contenuti postati su social network.

Il tema di come la comunicazione tramite social network renda più complessa l'interpretazione di certi usi non letterali del linguaggio è ripreso in "Identità e linguaggio discriminatorio nei social network", in cui Bianca Cepollaro e Paolo Labinaz si occupano del ruolo giocato dall'identità di chi scrive nel guidare chi legge verso una corretta interpretazione di alcuni usi del linguaggio apparentemente discriminatori: in particolare, si concentrano sugli usi non denigratori degli epiteti da un lato e su ironia e umorismo (a prima vista) discriminatori dall'altro. In questi

casi l'identità di chi scrive è un elemento cruciale nell'interpretazione di un certo contenuto: perderla di vista può modificare radicalmente il significato complessivamente inteso e in certi casi persino mettere in discussione la legittimità dell'ironia e della comicità rispetto a temi quali la discriminazione razziale, di genere, ecc. Inoltre, la difficoltà di stabilire l'identità di chi scrive pone seri problemi per quanto riguarda il controllo e la censura di post e commenti in cui occorrono usi del linguaggio come quelli trattati: lo sviluppo di *policies*, necessario nella gestione di un social network, richiede di affrontare questioni teoriche e pratiche insieme, che questo contributo cerca di illustrare.

I social network pongono sfide rilevanti anche agli studiosi di diritto. In "Moleste analogie? Social network e norme penali", Giovanni Tuzet si occupa di una di queste sfide, chiedendosi se sia possibile estendere una norma penale come quella sulle molestie recate "in luogo pubblico o aperto al pubblico", e quindi pensata per la comunicazione tradizionale, alla comunicazione attraverso social network. L'autore osserva preliminarmente come nel nostro ordinamento non sia consentita l'estensione analogica delle norme incriminatrici, mentre non è proibita la loro estensione "interpretativa". Il problema è che, nota Tuzet, è piuttosto difficile distinguere chiaramente tra analogia e interpretazione estensiva nei singoli casi. Per questo motivo, non basta invocare l'estensione di una norma incriminatrice ma è necessario dimostrare caso per caso che l'estensione sia di tipo interpretativo, e non analogico. Tuzet conclude esaminando un caso effettivamente accaduto di molestie (principalmente sessuali) su Facebook e mostra come la norma sulle molestie recate "in luogo pubblico o aperto al pubblico" possa essere effettivamente interpretata estensivamente nel caso in questione.

Il focus monografico si chiude con tre contributi che presentano nuovi metodi per lo studio empirico di fenomeni caratteristici della comunicazione via social network e alcuni risultati preliminari di ricerche sperimentali.

Il primo lavoro si inserisce nell'ambito della pragmatica della cosiddetta comunicazione mediata dal computer (CMC). In "La pragmatica di ☺ e ☹. Quando e quanto usiamo le emoticon su WhatsApp", Filippo Domaneschi, Luca De Vita e Simona Di Paola propongono un'indagine sull'uso delle emoticon nelle interazioni nella CMC. Per 'emoticon' si intendono quei segni grafici che accompagnano o sostituiscono il testo nella CMC. Il tipo di emoticon più diffuso è quello che rappresenta iconicamente le espressioni facciali, come ☺, ☹, 😐 e così via, frequenti soprattutto nelle chat e negli scambi via mail. Il contributo di Domaneschi, De Vita e Di Paola prende in esame l'applicazione di messaggistica WhatsApp per indagare l'idea che la frequenza con cui occorrono le emoticon positive (per esempio, ☺) e negative (per esempio, ☹) sia influenzata da almeno due fattori: lo scopo conversazionale e la natura della conversazione, che può essere cooperativa o conflittuale. I risultati

dell'esperimento comportamentale che hanno condotto suggeriscono alcune conclusioni preliminari: innanzitutto, le emoticon tendono ad occorrere più spesso in contesti positivi e cooperativi e in scambi verbali di natura socio-emotiva; in secondo luogo, le emoticon caratterizzate da valenza positiva occorrono più spesso nelle conversazioni orientate allo svolgimento di un compito, nonostante si osservi in quei frangenti una minore frequenza d'uso di emoticon in generale.

Il secondo contributo mette in luce un ruolo benefico che i social network possono ricoprire nell'ambito della salute. In "Salute e partecipazione: Facebook come strumento educativo per il coinvolgimento attivo della persona con diabete", Maria Grazia Rossi e Julia Menichetti descrivono il modo in cui Facebook può essere utilizzato come vero e proprio strumento di educazione e di supporto nella gestione di una malattia cronica. Rossi e Menichetti si concentrano sul caso del diabete: molte persone affette da diabete, così come i loro familiari, usano Facebook alla ricerca di informazioni o per discutere con altri utenti di questioni relative alla gestione di questa malattia. Rossi e Menichetti offrono delle analisi qualitative su un campione di scambi avvenuti all'interno di un gruppo Facebook dedicato a diabetici e ai loro familiari; l'osservazione di questi scambi suggerisce che i membri del gruppo riescono effettivamente a condividere informazioni e a supportarsi sia su un piano di conoscenze, sia in termini emotivi. Inoltre, Rossi e Menichetti hanno sottoposto 119 utenti di Facebook con diabete ad un questionario strutturato; l'analisi quantitativa delle risposte è in linea con le osservazioni qualitative circa le interazioni nel gruppo Facebook e mostra come le persone con diabete che utilizzano Facebook siano nel complesso attivamente coinvolte nelle cure e abbiano buone risorse e capacità informative. A partire da questi risultati, Rossi e Menichetti sottolineano l'importanza dei gruppi su Facebook e delle comunità on line in genere come veri e propri strumenti volti a promuovere il supporto tra pari, promuovendo il coinvolgimento attivo della persona nella gestione della malattia cronica.

In "Analizzare l'argomentazione sui social media. Il caso dei tweet di Salvini", Fabrizio Macagno presenta infine un metodo di analisi basato sulla teoria dell'argomentazione che è finalizzato alla costruzione di profili argomentativi di utenti di Twitter. Nel caso specifico egli esamina quali siano state le strategie che hanno caratterizzato il profilo argomentativo di Matteo Salvini nelle vesti di Ministro dell'Interno sulla base di un corpus di 843 messaggi pubblicati sul suo account Twitter, che sono compresi in un arco temporale che va dal 1º giugno 2018, giorno della sua nomina a ministro, fino al 12 gennaio 2019, data in cui nei messaggi si comincia a rimarcare il suo ruolo di capo politico della Lega in vista delle elezioni europee svoltesi a fine maggio di quello stesso anno. Il metodo proposto è strutturato secondo tre livelli: il primo prevede l'annotazione dei tipi di argomento e delle fallacie presenti nei messaggi; con il secondo si va a valutare la struttura degli argomenti; il

terzo livello infine mira all'identificazione delle strategie ridefinitorie, che si ricollegano all'utilizzo di parole emotive a fini manipolatori. I risultati dell'analisi mostrano, tra le altre cose, come solo il 21% degli argomenti identificabili nei *tweet* di Salvini abbia le caratteristiche formali di un argomento completo e la metà degli argomenti totali si configuri come manipolatoria.

Alcuni di questi studi hanno coinvolto termini offensivi e linguaggio volgare; quando necessario all'illustrazione dei fenomeni in questione, tali parole sono esplicitamente citate, nella speranza che questo non offenda la sensibilità dei lettori e che questo contesto di studio protegga da ambiguità: si tratta di menzione e non di uso.

Il numero prosegue con la prima uscita di una nuova sezione della rivista, intitolata *InterMenti* e dedicata alla pubblicazione di interviste con figure di rilievo nel panorama delle scienze cognitive, realizzate non da giornalisti, bensì da studiosi competenti nelle materie di interesse dell'intervistato, onde sollecitare un approfondimento critico dei temi trattati. Il contributo inaugurale è di Carlos Alós-Ferrer, intervistato da Pietro Terna a latere della conferenza di midterm dell'AISC di quest'anno (Lucca, 22-24 maggio 2019), in relazione anche ai contenuti del suo contributo recente apparso proprio sulle pagine di *Sistemi Intelligenti* (Alós-Ferrer, 2018). La loro discussione prende le mosse dall'ipotesi del cervello sociale e dal suo impatto sulle ricerche economiche, portando a riflessioni molto interessanti sul ruolo delle simulazioni ad agenti, sulla possibilità di inserire agenti artificiali in sistemi sociali umani (sia a scopi di ricerca, sia per influenzare in direzioni virtuose le dinamiche sociali), sull'auspicabilità o meno di strumenti di controllo dei mercati di tipo artificiale, e sulle implicazioni etiche e pratiche di tali sviluppi. Infine, la chiacchierata con Alós-Ferrer si conclude con alcuni consigli alle nuove generazioni di studiosi: quest'ultima parte, in particolare, costituisce un *must read* per chiunque stia valutando la possibilità di intraprendere una carriera di ricerca nelle scienze cognitive e sociali.

Il numero si chiude poi con vari interventi brevi di discussione. Il primo è di Massimo Egidi e prosegue il dibattito critico proprio sull'articolo di Alós-Ferrer citato poc'anzi. Nel suo squib "Dalla razionalità all'intelligenza", Egidi ricostruisce in modo sintetico e chiaro lo slittamento, nelle scienze economiche, da un'attenzione prevalente alla nozione di razionalità alla messa in luce del ruolo di una capacità più generale, eppure non meno importante – l'intelligenza. Mentre alcuni dei presupposti della razionalità classica, in particolare il carattere autointeressato delle scelte e la loro calcolabilità, sono stati messi in dura crisi dalle prove sperimentali, questi stessi risultati ci mostrano come i decisori siano adattabili al contesto e capaci di apprendere dai propri errori. Tratti, questi, decisamente ragionevoli e altamente desiderabili, a prescindere dalla dimostrata fallibilità dei decisori.

Lo squib successivo è di Pietro Terna: in "Gli scacchi e il computer: dalla *Drosophila* dell'Intelligenza Artificiale all'economia", Terna prende spunto da recenti sviluppi nei software per il gioco degli scacchi, in particolare AlphaZero, per sostenere, secondo una formula proposta dal campione Yuri Kasparov, che proprio gli scacchi rappresentino la *Drosophila melanogaster* dell'Intelligenza Artificiale. Proprio come il moscerino della frutta ha fornito un modello essenziale allo sviluppo della genetica, così il gioco degli scacchi è una costante fonte di ispirazione per la ricerca informatica. Ma il modello è il gioco stesso, non il modo in cui lo praticano gli esseri umani: come sottolinea Terna, la novità di AlphaZero sta nel fatto di non usare alcuna informazione sulle strategie usate dai maestri di scacchi umani, giacché si limita ad apprendere le strategie ottimali giocando milioni di partite contro se stesso, partendo inizialmente dalla sola conoscenza delle regole del gioco. E così facendo batte non solo gli scacchisti umani, ma anche i software basati sull'emulazione delle strategie umane. Terna poi discute brevemente l'organizzazione delle consegne di Amazon e recupera gli scritti di un importante economista dei primi del Novecento, Enrico Barone, per valutare le future applicazioni di simili approcci all'economia.

Infine abbiamo un trittico di contributi incentrati sull'attuale stato di salute delle scienze cognitive, prendendo spunto da un recente contributo in merito apparso su *Nature Human Behaviour* (Núñez *et al*., 2019). Domenico Parisi condivide la tesi di un sostanziale fallimento dell'originario progetto cognitivista e manifesta "Quattro perplessità sulla scienza cognitiva": in estrema sintesi, Parisi rimprovera alla scienza cognitiva di avere sottovalutato la rilevanza della dimensione fisica e corporea, avere escluso le scienze sociali dal proprio mix interdisciplinare, essersi appoggiata troppo sulla filosofia, e avere abbracciato una versione dell'informatica molto orientata all'Intelligenza Artificiale di tipo simbolico, finendo col sottovalutare il ruolo (per Parisi decisivo) delle simulazioni al computer nel comprendere la realtà. Cristiano Castelfranchi, in "Fine della scienza cognitiva? Ma non del cognitivismo", risponde a Parisi sottolineando la differenza e l'indipendenza fra la scienza cognitiva come fenomeno storico-sociologico e il progetto teorico del cognitivismo, incentrato sulla nozione di rappresentazione. Cercare di giustificare il superamento del cognitivismo con lo scarso successo (vero o presunto) della scienza cognitiva sarebbe come buttar via il bambino con l'acqua sporca, oltre ad una notevole confusione concettuale fra livelli diversi. Una tesi simile viene sostenuta, seppure con argomenti diversi, anche da Fausto Caruana: in "Stat scienza cognitiva pristina nomine, nomina nuda tenemus?", Caruana mostra come l'eventuale fallimento della scienza cognitiva riguardi più che altro la mancata nascita di un campo interdisciplinare autonomo rispetto alle discipline che lo compongono, talvolta vagheggiato agli albori dell'avventura cognitivista. Di questo afflato ecumenico vi è in effetti poca traccia nella struttura del sistema

accademico a livello internazionale, in cui non solo le discipline continuano a difendere tenacemente la propria autonomia, ma addirittura aumentano la frammentazione specialistica al proprio interno. Diverso, però, è il progetto scientifico di una collaborazione interdisciplinare per comprendere la mente e la società: quel progetto, fa notare Caruana, è oggi più vivo e fertile che mai, come attestano i numerosi risultati scientifici che esso ha prodotto e continua a produrre. Dunque, da questo punto di vista la scienza cognitiva può dirsi tutt'altro che estinta o fallita.

Riferimenti bibliografici

Alós-Ferrer, C. (2018). Analisi critica di «Neuroscienze sociali»: la ricerca sul cervello sociale e l'economia si influenzano a vicenda?. *Sistemi Intelligenti*, *30*, 2, 229-274.

Núñez, R., Allen, M., Gao, R., Miller Rigoli, C., Relaford-Doyle, J., Semenuks, A. (2019). What happened to cognitive science?. *Nature Human Behaviour*, *3*, 782-791.

O'Reilly, T. (2005). *What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software*. O'Reilly Network: http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html.

Pistolesi, E. (2004). *Il parlar spedito: l'italiano di chat, e-mail e SMS*. Padova: Esedra.

Roversi, A. (2004). *Introduzione alla comunicazione mediata dal computer*. Bologna: Il Mulino.

MARCO FASOLI

CACCIATORI (DI INFORMAZIONI) E PREDE (DI TRAPPOLE COGNITIVE) NEL WEB 2.0: UNA LETTURA COGNITIVO-EVOLUZIONISTA DELL'ATTRATTIVITÀ DEI SOCIAL NETWORK

1. Introduzione

Perché i social network hanno ottenuto così tanto successo e risultano essere spesso "irresistibili" (Pasquinelli, 2012)? Per quali ragioni essi ricevono investimenti di tempo così elevati da parte di milioni di persone? In questo articolo indagherò da una prospettiva cognitivo-evoluzionista le cause dell'attuale popolarità delle piattaforme social e della loro capacità di attrazione, che è testimoniata da fenomeni di natura diversa, ad esempio dal rapporto controverso che molti utenti hanno con esse. Negli ultimi anni, infatti, diversi soggetti hanno riscontrato delle difficoltà nella gestione del tempo di permanenza sulle pagine social, lamentando una tendenza a utilizzare questi strumenti più di quanto essi desiderano. Nella stragrande maggioranza di questi casi, fortunatamente, non si tratta di "dipendenza" (per una discussione della dipendenza da social si veda Andreassen, 2015), ma di sovra-utilizzo di tecnologia (Gui, 2014; Fasoli, 2019), cioè di un utilizzo che viene valutato a posteriori come eccessivo da parte degli stessi soggetti e che, in quanto tale, rischia di sottrarre tempo ad altre attività ritenute più rilevanti o significative. Ad esempio, nel 2016 in Gran Bretagna il 41% degli intervistati dichiarava di spendere più tempo di quello che avrebbe voluto online (OFCOM, 2016) e, tra le attività preferite degli internauti, spiccava la navigazione sui social.

Il sovra-utilizzo di internet o di social network rappresenta un caso di incoerenza cognitiva in cui si realizza un divorzio tra scelte e preferenze (Bianchi, 2004). Ad un istante X il soggetto decide di investire il proprio tempo in un certo modo, per esempio dedicando solo 30 minuti al giorno alla navigazione su piattaforme social. Ad un istante Y, tuttavia, lo stesso soggetto supera (a volte di gran lunga) il tempo di utilizzo di questi strumenti da lui stesso stabilito in precedenza, sacrificando altre attività lavorative, di intrattenimento o di socialità che possono essere più impegnative dal punto di vista cognitivo o richiedere di posticipare

le gratificazioni. Quindi, ad un istante successivo Z, il soggetto giudica come insoddisfacente la sua scelta. Alla base di comportamenti di questo tipo spesso è possibile individuare il fenomeno dell'inversione delle nostre preferenze (Ainslie, 2001), a sua volta causato da distorsioni cognitive come lo sconto temporale (si veda anche Paglieri, 2014). L'inversione delle preferenze è un fenomeno che si realizza quando i soggetti devono scegliere tra due opzioni, una meno vantaggiosa (per esempio il ricevere 99 euro) ma più ravvicinata nel tempo e un'altra più vantaggiosa (ricevere una somma di 100 euro) ma che richiede maggiore attesa. Dovendo scegliere tra l'opzione meno vantaggiosa tra 364 giorni o la somma più alta tra 365 giorni, tutti i soggetti optano per quest'ultima. Tuttavia, quando l'intervallo di tempo necessario a ricevere la somma di denaro si riduce e la scelta diventa se ricevere 99 euro domani o 100 euro dopodomani, alcuni soggetti optano per la ricompensa minore ma che richiede meno attesa, dimostrando così un'inversione delle preferenze.

Di fronte a fenomeni come questi è lecito domandarsi su che cosa si basa la capacità dei social network di essere così attraenti e immediatamente gratificanti per molti soggetti. In questo articolo sosterrò che alla base di tale potere di attrazione ci sono due tipologie di fattori, uno endogeno, cioè in qualche modo connaturato a questo tipo di piattaforme, e uno esogeno. Il primo tipo di fattori è costituito da elementi che sono intrinseci alle piattaforme social e che in parte dipendono dalla struttura degli strumenti attraverso cui accediamo a tali piattaforme, cioè le tecnologie digitali. Questo tipo di fattori sarà analizzato nel prossimo paragrafo. Il secondo tipo riguarda invece soprattutto il design delle piattaforme social. In molti casi, infatti, a causa dei meccanismi di monetizzazione dei dati, che creano un'equivalenza tra tempo speso sulle piattaforme da parte degli utenti e introiti, questi siti implementano alcuni meccanismi che possono essere considerati delle vere e proprie trappole cognitive.

2. I fattori endogeni di attrattività nei social network

In questo paragrafo mi occuperò di analizzare brevemente alcune caratteristiche che rendono i social network fortemente attraenti per gli utenti e che sembrano essere connaturati a questi strumenti.

2.1. *Le informazioni riguardanti la cerchia sociale*

Attraverso i social possiamo scambiare facilmente e in poco tempo una quantità enorme di informazioni con tutti i contatti della nostra rete sociale. Per centinaia di migliaia di anni, nel corso della nostra storia evolutiva, le informazioni riguardanti i componenti della propria comunità

e delle comunità vicine e riguardanti le loro attività, i loro spostamenti, le loro relazioni sociali hanno rivestito un ruolo molto più importante rispetto a quello attuale. Ciò in quanto queste informazioni potevano essere determinanti per individuare la presenza di cibo, di possibili minacce o di potenziali partner sessuali. Tuttavia, a causa dell'assenza di tecnologie dell'informazione e della comunicazione efficaci, per un periodo temporale molto lungo queste informazioni hanno costituito delle risorse molto scarse e difficili da reperire. In tale ambiente, le comunicazioni erano molto più difficili, meno frequenti, non potevano realizzarsi a distanza e raggiungevano nella maggior parte dei casi un pubblico limitato. In prospettiva evoluzionista, non è quindi difficile comprendere come mai l'essere umano abbia sviluppato una particolare sensibilità rispetto allo scambio di informazioni, tanto da poter essere caratterizzato – utilizzando un'analogia con le espressioni «erbivoro» e «carnivoro» – come «informivero» (Miller, 1984; Rigutti, Gerbino e Fantoni, 2016).

Da questa prospettiva, gran parte di ciò che in epoca moderna chiameremmo "gossip" sembrerebbe esercitare una forte attrazione su di noi in virtù del ruolo che ha giocato nel nostro passato evolutivo di esseri umani (Pinker, 2003; Dunbar, 2004). Una conferma di questa nostra predisposizione può essere ricavata indirettamente anche da alcuni risultati delle neuroscienze cognitive, dato che è stato appurato che il solo fatto di rivelare nostre informazioni ad altri rappresenta un'attività che è in sé stessa intrinsecamente gratificante (Tamir e Mitchell, 2012) e la gratificazione costituisce un espediente che è servito a spingere gli individui ai comportamenti che si rivelano essere funzionali per la sopravvivenza della specie, si pensi ad esempio alla riproduzione e all'alimentazione.

Quando siamo online e in particolare quando usiamo i social network, la nostra natura di "informiveri" sembra esprimersi in modo particolarmente marcato. Tali strumenti permettono di scambiarci dosi massicce di informazioni personali (a cui siamo particolarmente sensibili) molto più facilmente e velocemente rispetto a quanto abbiamo mai fatto in passato. Inoltre, essi ci permettono ad ogni istante di comunicare ad un gruppo di persone estremamente ampio e fisicamente distante da noi. La *facilità* di scambio, la *quantità massiccia* di informazioni, la *velocità* di comunicazione, la capacità di raggiungere un *pubblico estremamente ampio* e l'*accessibilità senza interruzioni* (possiamo accedere in ogni istante ad un profilo online) rappresentano quindi degli elementi chiave in grado di soddisfare il nostro bisogno di ottenere informazioni che riguardano le nostre cerchie sociali e di comunicare loro informazioni personali, che meritano di essere analizzati un po' più nel dettaglio.

2.2. *Progressività e non nello scambio di informazioni*

In particolare, vale la pena soffermarsi sul modo in cui lo scambio di informazioni personali che avviene attraverso i social si differenzia da quello non mediato tecnologicamente, cioè dal confronto orale che avviene faccia a faccia. Se accettiamo l'ipotesi secondo cui la scrittura non ha più di 5000 anni (Wolf, 2009), per migliaia di anni il dialogo faccia a faccia è stato l'unico canale comunicativo disponibile.

In questa forma di comunicazione, che si realizza soprattutto attraverso il linguaggio, lo scambio di informazioni personali è lento e progressivo: mano a mano che si dialoga si acquisiscono e sedimentano informazioni di diverso tipo sull'altra persona. Alcune di queste dinamiche "progressive" tipiche della presentazione di Sé e della conoscenza degli altri tradizionale si ripropongono anche online. Per esempio, l'amicizia social tra due persone che si sono appena conosciute per motivi di lavoro può avvenire prima in una piattaforma professionale, dove lo scambio di contenuti è più limitato, e solo successivamente in piattaforme più "informali". Di fronte alla richiesta di contatto di una persona conosciuta recentemente, è poi possibile che si decida di accettarla impostando un filtro che rende visibili solo alcuni contenuti. Ad ogni modo, nel complesso possiamo notare come mentre lo scambio di informazioni personali di tipo tradizionale sia caratterizzato dalla progressività e dalla lentezza, quello che avviene attraverso i social sia estremamente più rapido ed in grado di eliminare molti aspetti progressivi. Una volta ottenuto accesso ad un profilo social (senza restrizioni) è potenzialmente possibile attingere in breve tempo – e potenzialmente senza alcuno scambio linguistico diretto – ad una grande quantità di informazioni di diversa natura riguardanti l'altra persona, le sue cerchie sociali, la sua situazione sentimentale, i suoi gusti personali, il suo orientamento politico, i luoghi che ha visitato, ecc. Inoltre, è possibile scartare facilmente le informazioni che si ritengono irrilevanti per concentrarsi su quelle che si ritengono di maggiore interesse. Grazie ai dispositivi portatili e alle reti mobili, che permettono di accedere ai social e ai profili social altrui quasi in ogni luogo dello spazio, queste informazioni sono potenzialmente disponibili 24 ore su 24.

2.3. *Tipologie di informazioni scambiate e capacità di controllo sugli scambi comunicativi*

La comunicazione faccia a faccia si realizza attraverso gli scambi linguistici, i racconti, la comunicazione corporea. Al contrario, nei social essa avviene attraverso lo scambio di contenuti multimediali di natura estremamente variegata: foto, video, testi, messaggi, canzoni, link multimediali o siti web, gruppi, ecc. Un ruolo particolarmente impor-

tante, in questo contesto, sembrerebbe essere giocato dalle immagini, probabilmente anche in virtù della capacità di suscitare risposte emotive molto forti che esse hanno avuto nella nostra storia (si veda ad esempio Freedberg, 1982; Freedberg e Gallese, 2007). Tale preminenza trova conferma anche nel grande successo di quei social network che, come Instagram, si basano quasi esclusivamente sullo scambio di immagini e video.

Questa differenza in termini di contenuti informazionali si traduce anche in una differente capacità di controllo rispetto alle informazioni che i comunicanti desiderano trasmettere su sé stessi e rispetto a ciò che desiderano tenere privato. Nei social network è possibile selezionare accuratamente alcuni contenuti per la pubblicazione, per esempio immagini che mettono in risalto alcune caratteristiche fisiche o aspetti della propria socialità (come alcuni eventi a cui si è partecipato), e scartarne altri, per esempio le foto che mostrano alcuni particolari fisici.

Al contrario, la comunicazione e la presentazione di Sé tradizionali, faccia a faccia, si realizzano attraverso uno scambio di informazioni che offre minore capacità di controllo. Ciò in quanto la comunicazione avviene in diretta e in presenza cioè attraverso scambi linguistici che prevedono intervalli temporali molto brevi tra loro e che coinvolgono la corporeità e la gestualità dei parlanti. In questo modo, è più difficile operare una selezione di ciò che si intende comunicare (e ciò che si vorrebbe mantenere privato) così come delle emozioni che si manifestano. Questo accade soprattutto perché nella comunicazione faccia a faccia i tempi non sono dilatati e quindi le risposte devono essere elaborate immediatamente (Walther, 1996). Di fronte a una domanda che risulta essere imbarazzante, ad esempio, può essere difficile riuscire a mascherare le risposte emotive suscitate, soprattutto quando esse vengono espresse corporalmente in modo involontario (si pensi all'arrossimento). Al contrario, la stessa domanda posta attraverso un messaggio di testo permette al mittente così come al ricevente di non svelare le proprie emozioni e di selezionare il messaggio più appropriato avendo più tempo a disposizione. Come sottolinea Sherry Turkle (2016)[1], un problema che riguarda questo depotenziamento emotivo nella comunicazione mediata è che esso potrebbe andare a scapito del legame interpersonale tra i comunicanti. Quello scambio di emozioni impegnativo ma talvolta imbarazzante che caratterizza la comunicazione faccia a faccia potrebbe infatti costituire un fattore di coesione molto importante nella costruzione dei rapporti personali.

[1] Ringrazio uno dei due reviewer anonimi per aver segnalato questo riferimento.

2.4. *La creazione di un proprio Sé digitale ideale*

Il fatto che la comunicazione tradizionale offra minor controllo rispetto agli scambi di informazioni che avvengono online e in particolare attraverso i social, implica anche dei costi più alti in termini emotivi ed attenzionali (Turkle, 2011; 2016). I social network, in questa prospettiva, offrono una forma di comunicazione e di scambio di informazioni che risulta essere meno impegnativa e più controllata, fornendo più tempo per elaborare i contenuti dei nostri messaggi, eliminando la comunicazione non verbale (corporea), e permettendo di selezionare più facilmente ciò che si intende comunicare. Ma il fatto che i social permettano di comunicare con gli altri selezionando più facilmente ciò che si desidera esprimere significa anche che, nella gestione dei contenuti, essi permettono agli utenti di costruire facilmente dei propri "Sé ideali" in formato social, cioè dei costrutti digitali che rappresentano in buona parte l'immagine di sé stessi che desiderano comunicare agli altri. Alcune indagini empiriche confermano che, in molti casi, gli utenti dei social pubblicano i contenuti che ritengono essere ideali per la presentazione di Sé (Cramer *et al*., 2016), almeno per un determinato tipo di pubblico.

L'ipotesi che i profili social costituiscano in molti casi delle autorappresentazioni idealizzate trova conferma anche negli effetti che il semplice "contemplare il nostro profilo" ha sugli utenti. Tale attività, infatti, sembra rappresentare una forma di affermazione di Sé che aumenta l'autostima e riduce le risposte difensive quando ci si trova in difficoltà. L'affermazione di Sé (o *self affirmation*) è stata definita come il processo attraverso cui ci concentriamo sugli aspetti di noi stessi che riteniamo essere maggiormente positivi. Toma e Hancock (2013) hanno studiato gli effetti del trascorrere cinque minuti contemplando il proprio profilo Facebook dopo una delusione accademica, constatando come essi siano analoghi a quelli di altre attività che hanno dimostrato sperimentalmente di essere "auto affermanti". Interrogati rispetto al loro fallimento, infatti, il gruppo di studenti che aveva trascorso cinque minuti osservando il loro profilo Facebook nel periodo immediatamente successivo al feedback negativo elaborava risposte in media meno difensive rispetto ad un gruppo di controllo. Accettando più facilmente il fallimento, essi mostravano così che la loro autostima era nel frattempo aumentata (si veda anche Toma, 2016), al contrario dei soggetti sperimentali che avevano speso lo stesso tempo guardano profili social di sconosciuti.

2.5. *Forme di apprezzamento digitali e gratificazione*

Le forme di apprezzamento che vengono espresse nelle piattaforme social presentano molte differenze rispetto a quelle tradizionali. Esse si realizzano in particolare attraverso il *like*, che costituisce una vera

e propria *affordance*, cioè una possibilità d'uso (Gibson, 1979), che i social mettono a disposizione degli internauti. Altre *affordances* di questi strumenti sono le accettazioni di amicizia, i commenti, i messaggi, ecc.

Le caratteristiche principali del *like* sono l'intrinseca positività – in quanto i significati dei *like* sono per lo più di tipo positivo – e la vaghezza. Un *like* non ha un significato prestabilito. Spesso si manifesta attraverso un pulsante che raffigura un pollice verso l'alto, che nel linguaggio corporeo occidentale moderno è un segnale di approvazione, oppure tramite un'icona a forma di cuore. Di conseguenza, il *like* assume significati diversi a seconda del tipo di contenuto a cui è associato, della relazione che esiste tra chi lo esprime e il ricevente, del momento in cui è stato espresso, ecc. (si veda Scissors, Burke e Wengrovitz, 2016). Per questo motivo è stato è stato definito come una *affordance* di tipo para-linguistico[2] (Hayes *et al*., 2016) e come un segnale fàtico (Hayes *et al*., 2018, si veda anche Fasoli, 2019, 84), cioè un segnale che non comunica un'informazione precisa ma vaga e che spesso viene usato per mantenere viva una relazione.

La vaghezza del *like* implica anche l'esistenza potenziale di un gap tra interpretazione del ricevente e significato attribuito da chi lo esprime. A causa di questo divario è possibile che il ricevente, in modo conscio o inconscio, attribuisca alla ricezione di *like* il significato che egli desidera o di elabori delle fantasie attorno ai significati che questo segnale, di volta in volta, assume. Forse anche a causa di questo meccanismo, il *like* è in grado di suscitare risposte emotive molto forti, attivando circuiti cerebrali legati al piacere, come il nucleo *accumbens* (Sherman *et al*., 2016; Manago *et al*., 2008). Ma a contribuire a questa capacità di suscitare risposte piacevoli, molto probabilmente, è anche il significato di riconoscimento sociale che ha l'accumulo di *like*, che costituisce una sorta di ricompensa sociale della propria cerchia di contatti.

I social network, dunque, sono dei canali comunicativi potenti, che permettono agli utenti di rendere immediatamente disponibili attraverso investimenti di tempo e energie relativamente bassi una enorme mole di informazioni di diverso tipo. Essi permettono di costruire un proprio Sé ideale, così come di *accedere* a un quantitativo estremamente elevato di informazioni che riguardano le persone della propria cerchia sociale (e non). Attraverso i social, poi, vengono scambiati grandi quantità di apprezzamenti vaghi (*like*) che si rivelano essere molto gratificanti per il nostro cervello. In virtù di tali meccanismi, è lecito sospettare che i social sollecitino in modo intenso quei meccanismi di gratificazione legati alla comunicazione, meccanismi che nel corso sella nostra evoluzione sono emersi come risposta ad un ambiente informazionale

[2] I fenomeni para-linguistici sono tratti concomitanti alla comunicazione verbale, che veicolano informazioni aggiuntive. Un fenomeno para-linguistico tipico della comunicazione verbale è il tono della voce.

caratterizzato dalla scarsità e che oggi invece si presenta come completamente diverso.

Da questo quadro provvisorio emerge come esistano dei fattori connaturati ai social, e quindi endogeni, che li rendono strumenti attraenti, perché in grado di fornire grandi ricompense informative ed emotive in cambio di piccoli investimenti temporali e di energie. Accedere a un'applicazione social e scorrere il *newsfeed*, così come pubblicare una foto o un contenuto, sono operazioni che richiedono poco tempo, non necessitano di competenze complesse e gratificano il nostro bisogno di comunicare informazioni su noi stessi e di acquisire informazioni sugli altri. Tornando alla prospettiva evolutiva da cui ha preso piede questa breve analisi preliminare, si noti che è probabile che a motivare la ricerca di ricompense, informazionali o alimentari, non sia stata solo la gratificazione che esse forniscono, ma anche la gratificazione che proveniva dal processo di ricerca stessa: «durante l'inseguimento, il cacciatore è motivato dalla caccia stessa» (Eyal, 2015, 84). In quest'ottica, la ricerca di informazioni che avviene scorrendo il *newsfeed* potrebbe non essere gratificante solo in virtù delle ricompense informazionali, ma anche in quanto costituisce una nuova attività di "caccia" in grado di offrire ricompense al nostro lato più "informivoro".

3. La comparazione sociale nei social network

Una conseguenza della sovrabbondanza di informazioni riguardanti le nostre cerchie sociali che i social network ci mettono a disposizione è una maggiore facilità della comparazione sociale. Se, infatti, attraverso queste piattaforme è possibile ottenere rapidamente molte più informazioni riguardanti le proprie cerchie rispetto al passato, ciò significa che è molto più facile istituire dei confronti tra le proprie condizioni di vita e quelle degli altri. Non è chiaro, tuttavia, quale sia il ruolo giocato dalla comparazione nelle dinamiche di attrattività delle piattaforme social, cioè quanto la possibilità di operare questi confronti continuamente sia causa o effetto del successo di queste piattaforme. Ciò in quanto la comparazione può essere orientata in entrambe le direzioni, cioè verso l'alto (verso chi è ritenuto avere un successo sociale maggiore) o verso il basso, cioè verso le persone della cerchia sociale che si ritengono avere una vita in qualche modo meno soddisfacente della propria (ad esempio con un lavoro o delle relazioni personali che ci appaiono peggiori). In quanto aperta a entrambe le direzioni, la comparazione sociale può avere effetti piacevoli o spiacevoli, che derivano ad esempio dalla percezione che gli altri si trovino in una situazione economica, sentimentale, lavorativa, migliore o peggiore della propria.

Per quanto riguarda gli scopi di questo articolo, ciò che sarebbe importante capire è se la comparazione sociale che i social permetto-

no abbia radici evolutive e quanto essa sia connaturata in noi. In altre parole, quanto – come esseri umani – siamo animali intrinsecamente mimetici, cioè portati a confrontare sistematicamente ciò che abbiamo e che facciamo con gli altri, e quale sarebbe l'eventuale funzione evolutiva di questa predisposizione? Purtroppo, a questa domanda non sembrano esserci risposte univoche, anche perché la comparazione sociale sembra essere un'operazione che nasce come conseguenza dello scambio informativo e che è di natura per lo più mentale. Diversamente dallo scambio di informazioni, che avviene in forma inter-soggettiva e si concretizza attraverso atti linguistici, testi o contenuti multimediali, la comparazione sociale è un'operazione per lo più intra-soggettiva che non necessariamente si manifesta esplicitamente.

Ciò che sembra plausibile, quanto meno a livello congetturale, è che la comparazione sociale verso l'alto abbia giocato la funzione di propulsione ad un miglioramento delle condizioni di vita e che in questo modo abbia contribuito attivamente alla sopravvivenza della specie. Banalmente, la spinta a comparare la propria condizione materiale con chi vive in una condizione migliore (cioè chi mangia cibo migliore del nostro, vive in alloggi migliori, ecc.) sembrerebbe dover generare anche una spinta verso il miglioramento della propria condizione economica, aumentando così la probabilità di sopravvivenza. Ad ogni modo, come è stato detto, è difficile capire quanto questo meccanismo stia alla base del successo dei social e quanto la comparazione sia semplicemente una conseguenza dell'aumento della disponibilità di informazioni.

4. Il design della dipendenza nei social network

Una volta descritti gli elementi strutturali dei social che li rendono attrattivi, veniamo ora all'analisi di quei fattori che contribuiscono a spiegare il successo di queste piattaforme ma che ritengo debbano essere considerati come esogeni. Essi costituiscono una conseguenza del modello economico che attualmente sta alla base della capacità di fare profitto delle piattaforme social. Tale modello è semplice: maggiore è la quantità di tempo che un utente spende su una piattaforma, più numerosi sono i dati che si possono raccogliere e poi rivendere riguardo al suo comportamento e alle sue preferenze, così come maggiore è il numero di inserzioni e contenuti pubblicitari che è possibile proporgli. L'ovvia conseguenza è che i proprietari delle piattaforme (non solo social ma più in generale digitali), negli ultimi anni, hanno cercato di individuare gli espedienti più efficaci in grado di rendere attraenti i loro prodotti, perché tale attrattività si traduce in tempo speso dagli utenti e quindi in profitti. In alcuni casi, essi hanno cercato di sfruttare le scienze comportamentali copiando e adattando gli elementi del design di alcuni giochi d'azzardo (per esempio le slot machine) che favorisco-

no il consolidamento di comportamenti compulsivi negli utenti (Eyal, 2015; Schull, 2016).

Uno dei meccanismi basilari su cui si basa il funzionamento delle slot machine – e che è stato implementato da alcuni social – è la cosiddetta ricompensa randomizzata, che venne studiata prima di tutti da Skinner. Variando in modo aleatorio il tempo che intercorre tra un comportamento (ad esempio il beccare una piccola finestrella di plexiglass da parte di un piccione) e una ricompensa (come un piccolo seme), Skinner dimostrò di essere in grado di spingere alcuni animali a ripetere in modo compulsivo quel comportamento, centinaia di volte in pochi minuti. Se consideriamo le notifiche che i nostri dispositivi digitali ci forniscono, possiamo notare che esse sono "naturalmente" randomizzate, dato che nella maggior parte dei casi non è possibile prevederle. Inoltre, a volte tali notifiche segnalano l'arrivo di stimoli che possono essere percepiti come positivi, ad esempio messaggi che si ritengono interessanti o *like* ad un contenuto, mentre in altri casi sono notifiche meno significative. Questa aleatorietà intrinseca sembra spiegare, in prima battuta, come mai è così facile ritrovarsi a controllare i propri dispositivi ad intervalli temporali estremamente ridotti. In alcuni casi il nostro comportamento – cioè il controllare lo schermo dei nostri dispositivi – viene ricompensato con un messaggio, un *like* ad un contenuto, una richiesta di contatto da parte di una persona che si ritiene essere importante nella propria rete sociale, una informazione che suscita interesse, ecc., mentre altre volte ciò non accade.

La randomizzazione delle ricompense, però, può essere artificialmente potenziata e favorita dai designer di alcune piattaforme, tra cui i social network, nel tentativo di aumentare il tempo di utilizzo degli utenti e la frequenza di accesso la homepage. Ad esempio, la costruzione delle bacheche dove vengono raggruppati i contenuti mostrati agli utenti può essere costruita implementando questo tipo di meccanismo. Piuttosto che ordinare i contenuti in ordine cronologico, cioè dal più recente al meno recente, oppure dal più significativo al meno significativo – in modo tale che gli utenti possano ritrovare le ricompense informative sempre all'inizio del loro *newsfeed* – questi contenuti possono essere somministrati in modo casuale. Nelle parole di Nir Eyal: «L'affascinante giustapposizione di cose rilevanti e irrilevanti, stuzzicanti e banali, belle e ordinarie, mette in agitazione il sistema della dopamina nel cervello, con la promessa di una ricompensa» (Eyal, 2015, 13).

In questo modo un comportamento, cioè lo scorrere dei contenuti nel *newsfeed*, viene ricompensato con delle informazioni importanti in modo strutturalmente aleatorio. Tali contenuti non vengono quindi ordinati né cronologicamente né per rilevanza, ma vengono invece mescolati in modo da rendere impossibile prevedere se una informazione che risulta essere interessante apparirà in cima alla lista o più in basso. La randomizzazione può essere implementata anche a livelli diversi.

Per esempio, immaginiamo che un utente consulti il proprio *newsfeed* in un social network a dieci minuti di distanza. È possibile che esso sia totalmente coincidente rispetto al precedente, cioè che offra gli stessi contenuti, che offra solo contenuti leggermente diversi oppure contenuti totalmente nuovi. Queste stesse opzioni possono realizzarsi in modo randomizzato, rendendo così impossibile prevedere se consultando il *newsfeed* di un social più volte a distanza di poco tempo si verrà ricompensati ottenendo o meno informazioni nuove, interessanti, gratificanti oppure totalmente ridondanti.

Un secondo meccanismo che è stato implementato da molti social network per favorire un maggiore consumo da parte degli utenti è quello del *newsfeed* infinito. Molto spesso, le pagine web hanno una lunghezza stabilita. In questi casi, sullo schermo appare solo una porzione della pagina, che è visionabile completamente spostando la barra di navigazione verso il basso (anche attraverso lo scroll) in modo da mostrare altri contenuti, fino a giungere ad un termine. Alcuni siti, tuttavia, da alcuni anni hanno modificato la loro struttura. Nel momento in cui la barra laterale viene spostata verso la parte inferiore, che coincide con il termine della pagina web, nuovi contenuti vengono automaticamente inseriti ed essa subisce uno spostamento quasi impercettibile verso l'alto. La pagina diventa in tal modo potenzialmente "infinita", senza fondo. Questo espediente apparentemente privo di conseguenze, come confessato dal suo inventore[3], trae ispirazione da uno studio comportamentale che si concentra sul modo in cui interpretiamo i segnali visivi esterni per stabilire qual è il consumo di cibo che riteniamo essere soddisfacente e in grado di saziarci.

L'esperimento in questione (Wansik, Painter e North, 2005) prevedeva che gruppi di quattro soggetti (sconosciuti) si ritrovassero a consumare assieme un piatto di zuppa in un locale. Ad insaputa dei partecipanti, due delle quattro ciotole erano però "truccate", in quanto fissate al tavolo e collegate ad un sistema nascosto attraverso cui si aggiungevano piccole porzioni di zuppa nel piatto, senza che i commensali se ne accorgessero. Ai soggetti, a cui era vietato toccare le ciotole, non venivano forniti in precedenza i dettagli dello studio, ma venivano poste delle domane sui colori delle ciotole (che erano diverse) allo scopo di spingerli a credere che l'esperimento si stesse concentrando sulle influenze cromatiche. Una volta fatti sedere, veniva detto loro di mangiare quanto lo desideravano e di conversare con gli altri. Ai due partecipanti che utilizzavano la ciotola truccata, quindi, venivano aggiunte piccole porzioni di zuppa a distanza di brevi intervalli temporali, senza che se ne accorgessero. Al termine del pasto venivano quindi somministrati dei questionari riguardanti la percezione di quanto avevano mangiato e il loro senso di sazietà. I risul-

[3] https://www.bbc.com/news/technology-44640959, ultima consultazione 15 luglio 2019.

tati misero in evidenza due dati particolarmente interessanti. I soggetti che usavano la ciotola truccata mangiavano in media il 73% in più degli altri, ma non stimavano di aver mangiato più di loro né dichiaravano di essere più sazi (Wansik *et al*., 2005, 96-97).

L'esperimento citato mostra come «invece di monitorare quanto stiamo mangiando, ci affidiamo a segnali visivi o regole generali (come mangiare fino a quando la ciotola è vuota) per determinare quanto consumare» (p. 98). In questo senso, sembrerebbe realizzarsi una sorta di euristica che, nell'esperimento, potrebbe funzionare particolarmente bene anche per motivi contestuali. La conversazione con tre sconosciuti potrebbe essere impegnativa, tanto da rendere più difficile un monitoraggio intrapersonale del proprio senso di sazietà. Inoltre, la zuppa potrebbe essere un cibo leggero e non particolarmente saziante (o almeno non tanto quanto altri cibi), che può essere ingerito facilmente in dosi elevate. In ogni caso, l'esperimento sembra suggerire anche che la sazietà, che intuitivamente potrebbe apparire indipendente dai processi cognitivi superiori, può subire in realtà influenze di tipo *top down*.

5. Trappole e trappole cognitive

Tornando alle bacheche dei social network, possiamo quindi comprendere come l'assenza di un limite della pagina web costituisca anche l'assenza di un riferimento visivo del "consumo" di post, cioè di informazioni, degli utenti. Una volta costruito un *newsfeed* infinito "quantificare" il numero di contenuti che sono stati visualizzati risulta essere molto più difficile, perché l'unico riferimento possibile in questa operazione mentale è quello temporale, cioè quanto tempo abbiamo speso a scorrere la bacheca, mentre quelli visuo-spaziali vengono continuamente boicottati. La posizione della barra laterale destra, che di solito offre un segnale di riferimento indicando quanto siamo vicini al termine di una pagina web, perde invece gran parte della sua capacità di orientamento, non essendo mai previsto un termine.

Si può presumere che l'efficacia di questo meccanismo nell'aumentare il tempo speso su queste piattaforme da parte degli utenti dipenda da fattori diversi. Per esempio, dalla frequenza con cui gli utenti visualizzano da cima a fondo le pagine web che invece non implementano questo meccanismo. A proposito, va notato che la maggior parte dei siti web (al di fuori delle piattaforme social) hanno una struttura finita, e che questo potrebbe spingere gli utenti a essere più cognitivamente vulnerabili ai *newsfeed* infiniti.

Da un punto di vista teorico, questi espedienti possono essere efficacemente considerati delle vere e proprie trappole cognitive (Seaver, 2018). Il concetto di "trappola", infatti, sembra essere molto efficace da un punto di vista esplicativo, una volta abbandonata la visione *naif*

delle trappole come meri dispositivi brutali, coercitivi. Tale concezione in realtà trascura la sua complessità e alcune sue caratteristiche essenziali, in particolare il fatto che ogni trappola presuppone una teoria della mente della preda. Una trappola, infatti, «deve persuadere la sua preda a svolgere il ruolo scritto per essa dal suo design» (Seaver, 2018, 6; si veda anche Mason, 1900, 659-660). Se la preda, animale o umano che sia, non gioca il ruolo prestabilito, la trappola non funziona. Considerando la complessità del comportamento di molti animali che sono stati cacciati nella nostra storia, è evidente come la trappola richieda prima di tutto un'ottima conoscenza del comportamento della preda. In questo senso, la trappola è un artefatto attraverso cui si materializzano contemporaneamente una teoria della mente e una strategia cognitiva, che possono essere di notevole sofisticazione.

Tornando all'ambito digitale e ai social network, concepire gli espedienti cognitivi descritti in precedenza come trappole significa attribuire un senso molto più letterale alle espressioni "catturati da questa piattaforma" o da qualche gioco, applicazione, ecc. Le trappole cognitive sarebbero quindi degli espedienti sviluppati sulla base di una articolata teoria della mente umana e volti a catturare l'attenzione degli utenti nel modo più efficace possibile. In questo senso, le trappole cognitive spesso nascono per motivi commerciali e tendono a sfruttare gli aspetti della nostra vita mentale che tendiamo di solito a trascurare, come i bias e le euristiche (si veda ad esempio Motterlini, 2008).

Esistono, ovviamente, alcune peculiarità delle trappole cognitive che si realizzano sui social e sulle piattaforme online rispetto alle trappole per animali. Mentre per il cacciatore la preda dev'essere catturata in senso fisico, per il progettatore di piattaforme il bersaglio primario è l'attenzione dell'utente social. Non si tratta, però, di un obiettivo che viene raggiunto in modo permanente ma di una cattura temporanea, limitata ad un intervallo di tempo. Lo scopo delle trappole cognitive implementate nei social è quindi duplice: da un lato esse cercano di motivare l'utente a ritornare il più spesso possibile sulle piattaforme, dall'altro lato, una volta "atterrati" sulla piattaforma, l'obiettivo diventa quello di trattenerli il più possibile. In quanto tale, probabilmente essa richiede costantemente nuovi espedienti per essere ripetuta e rinnovata.

Considerando la complessità del comportamento umano, è chiaro che le trappole cognitive che vengono progettate *online*, o almeno quelle più efficaci e sofisticate, possono essere create solo assumendo pienamente quel modello complesso di mente come "prevedibilmente irrazionale" (Ariely, 2008) che è stato elaborato dalle scienze cognitive contemporanee.

5.1. *Big data e trappole cognitive personalizzate*

Al di là dei due esempi, viene spontaneo chiedersi se non siano presenti altre "trappole cognitive", magari più efficaci ma invisibili agli occhi degli utenti e anche agli esperti di scienze cognitive. Se, come ho sostenuto, le trappole cognitive si basano sempre su una teoria della mente, e se le aziende proprietarie dei social sembrano aver ben assimilato il modello di mente che le scienze cognitive e comportamentali hanno elaborato negli ultimi anni, è lecito quanto meno sospettare che esistano altre trappole cognitive più sofisticate e trasparenti. Poniamo, per ipotesi, che sfruttare il *bias* della conferma, mostrando nelle bacheche più spesso i contenuti che confermano le credenze degli utenti rispetto a quelli che le contraddicono, possa spingerli ad aprire più spesso la homepage. Perché mai queste aziende dovrebbero esimersi dal farlo, rinunciando a possibili guadagni, solo perché una trappola del genere si rivelerebbe deleteria per la nostra capacità critica?

Inoltre, i big data disponibili ai giganti del web e la possibilità di personalizzare i contenuti permetterebbero a queste aziende anche di andare facilmente oltre le teorie cognitive oggi disponibili e al di là dei fenomeni maggiormente noti. Per esempio, l'analisi di questi dati potrebbe mettere in evidenza che per alcuni utenti risulta essere più efficace, in termini di aumento d'uso, una leggera sovra-esposizione a contenuti emotivamente salienti piuttosto che una sovraesposizione ai contenuti che confermano il proprio set di credenze. In generale, i big data conferiscono alle piattaforme social un potere previsionale sul comportamento degli utenti estremamente elevato, in quanto, in un certo senso, permettono di sviluppare delle teorie della mente accurate ritagliate su gruppi molto ristretti di utenti. Tornando alla metafora della caccia e della trappola, è come se un cacciatore avesse a disposizione un database molto accurato dei dati relativi ai comportamenti delle proprie prede, e potesse progettare trappole diverse per ognuna di esse.

Il fatto che le informazioni che riteniamo essere attraenti e interessanti, e che ricaviamo dalle attività di ricerca nei social, rappresentino delle ricompense informazionali, così come i *like* costituiscono delle ricompense sociali, ha anche un'altra delicata conseguenza. Se, infatti, attraverso il loro algoritmo i social hanno il potere di modificare la frequenza e la quantità delle ricompense sociali e informazionali che otteniamo, essi hanno anche il potere di ricompensare o punire alcuni *comportamenti* degli utenti. Ad esempio, essi hanno il potere di mettere in evidenza o di nascondere un contenuto particolare nelle bacheche degli altri utenti, di farlo comparire una volta sola o più volte, incidendo così anche sul numero di *like* e di interazioni che esso riceve. Espedienti di questo tipo, in linea teorica, sembrerebbero poter essere impiegati anche per rafforzare alcuni comportamenti, ad esempio per favorire la pubblicazione di alcuni contenuti piuttosto che altri.

Evidenziare questo tipo di possibilità non significa assumere un atteggiamento necessariamente complottista, che presuppone una manipolazione pesante da parte di chi controlla i social, ma vuol dire semplicemente dedurre delle possibilità tecniche che è bene quanto meno discutere e tenere in considerazione quando si analizza il potere e le responsabilità etiche di questi strumenti. Da questo punto di vista, un grosso problema che riguarda le autorità e la protezione degli utenti è l'impossibilità strutturale di monitorare questi processi. Ogni *newsfeed* è diverso dagli altri, proprio in quanto personalizzato, ed è difficile da controllare perché richiederebbe una registrazione delle attività dello schermo durante lo scorrimento, che è poco utilizzato e, tra l'altro, richiederebbe molta memoria nei propri dispositivi. Il risultato è che ogni *newsfeed* appare e scompare senza lasciare nessuna traccia, ma ciò conferisce ai gestori dei social una enorme libertà di gestione (e quindi anche di potenziale manipolazione) dei contenuti nelle singole bacheche.

6. Conclusioni

Le grandi aziende tecnologiche, contrariamente a molte istituzioni contemporanee, hanno abbandonato quel modello olimpico della razionalità che è stato demolito teoricamente e empiricamente dalle scienze cognitive moderne, assumendo quel nuovo modello di mente che la concepisce come sistematicamente irrazionale ma prevedibile (Ariely, 2008). Nel fare ciò, esse hanno progettato strumenti come i social network che hanno avuto un successo enorme, alla cui base stanno dei fattori endogeni e dei fattori esogeni.

Tuttavia, per ragioni economiche, negli ultimi anni esse hanno partecipato ad una vera e propria corsa all'implementazione del cosiddetto design della dipendenza (Eyal, 2015). Tale design cerca di individuare e riprodurre alcuni meccanismi tipici di oggetti come le slot machine, con lo scopo di favorire il più possibile comportamenti compulsivi negli utilizzatori. In questo articolo, ho attribuito alle trappole cognitive che nascono dalla logica del profitto una parte non trascurabile della capacità di attrazione che le piattaforme social esercitano sugli utenti. A monte, però, ho mostrato come esse siano prima di tutto strumenti in grado di soddisfare i bisogni informativi ancestrali che abbiamo ereditato dai nostri predecessori.

Si giunge così ad una prospettiva profondamente paradossale. Agganciandosi alla nostra predisposizione evolutiva verso la raccolta di informazioni che riguardano le nostre cerchie sociali, cioè facendo leva sulla nostra natura di informiveri sociali, i social network ci attraggono fornendoci una nuova possibilità di "cacciare informazioni", che consiste essenzialmente nel consultare le nostre bacheche online. Specularmente, essi ci danno anche la possibilità di "dare in pasto" a molti informiveri

della nostra cerchia sociale una mole enorme di informazioni che ci riguardano. Paradossalmente, però, questa forma di caccia risulta essere ancora più attraente proprio perché nasconde diverse trappole cognitive. In quanto informiveri i social ci rendono quini contemporaneamente cacciatori (di informazioni) e prede (delle trappole cognitive implementate dai social stessi). Comprendere se sia possibile riuscire a sfuggire alle trappole cognitive che i social implementano e quali tecniche possano essere efficaci per neutralizzarle sembra essere complesso, soprattutto perché si tratta di meccanismi che agiscono in modo inconscio. Per quanto riguarda il sovra-utilizzo, è probabile che un ruolo regolatorio importante sia giocato dalle abitudini e dalle strategie meta-cognitive, come accade per quanto riguarda la capacità di controllo nel caso del Marshmallow dell'esperimento di Mischel (Mischel, Shoda e Rodriguez, 1989). Resta da capire, però, se e come sia possibile neutralizzare specifici meccanismi che fungono da trappole e che intervengono in attività che non richiedono un livello di attenzione elevato e che spesso, anzi, sono eseguite in modalità multitasking.

Riferimenti bibliografici

Ainslie, G. (2001). *Breakdown of will*. Cambridge: Cambridge University Press.

Andreassen, C.S. (2015). Online social network site addiction: A comprehensive review. *Current Addiction Reports*, *2*, 2, pp. 175-184.

Ariely, D. (2008). *Predictably irrational*. New York: HarperCollins.

Bianchi, M. (2004). Se la felicità è così importante, come mai ne sappiamo così poco?. In L. Bruni e P. Porta, *Felicità ed economia. Quando il benessere è ben vivere*. Milano: Guerini e Associati, pp. 170-191.

Cramer, E.M., Song, H., Drent, A.M. (2016). Social comparison on Facebook: Motivation, affective consequences, self-esteem, and Facebook fatigue. *Computers in Human Behavior*, *64*, pp. 739-746.

Dunbar, R.I. (2004). Gossip in evolutionary perspective. *Review of general psychology*, *8*, 2, 100.

Eyal, N. (2015). *Catturare i clienti (Hooked)*. Milano: Edizioni LSWR.

Fasoli, M. (2019). *Il benessere digitale*. Bologna: Il Mulino.

Freedberg, D. (1989). *The power of images: Studies in the History and Theory of Response*. Chicago: University of Chicago Press.

Freedberg, D., Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in cognitive sciences*, *11*, 5, pp. 197-203.

Gui, M. (2014). *A dieta di media*. Bologna: Il Mulino.

Hayes, R.A., Carr, C.T., Wohn, D.Y. (2016). One click, many meanings: Interpreting paralinguistic digital affordances in social media. *Journal of Broadcasting & Electronic Media*, *60*, 1, pp. 171-187.

Hayes, R.A., Wesselmann, E.D., Carr, C.T. (2018). When Nobody "Likes" You: Perceived Ostracism Through Paralinguistic Digital Affordances Within Social Media. *Social Media+Society*, *4*, 3, 2056305118800309.

Manago, A.M., Graham, M.B., Greenfield, P.M., Salimkhan, G. (2008). Self-pre-

sentation and gender on MySpace. *Journal of Applied Developmental Psychology*, *29*, 6, pp. 446-458.
Mason, O.T. (1900). Traps of the Amerinds. A Study in Psychology and Invention. *American Anthropologist*, *2*, 4, pp. 657-675.
Miller, G.A. (1984). Informavores. In F. Machlup e U. Mansfield, *The study of information: Interdisciplinary messages*. New York: Wiley-Interscience, pp. 111-113.
Mischel, W., Shoda, Y., Rodriguez, M.I. (1989). Delay of gratification in children. *Science*, *244*, 4907, pp. 933-938.
Motterlini, M. (2011). *Trappole mentali*. Milano: Rizzoli.
Ofcom (2016). *The Communications Market Report*, https://www.ofcom.org.uk/research-and-data/multi-sector-research/cmr/cmr16.
Paglieri, F. (2014). *Saper aspettare*. Bologna: Il Mulino.
Pasquinelli, E. (2012). *Irresistibili schermi: fatti e misfatti della realtà virtuale*. Milano: Mondadori.
Pinker, S. (2003). *The language instinct: How the mind creates language*. London: Penguin.
Rigutti, S., Gerbino, W., Fantoni, C. (2016). Websites Eco-Usability. *Sistemi intelligenti*, *28*, 2-3, pp. 343-362.
Schull, N. (2016). *Architetture dell'azzardo: Progettare il gioco, costruire la dipendenza*. Bologna: Luca Sossella Editore.
Scissors, L., Burke, M., Wengrovitz, S. (2016). What's in a Like?: Attitudes and behaviors around receiving Likes on Facebook. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pp. 1501-1510.
Seaver, N. (2018). Captivating algorithms: Recommender systems as traps. *Journal of Material Culture*, 1359183518820366.
Sherman, L.E., Payton, A.A., Hernandez, L.M., Greenfield, P.M., Dapretto, M. (2016). The power of the like in adolescence: Effects of peer influence on neural and behavioral responses to social media. *Psychological science*, *27*, 7, pp. 1027-1035.
Toma, C.L. (2016). Taking the good with the bad: effects of Facebook self-presentation on emotional well-being. In L. Reinecke e M.B. Oliver, *The Routledge Handbook of Media Use and Well-Being*. London: Routledge, pp. 170-182.
Toma, C.L., Hancock, J.T. (2013). Self-affirmation underlies Facebook use. *Personality and Social Psychology Bulletin*, *39*, 3, pp. 321-331.
Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York, Basic Books.
Turkle, S. (2016). *Reclaiming conversation: The power of talk in a digital age*. New York: Penguin.
Walther, J.B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication research*, *23*, 1, pp. 3-43.
Wansink, B., Painter, J.E., North, J. (2005). Bottomless bowls: why visual cues of portion size may influence intake. *Obesity research*, *13*, 1, pp. 93-100.
Wolf, M. (2009). *Proust e il calamaro. Storia e scienza del cervello che legge*. Milano: Vita e Pensiero.

Hunters (of information) and preys (of cognitive traps) in web 2.0: An analysis of the attractivity of social networks from a cognitive-evolutionary perspective

The article provides an explanation of the current success of social platforms from a cognitive-evolutionary perspective. On the one hand, some elements that make them particularly attractive to users – and that therefore may be considered as intrinsic (or endogenous) to these sites – are identified. For instance, the ability of social networks to provide us with an enormous amount of information about our social circle seems to be a salient element. This is because, in our evolutionary history, the possession of such information offered advantages in terms of survival. On the other hand, it is possible to identify some technical expedients often implemented by these platforms (e.g. the randomisation of rewards, the construction of an infinite newsfeed) that try to "capture" users' attention and that can be described as cognitive traps. These traps are exogenous elements, namely they are not part of the intrinsic nature of these platforms but are implemented through specific design choices for economic reasons.

Keywords: social networks, cognition, evolution, traps, cognitive traps, addiction design, digital over-consumption.

Marco Fasoli, Università di Milano-Bicocca, Dipartimento di Sociologia e della ricerca sociale, Piazza dell'Ateneo Nuovo 1 - 20126 Milano, marco.fasoli@unimib.it

FRANCESCA ERVAS

METAFORE VISIVE, COMUNITÀ IMMAGINATE E RAZIONALITÀ DIFFERITA

1. Le comunità immaginate nei social network

In *Imagined Communities*, Benedict Anderson così definiva una nazione: «si tratta di una comunità politica immaginata, e immaginata come intrinsecamente insieme limitata e sovrana» (1983, 24). La nazione è innanzitutto una comunità, perché «malgrado ineguaglianze e sfruttamenti di fatto che possono predominarvi, la nazione viene sempre concepita in termini di profondo, *orizzontale cameratismo*» (1983, 26, corsivo mio). È una comunità immaginata come «limitata» perché, per quanto grande, non s'immagina mai coincidere con l'intera umanità ed è dunque consapevole dell'esistenza di limiti oltre ai quali esistono altre comunità. È una comunità immaginata come «sovrana» perché, consapevole della propria identità e dell'esistenza di un pluralismo di identità diverse dalla propria, si immagina libera.

Il pensiero di Anderson sulle comunità immaginate viene qui ripreso perché si ritiene che possa essere ancora attuale e utile per pensare alle comunità create in seno agli odierni social network come *Facebook* (FB) (Acquisti e Gross, 2006). Mutatis mutandis, gli internauti costituiscono nei social network delle comunità («Gruppi» o «Fan page» nella terminologia di FB) che si percepiscono, o meglio, si immaginano come limitate e autonome: ciascun gruppo ha un numero determinato (per quanto estendibile) di membri che condividono certi interessi, credenze, opinioni, ecc. e sono consapevoli dell'esistenza di altri gruppi caratterizzati da diversi interessi, credenze, opinioni, ecc. Malgrado le possibili e inevitabili differenze tra i membri del gruppo, anche le relazioni sociali delle comunità FB sono connotate da un simile «orizzontale cameratismo»: i contatti di ciascun membro sono detti «amici» (*friends*) e ciascun membro di un gruppo è libero di inviare («postare») immagini, messaggi verbali o audio-visivi («post») o di rispondere con immagini, messaggi verbali o audio-visivi ai post altrui, secondo le regole condivise e accettate da tutti coloro che aderiscono al gruppo (Caers *et al*., 2013). Attraverso le relazioni sociali create «in rete» su temi e problemi condivisi, ciascun

membro del gruppo può farsi un'immagine della propria comunità, sebbene molti dei membri di un gruppo FB (a volte persino tutti) non si siano mai visti. In questo senso si può pensare che i gruppi FB siano una sorta di «comunità immaginata».

Non a caso Anderson aveva scelto il termine «immaginata», e non «immaginaria» o «inventata». Il concetto di comunità immaginata era in esplicita contrapposizione al concetto di nazione proposto da Ernest Gellner (1964), per cui il nazionalismo aveva «inventato» le nazioni dove non esistevano. Gli abitanti della più piccola nazione sono invece reali: la comunità è immaginata perché gli abitanti «non conosceranno mai la maggior parte dei loro compatrioti, né li incontreranno, né ne sentiranno mai parlare, eppure nella mente di ognuno vive l'immagine del loro essere comunità» (Anderson, 1983, 25). La comunità è dunque immaginata solo perché supera quella «taglia limite» che consente ai suoi abitanti di vedersi faccia a faccia. Segue da questa definizione che una nazione può essere un villaggio o un intero continente, e anche, nel senso sopra descritto, un gruppo FB, per quanto ad un gruppo FB non corrisponda un'estensione geografica. I membri del gruppo FB esistono nella realtà, altrimenti non potrebbero accedere a FB («registrarsi») e a uno specifico gruppo, e infine «postare» in quel gruppo. Esistono naturalmente profili non-autentici (*fake*), ma – oltre al fatto che un intero gruppo FB di profili *fake* sarebbe disfunzionale al compito di ingannare i membri autentici – ciò non impedisce al gruppo FB di costituirsi come «comunità immaginata» nella mente di ciascuno dei suoi membri autentici.

Ad ogni modo, spiega Anderson, le comunità non si distinguono tanto per l'ordine di grandezza o per l'estensione geografica, né tantomeno in base alla loro presunta «autenticità» o «falsità». Si distinguono invece in base allo *stile in cui sono immaginate* (Anderson, 1983, 25). Ciascuna comunità si immagina attraverso le metafore e i miti di cui dispone, e più in generale, attraverso i linguaggi che crea e le memorie che condivide tramite determinate tecnologie. Le nazioni, come comunità immaginate, erano nate dalla convergenza storica del capitalismo e della tecnologia di stampa sulla «fatale diversità delle lingue umane». In particolare, in Occidente, quello che Anderson chiama «capitalismo della stampa» (*print capitalism*) ha reso possibile nel XV secolo la traduzione, la diffusione e dunque la condivisione – attraverso linguaggi e codici comunicativi altri rispetto al Latino istituzionale – di nuovi sistemi di credenze, ma soprattutto di nuove immagini della realtà. I membri delle comunità immaginate di FB dispongono di una nuova tecnologia (computer e connessione internet) necessaria per accedere alla rete virtuale che li mette in contatto. La nuova tecnologia dà loro la possibilità di inviare e condividere messaggi in una pluralità di formati diversi e interposti (testi verbali, audio, immagini, video), ma soprattutto – diversamente da tecnologie precedenti – di poterli commentare e discutere online «in diretta». Se si dovesse dunque caratterizzare lo stile attraverso cui ven-

gono immaginate le comunità FB, diremmo non solo che è *per natura multimodale* ma anche che è un'opera di *traduzione collettiva simultanea* di contenuti multimodali, ovvero una continua riformulazione di contenuti multimodali online da parte del gruppo di utenti. Un esempio di tale opera di traduzione collettiva simultanea sono gli stessi «post» di formato multimodale, spesso condivisi con altri membri del gruppo FB e commentati e (ri)discussi in formato multimodale dai membri del gruppo FB in un *thread* «in diretta», potenzialmente infinito.

Secondo Anderson, il processo di vernacolarizzazione messo in atto dai mezzi di comunicazione di massa, a partire dalla stampa, ha inizio dalla traduzione della Bibbia in tedesco ad opera di Martin Lutero (1522-1534) – ma pensiamo anche a quella contemporanea in lingua inglese di William Tyndale (1526). La traduzione è vista qui non tanto in senso proprio, come atto del tradurre un singolo testo da una lingua a un'altra, quanto piuttosto in senso esteso (Ervas, 2008), come strumento di ibridazione attraverso il quale ci si confronta con la diversità della lingua, della storia e della cultura di una comunità linguistica, ma nello stesso tempo la si integra nella propria comunità linguistica. In un certo senso anche oggi le lingue e i linguaggi utilizzati dalle comunità reali «risentono» dello (e sono forgiati dallo) stile delle comunità immaginate nei social network: testi brevi, commenti immediati, ricchezza di immagini ed elementi fatici.

Nel prosieguo di questo saggio si cercherà di capire quando e perché nei social network, e in FB in particolare, una «comunità immaginata» può diventare una «comunità immaginaria» (§ 2). L'idea che guida questo lavoro è che proprio per la natura dello stile con cui le comunità FB vengono immaginate, esse siano più esposte, rispetto alle comunità immaginate del passato, al rischio di degenerare in comunità immaginarie o inautentiche (§ 3). Si prenderà in esame lo stile con cui le comunità FB vengono immaginate, dedicando particolare attenzione a come vengono comunicate le immagini, sotto forma di metafore visive (mono o multimodali) (§ 4). Diversi *frames* teorici saranno utilizzati per spiegare i meccanismi concettuali (Lakoff e Johnson, 1980) e gli effetti immaginistici (Carston 2010, 2018) e persuasivi (Black, 1955) delle metafore visive. In particolare, si cercherà di mostrare che, paradossalmente, lo stile in cui si immaginano le comunità FB lascia meno spazio all'immaginazione dei singoli utenti FB e mina la razionalità argomentativa relazionale (Mercier e Sperber, 2017), sfociando spesso in una sorta di «razionalità differita», in cui – invece di anticipare un possibile disaccordo nella discussione online – si rimanda il disaccordo a un momento che potrebbe non darsi mai (§ 5). Si porteranno infine due esempi di metafore visive, commentate nei social network, come casi di studio di razionalità differita (§ 6) e se ne trarranno le conclusioni (§ 7).

2. Comunità reali, immaginate e immaginarie

Centrale per il pensiero di Anderson sulle comunità immaginate è l'*aspetto performativo* dei loro linguaggi o, più precisamente, dello stile con cui vengono immaginate: è esattamente l'evocazione di quell'insieme di pratiche linguistiche, metafore, simboli che rimandano a credenze condivise che crea quella moderna comunità chiamata nazione. Ciò non vuol dire – come sopra ricordato – che la comunità immaginata non abbia fondamento nella realtà. Anzi, al contrario, la comunità affonda le proprie radici nella realtà dei singoli membri ma la travalica, tramite processi di immaginazione. Gellner, nemico acerrimo della filosofia linguistica di Oxford, aveva invece assimilato «"invenzione" a "fabbricazione" e "falsità", piuttosto che a "immaginazione" e "creazione"» (Anderson, 1983, 25). Anderson osserva invece che il tipo di processo immaginativo in atto nella costruzione della comunità immaginata non ha niente di falso, fabbricato o inventato, perché tutti gli individui della comunità si immaginano la comunità in base a proprietà che sono comuni e condivise nelle comunità reali. Le stesse comunità immaginate nei social network non sono disancorate dalla realtà: per esempio, la maggior parte delle notizie condivise su FB (70%) provengono da amici o membri della propria famiglia, ovvero da persone di cui gli utenti hanno già fiducia nella «comunità reale», piuttosto che da nuovi gruppi o organizzazioni che gli utenti seguono sulla loro pagina FB (Oeldorf-Hirscha e Sundar, 2015). Sebbene ci sia un numero massimo di cinquemila contatti o amici per profilo, comunque gli utenti riescono e di fatto preferiscono gestire un numero di contatti inferiore, di circa 150/200 persone (Dunbar, 2016), ovvero un numero molto vicino a quello delle comunità reali di persone che si possono riconoscere e incontrare faccia a faccia. Nonostante ciò, come si intende argomentare, le comunità immaginate su FB sono più esposte al rischio di degenerare in comunità immaginarie, proprio per lo stile in cui vengono immaginate.

Nelle comunità reali ciascun membro della comunità può interagire con l'altro in forma dialogica, avanzando le proprie richieste, opinioni, commenti, spiegazioni e ragioni e aspettando dall'altro membro (o dagli altri membri) della comunità delle risposte che moduleranno o comunque inevitabilmente influenzeranno le successive richieste, opinioni, commenti, spiegazioni e ragioni. Il confronto e talvolta lo scontro con qualcuno che ha opinioni diverse sta alla base sia delle singole interazioni costruttive per i singoli membri di una comunità reale, sia della crescita della comunità intera. Proprio nelle discussioni in cui emerge un disaccordo i membri della comunità cercano delle ragioni per difendere le proprie idee o per contrastare quelle altrui e, così facendo, migliorano e raffinano non solo le proprie posizioni ma anche quelle (possibilmente diverse) di tutta la comunità. Nelle comunità FB ci sono i mezzi tecnologici per dialogare in modo simile a quanto avviene

nelle comunità reali, ma di fatto – non avendo il proprio interlocutore di fronte – c'è sempre la possibilità che risponda molto più tardi, non acceda al proprio profilo FB, acceda al proprio profilo FB ma non legga il commento, o addirittura non risponda mai. Ciò sicuramente influisce non solo sull'interazione che il singolo utente FB ha con i suoi contatti, sul suo modo di conoscere, di argomentare e di comunicare, ma anche sullo stile in cui si immagina la propria comunità FB.

Il rischio che corre il singolo membro della comunità immaginata di FB è quello di ritrovarsi a essere «solo» o a pensare e ragionare in vista di un possibile disaccordo in una discussione che potrebbe non arrivare mai. A tal proposito Mercier e Sperber (2017) notano che non c'è nulla di male nel «ragionamento solitario» che anticipa un dialogo, se tale dialogo avrà effettivamente luogo. Il ragionamento solitario diventa tuttavia problematico se rimane solitario:

> Gli ambienti moderni influenzano la nostra capacità di *anticipare il disaccordo*. [...] Per esempio, prima dell'invenzione della stampa e dell'avvento dei media moderni, le persone erano solitamente consapevoli che qualcuno del loro gruppo aveva opinioni diverse dalle loro grazie all'interazione con quella persona. Accadeva spesso di trovare una differenza di opinione e cercare di risolverla attraverso ripetuti scambi di argomenti che potevano essere anticipati e mentalmente inscenati. Oggi siamo inondati di opinioni di persone che non incontreremo mai: editorialisti, conduttori, bloggers. Ci si aspetta anche che abbiamo un'opinione su molti temi diversi – dalla politica alla musica e al cibo – e che siamo capaci di difendere questa opinione quando è messa in discussione, dandoci delle ragioni per preparare una varietà di dibattiti che non accadranno mai (Mercier e Sperber, 2017, 249-250, trad. mia, corsivo mio).

Anderson vedeva infatti proprio nell'avvento della stampa uno dei principali motivi della costruzione delle nazioni come comunità immaginate. In effetti la stampa metteva le idee dell'autore del testo in mano a persone che non avrebbe mai incontrato e con cui dunque non avrebbero mai discusso. La stampa ha dato un forte impulso anche a quell'opera di «traduzione collettiva» che sta alla base delle comunità immaginate, a cui hanno partecipato tanti traduttori e a cui hanno risposto le intere comunità reali. Tuttavia non si trattava di un'opera di traduzione collettiva simultanea: le traduzioni nella lingua volgare nazionale richiedevano tempo e ancora più tempo serviva alla lingua e alla cultura tradotta per entrare nella lingua volgare nazionale.

Il caso dei social network sembra diverso dalle comunità del capitalismo della stampa: si fonda su una tecnologia che dà la possibilità di interagire con la persona che forse non si incontrerà mai, sebbene la limiti in molti modi, per esempio per formato e quantità di contenuti. Il processo di traduzione dei linguaggi usati in rete – e non solo del linguaggio verbale – ha dunque la possibilità di essere un'opera di *traduzione collettiva simultanea* che costituisce la modalità specifica o lo stile in

cui si immaginano le comunità dei social network. Ci si aspetta che ciascun membro della comunità immaginata partecipi a questa traduzione nei tempi velocissimi concessi dalla tecnologia, in «diretta». Tuttavia il possibile disaccordo dell'interlocutore può essere «differito» a un *tempo imprecisato*: l'interlocutore potrebbe rispondere domani, come accade sovente, ma anche fra un mese, un anno, dieci anni, oppure mai. Molte volte, nell'incessante fluire dei post, una mancata risposta immediata da parte dell'interlocutore, per qualsiasi motivo, può portarlo a non dare mai più una risposta: troppo tardi non avrebbe più senso per quell'opera di traduzione collettiva simultanea. Se nelle comunità reali la conversazione faccia a faccia obbliga l'interlocutore a una risposta immediata e se nelle comunità immaginate della stampa l'interlocutore non dà una risposta immediata o comunque per la maggioranza non c'è possibilità di rispondere sotto forma di dialogo, nelle comunità immaginate dei social network ci si aspetta che l'interlocutore dia una risposta in diretta e sotto forma di dialogo, ma potrebbe anche non farlo mai, lasciando «girare a vuoto» il ragionamento o il tentativo di dialogo del mittente. Non ha senso di conseguenza anticipare il disaccordo, immaginandoselo radicato nella realtà delle relazioni, come invece ha senso fare nelle comunità reali. Non ha senso immaginare il disaccordo nemmeno in forma diversa dal dialogo faccia a faccia, come accadeva nelle comunità immaginate dopo l'avvento della stampa. La mancata anticipazione del disaccordo che si dà nel dialogo e che permette di ancorare i ragionamenti alla relazione dialogica con gli altri è il meccanismo che porta le comunità immaginate su FB a dimenticare più facilmente quelle differenze che producono il disaccordo stesso e che costituiscono qualsiasi comunità immaginata. In questo senso, lo stile in cui le comunità dei social network si immaginano favorisce la deriva della comunità immaginata in comunità immaginaria.

3. Bolle epistemiche e razionalità «differita»

Come argomentano Mercier e Sperber (2017), il ragionamento solitario è destinato a incepparsi, producendo interpretazioni fuorvianti, proprio perché il ragionamento si è evoluto come strumento di interazione e di dialogo. Portando avanti una reinterpretazione sistematica dei precedenti studi sul ragionamento (Frixione, 2007; Labinaz, 2013), Mercier e Sperber mostrano che mentre il ragionamento individualistico viene influenzato da *bias* come quello di conferma (*confirmation bias* o *myside bias*), il ragionamento collettivo, in particolare all'interno di gruppi tra pari, migliora l'argomentazione dei singoli individui. La pervasività del *myside bias*, ovvero la tendenza a cercare quelle informazioni che convalidano le proprie convinzioni, credenze o ipotesi precedenti (Nickerson, 1998;

Correia, 2011), viene spiegata come la naturale conseguenza di un meccanismo cognitivo che cerca giustificazioni post-hoc, più che correggere le intuizioni di partenza: «le razionalizzazioni sono il prodotto naturale di un meccanismo di ragionamento argomentativo dotato di un consistente *confirmation bias*» (Mercier, 2011, 133, trad. mia). Da questo punto di vista, il ragionamento è un prodotto dell'evoluzione che non solo ha permesso al singolo di difendere le proprie credenze e ad accettare le credenze altrui qualora fossero fondate, ma anche alle comunità di consolidarsi: le stesse norme sociali che costituiscono e regolano la vita della comunità si instaurano perché in questo modo i singoli riescono a risolvere problemi che da soli non riuscirebbero a risolvere (Mercier e Sperber, 2017). Nella trasformazione della comunità immaginata in comunità immaginaria, il ragionamento diventa non solo solitario, ma anche legato all'illusione che sia realmente di tipo relazionale e collettivo. Le comunità immaginarie, paradossalmente, «ragionano» come un grande individuo singolo e sono dunque, al contrario delle comunità immaginate, esposte ai *bias* del ragionamento individualistico. In un circolo vizioso, questo meccanismo può essere reiterato *ad libitum* nei social network, che offrono gli strumenti per condividere l'informazione creata da un altro membro della comunità come se fosse propria.

Diversamente dai membri delle comunità immaginate dopo l'avvento della stampa, i membri delle comunità immaginate nei social network non sono semplicemente i destinatari delle informazioni, ma sono al contempo produttori dei contenuti delle informazioni che condividono («produsers», Bruns e Highfield, 2012). Nella condivisione delle notizie da un membro all'altro della comunità, le notizie stesse vengono elaborate in quel processo che abbiamo chiamato *traduzione collettiva simultanea*, che sovralimenta un certo «sense of agency» dei membri della comunità (Sundar, 2008), ovvero la sensazione di «avere un qualche controllo sull'informazione che condividono» (Arfini, Bertolotti e Magnani, 2017, 11, trad. mia). Si tratta naturalmente di un controllo illusorio, pari a quello che si ha con il ragionamento solitario, con l'aggiunta dell'illusione di avere un qualche «potere epistemico» sull'informazione data alla comunità cui si appartiene. Il paradosso del ragionamento solitario a dimensione collettiva insieme a questa sensazione di potere epistemico sull'informazione spiega almeno in parte la proliferazione di esperti auto-proclamati nelle comunità immaginarie dei social network. Basta ricordare i movimenti antivaccinisti per dire quanto i social network possano favorire la diffusione di comunità immaginarie che ruotano attorno a sedicenti leaders esperti nel tema di interesse collettivo (Ervas, 2018).

La difficoltà e a volte l'incapacità di distinguere tra conoscenza e ignoranza è la caratteristica centrale delle «bolle epistemiche» (Woods, 2005). Come spiegato da Selene Arfini e Tommaso Bertolotti:

> Una bolla epistemica è un fenomeno di autoinganno epistemico, attraverso il quale l'agente diventa inconsapevole della differenza tra conoscere qualcosa e credere di conoscere quella stessa cosa. Deriva dal fatto che credere di avere una qualche conoscenza è una condizione piacevole per l'individuo: gli permette di agire secondo le sue credenze e di alleviare l'irritazione che potrebbe dare la mancanza di qualche informazione importante (Arfini e Bertolotti, 2018, 82, trad. mia).

Nelle comunità immaginarie che si trovano all'interno di una bolla epistemica, manca anche l'irritazione che potrebbe venire da un reale disaccordo. I membri di tali comunità vivono perciò nell'illusione di possedere la conoscenza rilevante su argomenti centrali della vita pubblica delle comunità reali, senza di fatto possederla o, peggio, pensando che sia valida anche per gestire la realtà. Infatti, se da un lato la tecnologia messa a disposizione dai social network ha consentito ai membri delle comunità di avere accesso a più informazioni con la congiunta possibilità di condividerle e discuterle, dall'altro lato ha contribuito a diminuire l'attenzione e la responsabilità che ci si assume nella vita reale quando si applicano quelle stesse informazioni, acquisite normalmente dagli esperti delle comunità reali in maggior tempo e con maggiori risorse intellettive (Arfini e Bertolotti, 2018).

D'altro canto, la velocità con cui si propagano le informazioni nei social network grazie alla possibilità di condividerle e la velocità con cui vengono consumate e deteriorate in situazioni di comunità immaginarie, rende difficile l'intervento dei «veri esperti». Se le notizie inautentiche non vengono contraddette subito, è molto difficile poi intervenire all'interno della bolla epistemica, dove tutti i membri della comunità immaginaria sostengono un argomento (fallace) e guadagnano consenso e fiducia da parte della comunità nell'atto stesso di condividerlo. Inoltre, la mancanza del lessico specialistico e di rigore argomentativo rende ancora più difficile, per un esperto, entrare nel merito delle questioni poste dalla comunità immaginaria. La razionalità dell'esperto viene perciò «differita», come si fa con l'interlocutore il cui disaccordo si è smesso di anticipare. In tali situazioni, la maggior parte delle volte gli esperti rinunciano a intervenire o rimandano l'intervento. L'esperto dovrebbe infatti tenere insieme sia un livello accettabile di rigore sia una certa efficacia comunicativa per sradicare i meccanismi che portano alla creazione delle «bolle epistemiche». Dovrebbe riuscire a comunicare quanto è corroborato dalla ricerca scientifica, mantenendo il rigore della metodologia e del lessico specialistico del proprio settore disciplinare, ma al contempo rendere fruibile il sapere senza ridurlo o banalizzarlo, facendo in modo che susciti interesse, stimolando l'immaginazione dell'uditorio verso un'ulteriore ricerca personale.

Gli esperti hanno spesso trovato nel linguaggio figurato un modo per mantenere il rigore metodologico e al contempo avvicinarsi al lessico

del grande pubblico: per esempio, si è sostenuto che la metafora svolga un ruolo cognitivo fondamentale nella crescita del sapere scientifico (Hesse, 1974; Kuhn, 1979). Al contempo la metafora ha un forte potere comunicativo perché aiuta a spiegare un fenomeno scientifico complesso, estendendo la conoscenza da un dominio noto al grande pubblico a un dominio ignoto e di difficile comprensione, senza ricorrere a un linguaggio letterale iperspecialistico (Ervas, 2018). La metafora sembra un mezzo di comunicazione efficace anche nelle comunità immaginate: come ampiamente argomentato (Murphy, 2012; Tarzia, 2013), le metafore sono i mezzi per eccellenza per costruire un «immaginario collettivo». Tuttavia anche le metafore possono essere un'arma a doppio taglio: la metafora infatti non risolve in toto il problema di ignoranza o di presunta conoscenza aggiungendo conoscenza (giustificata) a ciò che è ignoto, ma propone piuttosto una comparazione con un ambito concettuale noto da cui ci si aspetta che l'interlocutore tragga le giuste inferenze. Tuttavia, non è detto che l'interlocutore lo faccia e l'incomprensione o l'interpretazione deviata sono molto più probabili lì dove esistono i meccanismi della «bolla epistemica» tipici delle comunità immaginarie. In quanto segue, si prenderà in esame la metafora come meccanismo di creazione dell'immaginario collettivo, per mostrare in che modo i membri delle comunità immaginano e condividono i contenuti delle informazioni. In particolare, ci si concentrerà sulle metafore visive piuttosto che su quelle verbali: per la natura multimodale della comunicazione nei social network, si ritiene infatti che le metafore visive rappresentino lo stile specifico delle comunità immaginate nei social network.

4. Immagini e metafore visive

Lo stile con cui le comunità FB si immaginano, come detto, è intrinsecamente multimodale: immagini (statiche o in movimento) e testi si intersecano a sequenze audio. Tuttavia, sembra che la modalità predominante dei post sia quella visiva (immagini e video) piuttosto che quella verbale (testuale o audio). Secondo le infografiche dei social media (Jakus, 2018), l'84% della comunicazione è visiva e spesso gli utenti preferiscono disattivare altre modalità o canali per usufruire esclusivamente di quella visiva: per dare un esempio, l'85% dei video FB è visto senza audio (Patel, 2016). I testi sono normalmente molto brevi e comunque limitati (per esempio Twitter consentiva un limite di 140 caratteri e solo recentemente, nel 2017, ha concesso 280 caratteri), mentre le immagini possono essere inserite senza «sprecare» caratteri, anche per la loro capacità di rappresentare simultaneamente diversi elementi in un unico contenuto. Altre ricerche dimostrano che uno dei modi per rafforzare la visibilità di un post FB è quello di rendere le sue caratteristiche più «vivide» (De Vries, Gensler e Leeflang, 2012).

Sebbene tali indagini abbiano finalità di mercato, sono utili ai nostri fini perché mostrano come tale «vividezza» sia maggiore lì dove ci sono più elementi dinamici (animazioni), colori e, soprattutto, immagini (Fortin e Dholakia, 2005; Lohtia, Donthu e Hershberger, 2003).

In particolare, alcuni di questi studi hanno dimostrato che i post più «vividi» sono anche quelli più efficaci rispetto alla percentuale di *click* ottenuti (Lohtia, Donthu e Hershberger, 2003), e quelli che più rafforzano la propria posizione e le proprie aspettative verso la pagina web che li contiene (Fortin e Dholakia, 2005). Questo effetto è ancora più forte in presenza di immagini «metaforiche», ovvero quelle immagini che non rappresentano in modo «fotografico» la realtà ma che la lasciano immaginare all'utente oppure, come si vedrà in seguito, all'intero gruppo. Le metafore in genere permettono a chi comunica di esprimere un concetto *target* nei termini di un concetto *source* (Ervas e Gola, 2016). Le metafore possono essere viste – oltre che come abbellimenti o mezzi letterari – anche come meccanismi che ci mettono in grado di pensare qualcosa di generalmente meno noto o sconosciuto (il concetto *target*) nei termini di qualcosa di familiare e conosciuto (il concetto *source*) (Bowdle e Gentner, 2005). Solo alcune proprietà del concetto *source*, pertinenti per l'interpretazione della metafora, vengono selezionate per comprendere che cosa si intende dire del concetto *target* (per una panoramica delle teorie dei processi cognitivi sottesi alla selezione delle proprietà, si veda Gibbs e Colson, 2012): per esempio in una metafora verbale come «Giulietta è il sole» pronunciata da Romeo, solo alcune proprietà del concetto di sole vengono utilizzate per capire chi è Giulietta per Romeo, come per esempio l'essere fonte di luce, calore, bellezza, ecc.

Le metafore verbali sono metafore monomodali che si presentano in formato testuale (o audio), ma si possono presentare anche in formato visivo. Una metafora visiva, detta anche *metafora pittorica*, è una metafora monomodale in cui i concetti *target* e *source* vengono rappresentati e comparati visivamente, esclusivamente tramite immagini. Quando un testo accompagna la metafora visiva, modulando o cambiando il suo significato, parliamo invece di *metafora pittorica multimodale* (Forceville, 1996; 2008; Ojha, 2015). La metafora visiva differisce dunque da un'immagine «letterale» perché rappresenta e compara due concetti distinti e anche dalla metafora pittorica multimodale perché non presenta elementi verbali. Per esempio, l'immagine di un gelato che si scioglie è letterale rispetto a quella metaforica del gelato che al posto della pallina presenta la Terra che si scioglie in fig. 1. L'immagine in fig. 1 è monomodale (esclusivamente visiva) rispetto, per esempio, a un'immagine metaforica multimodale della «Terra/gelato» con il testo «Melting away» (autore: Joe Antcliff; www.deviantart.com/joe-antcliff/art/Save-the-Earth-3-99668436) che ne precisa l'interpretazione e rimanda dunque più esplicitamente al problema del riscaldamento globale.

FIG. 1. Esempio di metafora pittorica o metafora visiva (123RF, Creative agency: Aunt Spray).

Nel caso della metafora verbale, la stessa distinzione tra «letterale» e «metaforico» è alquanto problematica (Carston, 2002; Stern, 2006) e altrettanto si può dire per la metafora visiva: sono il contesto e l'intenzione del mittente che fanno di una determinata metafora una metafora, piuttosto che un enunciato/immagine letterale. Fuor di contesto e d'uso da parte di un parlante, la metafora di per sé ha natura ambigua e, soprattutto nel caso delle metafore visive, senza contesto è difficile stabilire la *direzionalità della metafora*, ovvero individuare i concetti *source* e *target* (Indurkhya e Ojha, 2017). Nel caso delle metafore visive, per individuare quali sono i concetti *source* e *target* serve l'immagine nel suo insieme e a volte degli elementi testuali aggiuntivi oppure il marchio dell'azienda nel caso della pubblicità o dell'istituzione/organizzazione nel caso delle campagne sociali.

Nonostante la sua intrinseca ambiguità, la metafora verbale viene ampiamente utilizzata nella sfera pubblica, social network compresi (Ervas e Gola, 2016), dove invece sembreremmo aver bisogno di messaggi chiari, non equivoci, per garantire una comunicazione trasparente. Seppur potenzialmente più ambigua di un messaggio letterale, la metafora ha infatti un potere comunicativo maggiore: rispetto ad un equivalente letterale ci permette di comunicare in modo più breve (e dunque cognitivamente

più economico) e suggestivo il messaggio del parlante (Sperber e Wilson, 1995; Wilson e Carston 2007; Carston, 2002). Anche le metafore visive, così come quelle pittoriche multimodali, sono usate spesso nelle pubblicità, nell'arte, nelle campagne sociali e politiche, nelle vignette satiriche (Forceville, 1996; Forceville e Urios-Aparisi, 2009; Refaie, 2003; Fahlenbrach, 2015; Meijers, Remmelswaal e Wonneberger, 2018). La metafora in genere viene utilizzata per il suo potere cognitivo di stabilire connessioni inedite e creative tra concetti (Indurkhya, 1992), per il suo potere epistemico di estendere la conoscenza (Hesse, 1974; Kuhn, 1979) e per il suo potere retorico-persuasivo (Sopory e Dillard, 2002). Le metafore visive sembrano tuttavia essere più persuasive di quelle verbali (Bulmer e Buchanan-Oliver, 2006): alcuni studi sperimentali in ambito pubblicitario hanno mostrato che le metafore visive aumentano la credibilità di chi le ha comunicate (Jeong, 2008), l'apprezzamento della pubblicità (Chang e Yen, 2013) e la fiducia del consumatore nel prodotto (Phillips e McQuarrie, 2009).

È stata proposta una classificazione delle metafore visive in base a come vengono rappresentati i concetti *source* e *target* (Maes e Schilperoord, 2008): 1) tramite *giustapposizione*, qualora vengano rappresentati uno a fianco all'altro, come nel caso della pubblicità della chitarra Gibson, accostata ad un tridente diabolico; 2) tramite *fusione* in un unico oggetto come nel caso della campagna sociale dell'Unicef in cui la forchetta assume la forma di una mano; 3) tramite *sostituzione* dell'uno con l'altro, come nel caso della pubblicità della giornata mondiale per l'ambiente, in cui il *target* (la Natura rappresentata dall'albero) sostituisce il *source*: il ventre gravido della madre umana, assente ma inferito dal gesto della donna (fig. 2). Secondo la classificazione classica delle metafore visive proposta da Charles Forceville (1996, 2008), si avranno rispettivamente 1) una similitudine pittorica; 2) una metafora visiva ibrida e 3) una metafora visiva contestuale.

Si può ipotizzare che a questi diversi tipi di metafora visiva corrispondano diversi processi d'immaginazione. Nel caso della *similitudine pittorica* non si devono immaginare oggetti ma proprietà di oggetti rilevanti per cogliere l'interpretazione della similitudine: gli oggetti sono già dati e vengono «comparati visivamente» in base a delle somiglianze percettive, quali ad esempio il colore (gradazione di rosso e nero) e la forma, chiedendo all'interprete di immaginare e comparare le proprietà degli oggetti che rappresentano *source* e *target*. Contra Forceville (1996, 2008), si ritiene tuttavia che – come per le similitudini vs. metafore verbali (Glucksberg e Haught, 2006; Carston e Wearing, 2011) – le similitudini pittoriche non dovrebbero essere considerate come metafore *tout court*, dove gli oggetti che rappresentano *source* e *target* non vengono semplicemente offerti alla comparazione (Ojha *et al*., 2018). Infatti, nel caso della *metafora visiva ibrida* l'interprete ricostruisce tramite immaginazione le parti mancanti degli oggetti *source* e *target*:

1) Similitudine pittorica

2) Metafora visiva ibrida

3) Metafora visiva contestuale

FIG. 2. Esempi di metafore visive secondo il tipo di rappresentazione dei concetti *source* e *target* Per gentile concessione di 1) Gibson Italia e USA, Creative agency: Carmichael Lynch, USA; 2) UNICEF.ch, Creative agency: Saatchi & Saatchi; 3) World Environment Day, Creative agency: Valappila Communication, India.

un «dente» della forchetta e la «sostanza» della mano umana, la cui forma – già data nell'immagine e dunque non immaginata – rimanda sia all'azione dell'imboccare per sfamarsi sia a quella del tendere la mano per aiutare. Nel caso della *metafora visiva contestuale* l'interprete deve immaginare un intero oggetto con le sue proprietà rilevanti: è il contesto insieme all'azione compiuta dalle mani femminili che porta l'interprete a immaginare il concetto *source*, la gravidanza della donna.

5. Immaginazione e persuasione nelle metafore

Nella teoria contemporanea della metafora, c'è stato un acceso dibattito tra l'idea che la metafora abbia natura concettuale, come proposto da filosofi quali Max Black (1955) o dalla stessa teoria concettuale della metafora di George Lakoff e Mark Johnson (1980), e l'idea di Donald Davidson (1978) che la metafora sia un'immagine irriducibile a concetti. Recentemente Robyn Carston (2010; 2018) ha offerto una rilettura della tesi davidsoniana, che in qualche modo riconcilia le alternative di quel dibattito: la metafora dà accesso a un contenuto concettuale – calcolabile a partire dal significato convenzionale e dall'informazione disponibile nel contesto – nel caso delle metafore convenzionali o lessicalizzate come per esempio «Una mamma è una chioccia», ovvero quelle metafore il cui uso è già consolidato in una comunità linguistica e il cui significato è stato già codificato nel dizionario (Ervas e Gola, 2016). Oltre a dare accesso a un contenuto concettuale, le metafore vive o creative – come per esempio «Una mamma è un arcobaleno» – possono invece evocare delle immagini:

> Le immagini non sono comunicate ma attivate o evocate quando si ha accesso a certi concetti lessicali e possono essere ulteriormente sviluppate dall'immaginazione (per esempio, cambiando prospettiva, focalizzando l'attenzione su

un dettaglio, o formando una nuova sequenza dinamica) quando il contenuto concettuale del proferimento è stato colto (Carston, 2010, 319, trad. mia).

Il significato letterale non viene soppresso nell'interpretazione della metafora e produce una serie di effetti comunicativi di tipo affettivo, estetico e immaginativo (Carston, 2018; Carston e Wearing, 2011; Rubio-Fernandez, 2007; Weiland *et al*., 2014). In particolare, quando il termine (*vehicle*) del dominio concettuale *source* è concreto, risulta più comprensibile e memorabile rispetto a un termine astratto (Paivio, Walsh e Bons, 1994; Sadoski, Goetz e Fritz, 1993) e al contempo più «immaginabile», ove per «immaginabilità» si intende la capacità di un termine di evocare immagini mentali (Paivio, 2013). Per esempio, il termine «arcobaleno», *vehicle* nella metafora «Una mamma è un arcobaleno», è più concreto, comprensibile e «immaginabile» rispetto al termine astratto «rasserenamento», *vehicle* nella metafora «Una mamma è un rasserenamento».

Come argomentato altrove (Ervas, 2019), l'immaginabilità è vincolata dall'architettura del nostro sistema percettivo e cognitivo (Pylyshyn, 1973; 2002; Block, 1981). Tuttavia, sembra diverso il caso della metafora visiva, dove l'immagine stessa con i propri dettagli vincola ancora di più l'immaginazione agli oggetti raffigurati nell'immagine stessa. Per esempio, nei casi di metafora visiva ibrida e contestuale nella figura 2, l'immaginazione è vincolata da dettagli specifici dell'immagine, come il colore grigio della forchetta nella metafora visiva ibrida o la consistenza dell'albero nella metafora visiva contestuale. Recentemente, alcuni studi sperimentali hanno avuto per oggetto l'interpretazione delle metafore visive (Indurkhya e Ojha, 2013; Ojha e Indurkhya, 2016), mostrando che le immagini evocate dalle metafore visive sono vincolate da specifiche caratteristiche percettive delle metafore stesse. Non tutte le caratteristiche dell'immagine metaforica sono infatti rilevanti o pertinenti per l'interpretazione della metafora. Nei casi di metafora visiva ibrida e contestuale in fig. 2, nessuna delle caratteristiche sopra elencate è determinante per l'interpretazione della metafora: è la forma della forchetta nella metafora visiva ibrida a suggerire l'immagine della mano che si tende a donare o l'abbraccio «femminile» all'albero nella metafora visiva contestuale a suggerire l'immagine della gravidanza.

Il potere delle metafore di evocare immagini (Carston, 2010; 2018) può essere collegato al loro potere persuasivo (Ervas, 2019). Seguendo la teoria interattiva (Black, 1955; Indurkhya, 1992), una metafora come per esempio «l'uomo è un lupo» non è ricavabile dai singoli termini o dai singoli domini concettuali *source* e *target*, ma è prodotta dalla loro interazione che ristruttura i concetti coinvolti. Al concetto *source* è associato infatti un «sistema di luoghi comuni» che viene applicato al concetto *target* e in questa operazione si selezionano, enfatizzano, sopprimono, organizzano le sue proprietà. Il sistema di luoghi comuni

non è vero, non coincide cioè con le proprietà reali di ciò che istanzia il concetto *source*, né con il concetto dell'esperto, ma non importa sapere se è vero o conoscere il concetto dell'esperto perché la metafora sia efficace e persuasiva:

> Dal punto di vista dell'esperto, il sistema di luoghi comuni può includere mezze verità o errori madornali (come quando una balena viene classificata come un pesce); ma la cosa importante per l'efficacia della metafora non è che i luoghi comuni siano veri, ma che possano essere facilmente e prontamente evocati. [...] Qualsiasi tratto umano che senza troppi sforzi possa essere espresso nel «linguaggio-lupo» sarà reso evidente, qualsiasi tratto che non possa essere espresso rimarrà nello sfondo. La metafora del lupo sopprime certi dettagli e ne enfatizza altri – in poche parole, organizza la nostra visione dell'uomo (Black, 1955, 287-288, trad. mia).

È stato dato il nome di *framing* all'effetto metaforico di mettere in evidenza delle caratteristiche del concetto *source* utili a comprendere il concetto *target*, ma contemporaneamente mettere in ombra altre caratteristiche del concetto *source*, ritenute non salienti o rilevanti per la comprensione del *target* (Entman, 1993; Semino, 2008).

In modo molto simile, come si è visto, le metafore visive propongono delle caratteristiche percettive salienti, che mettono in ombra altre caratteristiche percettive dell'immagine metaforica, irrilevanti per l'interpretazione della metafora. Christian Burgers e colleghi (2015) hanno dimostrato sperimentalmente che l'efficacia persuasiva, fine ultimo delle azioni promozionali, è data dalla convenzionalità dei *frames* richiamati implicitamente nel messaggio, che aiutano a rendere concreti e più comprensibili i messaggi. Tuttavia, altri studi recenti hanno argomentato che le metafore non sono persuasive per questo motivo, ma piuttosto perché creative e devianti rispetto alle aspettative (McQuarrie e Mick, 1996) e perché necessitano di un'ulteriore elaborazione per «risolvere il puzzle» e comprendere il messaggio (Jeong, 2008; Chang e Yen, 2013). Tuttavia altri studi hanno mostrato che anche la creatività è frutto di regole applicabili all'interno di schemi concettuali familiari, più che un processo che avviene al di fuori di tali schemi (Goldenberg, Levav, Solomon e Mazursky, 2009). L'uso di *frames* più o meno creativi sembra dunque una questione di grado, come del resto per le metafore verbali e dunque le strategie persuasive di *framing* sono molto simili. Altri studi hanno sostenuto che le metafore visive, facendo uso dell'argomentazione e della persuasione visiva, sono più indirette e aperte all'interpretazione (McQuarrie e Phillips, 2005; Phillips e McQuarrie, 2004). Se le metafore visive risultano essere più persuasive di quelle verbali è piuttosto perché attirano l'attenzione (visiva) degli interlocutori e sono dunque più memorabili. In quanto segue, tramite l'analisi di due casi di studio si osserverà che, come sostenuto sopra, le immagini metaforiche vincolano

maggiormente l'immaginazione rispetto alle metafore verbali rendendole meno aperte all'interpretazione e più soggette a interpretazioni «devianti» solo apparentemente legate all'immagine data. In questi casi, le comunità immaginate dei social network possono diventare comunità immaginarie, dove il possibile razionale disaccordo nella discussione – anziché essere anticipato e previsto – diventa «differito» o posticipato virtualmente all'infinito. In questi casi, a meno che un «vero esperto» non intervenga, le interpretazioni ancorate a irrilevanti dettagli visivi alimentano la nascita delle «bolle epistemiche».

6. Come preservare l'ignoranza ancorandola a dettagli

Le metafore verbali sono state interpretate anche come un *argomento implicito* in cui si traggono alcune inferenze dalla comparazione tra concetti *source* e *target* (Oswald e Rihs, 2014; Macagno e Zavatta, 2014; Wagemans, 2016). Come ricordato, la selezione delle proprietà è una strategia di *framing* che forza implicitamente l'interprete a considerare il *target* sotto una prospettiva specifica e che può fortemente influenzare il ragionamento e la valutazione dell'argomento sottostante (Thibodeau e Borodisky, 2013; Semino, Demjen e Demmen, 2016). Qui si sostiene che anche le metafore visive possano essere pensate come argomenti impliciti che sottendono delle strategie di *framing*, sebbene abbiano specifiche caratteristiche dovute alla rappresentazione visiva «superficiale» che le distinguono da quelle verbali. Non si intende qui dire che tutte le metafore visive siano argomenti impliciti. Come per altri tipi di immagine, alcune metafore sono segnali o «bandiere» (*flags*), altre sono effettivamente argomentative (Tseronis e Forceville, 2017). Una metafora «bandiera» mira solo a catturare l'attenzione e spesso la dirige verso un messaggio verbale, incluso nell'immagine quando si tratta di una metafora pittorica multimodale. Non ci dà una ragione (nel senso di una premessa) per concludere che dovremmo fare o pensare ciò che ci vuole far fare o pensare, quindi non è parte di un argomento (Birdsell e Groarke, 2007). Tuttavia la metafora visiva può essere – ed anzi è spesso – sia una «bandiera» per attirare l'attenzione sia un argomento (o parte di un argomento nel caso delle metafore multimodali). Sicuramente, come ricordato precedentemente, nei social network le metafore visive funzionano come «bandiere» e vengono utilizzate per attirare l'attenzione del gruppo su una notizia e/o per aumentare la visibilità di un post favorendone l'*interattività*. Le metafore visive sono perciò un mezzo molto diffuso per comunicare via social network e sono diventate probabilmente l'esempio per eccellenza dello stile in cui si immaginano le comunità come quelle di FB.

Un esempio che demarca la differenza tra le comunità immaginate dopo l'avvento della stampa e quelle immaginate dopo l'avvento dei

Fig. 3. Post tratto dalla pagina FB del settimanale *The Economist* (www.facebook.com/TheEconomist/photos/a.10150279872209060/10156604802199060/?type=3&theater, ultimo accesso: 20/02/2019).

social network è dato dalla fig. 3, che riporta un post della pagina FB del settimanale *The Economist* dedicato alla più lunga crisi istituzionale italiana della storia della Repubblica, conseguente alle elezioni di marzo 2018. La metafora visiva propone un'Italia, rappresentata da un gelato tricolore (bianco, rosso, verde), alle prese con una crisi «a due velocità»: una crisi immediata, rappresentata dalle micce esplosive inserite nel gelato, e una crisi a lungo-termine, rappresentata dallo sciogliersi stesso del gelato. Il titolo (*Handle with care*) rende esplicita la necessità di cura e attenzione per la situazione di crisi. Il post precisa poi verbalmente che la crisi immediata sembra risolvibile e che la crisi a lungo termine è più preoccupante: si lascia all'interprete immaginare il lento degenerare della situazione politica italiana a partire dallo sciogliersi del gelato tricolore.

Si tratta di una metafora visiva ibrida simile a quella in fig. 1, ma che, diversamente da quella in fig. 1 fa leva su elementi metonimici (i colori della bandiera italiana) e *frames* stereotipici ben noti a livello mondiale per rappresentare l'Italia (le prelibatezze culinarie). L'uso di *frames* consolidati corrisponde alla necessità di rendere più comprensibile il messaggio e insieme di rendere concreto e legato all'esperienza quotidiana un concetto *target*, come la crisi, astratto e dunque difficile da rendere in immagini. Non manca però chi polemizza con l'uso dello stereotipo o chi lo avverte come «meno grave» di altri stereotipi diffusi sull'Italia (fig. 3.1), e non solo nelle testate giornalistiche:

FIG. 3.1. Commenti al Post in fig. 3 (*ibidem*).

Se fosse stata la copertina di un settimanale a stampa, non ci sarebbe stata la possibilità di commentare in diretta e/o di cercare «traduzioni simultanee» della metafora alternative a quella fornita dall'autore del post. Invece la comunità FB reinterpreta la metafora visiva, annullando dettagli visivi rilevanti o generalizzando (fig. 3.2).

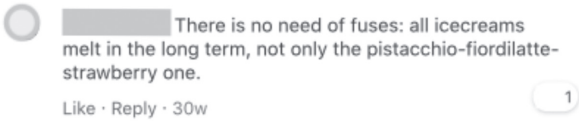

FIG. 3.2. Commenti al Post in fig. 3 (*ibidem*).

In questo commento, invece di discutere il problema sollevato dalla comunità immaginata di FB e rappresentato dalla metafora visiva come problema della comunità italiana reale, non si coglie la rilevanza delle micce per comprendere che esiste una crisi a breve termine, quella generata dall'esito delle elezioni, e si applica la proprietà dello «sciogliersi» rilevante per la situazione italiana «a lunga durata» indiscriminatamente a qualsiasi nazione. In questo modo si distorce la situazione dell'Italia all'interno della realtà europea (rappresentata dal simbolo dell'euro nel cono), e si annulla anche il problema reale sulla governabilità del Paese – ingenerato allora dai risultati delle elezioni e riconosciuto al di là delle distinzioni tra partiti – generando online l'idea di una comunità immaginaria.

Fig. 4. Post FB tratto dal gruppo *Italia Unita per la Scienza* (https://www.facebook.com/IUXLS/posts/941433429272813, ultimo accesso: 19/02/2019).

Una metafora visiva multimodale più creativa, meno legata a *frames* e stereotipi già consolidati, è quella proposta dal gruppo FB *Italia Unita per la Scienza* in fig. 4. Di fronte al calo delle vaccinazioni e alla necessità di recuperare le soglie di vaccinazione necessarie a preservare l'«immunità di gregge», si cerca di spiegare la vaccinazione attraverso

la metafora dell'ombrello (Ervas, 2018). L'ombrello ripara dalla pioggia, alias le malattie contagiose, non solo chi ce l'ha, ma anche chi non ce l'ha, chi non può averlo e addirittura chi non vuole averlo, alludendo ai sostenitori del movimento no-vax.

Il post condivide un post precedente e precisa l'interpretazione della metafora visiva attraverso la similitudine «La vaccinazione è come un ombrello». In questo caso c'è chi critica la metafora fornendo un'interpretazione che si ancora a dettagli non pertinenti, come la direzione dell'ombrello o addirittura immaginari, ovvero fabbricati o inventati, ma non suggeriti o evocati dall'immagine, i.e. ombrelli bucati o arrugginiti (fig. 4.1).

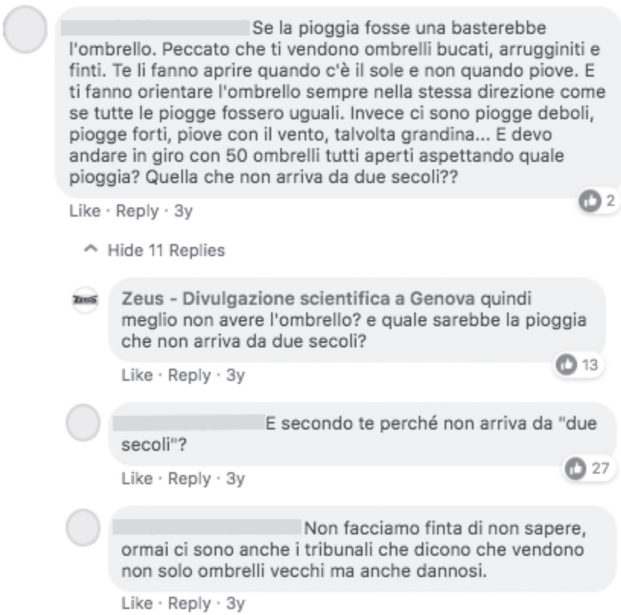

Fig. 4.1. Commenti al Post in fig. 4 (*ibidem*).

La metafora viene dunque (re)interpretata evocando delle immagini totalmente svincolate dall'immagine data o ancorate a dettagli irrilevanti o inventati, preservando sostanzialmente l'ignoranza sull'argomento delle vaccinazioni e cercando una conferma ad hoc a tesi già ritenute vere (*confirmation bias*), indipendentemente dal confronto con la comunità, e con un esperto in questo caso. Il divulgatore scientifico risponde inizialmente mantenendo il *frame* creato dalla metafora e ricevendo meno *likes* di un altro membro della comunità non-esperto. Dopo la reinterpretazione dell'immagine dell'ombrello che ancora si appella a dettagli slegati dall'immagine proposta (ombrelli vecchi e dannosi), esce dal *frame* creativo della metafore e lascia cadere la conversazione.

7. Conclusioni

I social network come FB hanno reso «liquidi» e facilmente superabili i confini tra comunità reale, comunità immaginata e comunità immaginaria. Si è cercato di mostrare che ciò è dovuto principalmente allo stile in cui le comunità si immaginano nei social network. Da un lato la tecnologia offerta dai social network rende possibile una sorta di traduzione collettiva simultanea a cui ciascun membro della comunità può contribuire commentando, condividendo e reinterpretando in diretta quanto viene proposto dalla comunità. Dall'altro lato però, la mancanza della relazione reale, faccia a faccia, può rendere la comunicazione «solitaria» e il ragionamento collettivo sottoposto agli stessi *bias* di conferma del ragionamento individualistico. Questo meccanismo sta alla base della creazione delle «bolle epistemiche», in cui la razionalità di chi ha un'opinione diversa o, in taluni casi, degli esperti, è «differita» anziché anticipata così da lasciar propagare notizie inautentiche e rafforzare le comunità immaginarie.

Infine, lo stile in cui le comunità immaginate si immaginano nei social network è strettamente legato alla natura multimodale e, in particolare, visiva della comunicazione. Se le metafore sono state viste come fonte di creazione degli immaginari collettivi del passato, le metafore visive sono oggi un mezzo potente di argomentazione ma anche di persuasione. Si è cercato di mostrare che, pur basandosi sugli stessi meccanismi di *framing* delle metafore verbali, le metafore visive – al contrario delle verbali – limitano l'immaginazione. Se nelle metafore verbali il limite all'immaginazione è quello imposto dall'architettura del sistema percettivo e cognitivo umano, nelle metafore visive è imposto dagli stessi elementi percettivi delle immagini metaforiche. I casi di studio presi in esame mostrano come a volte le reinterpretazioni delle metafore visive preservino l'ignoranza ancorandola a dettagli, soprattutto lì dove l'utente rimane nella «bolla epistemica» di una comunità immaginaria.

Riferimenti bibliografici

Anderson, B. (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso; trad. it. *Comunità immaginate. Origini e fortuna dei nazionalismi*. Roma: Manifestolibri, 1996.

Arfini, S., Bertolotti, T. (2018). The expert you are (not). Citizens, experts and the limits of science communication. In P. Barrotta e G. Scarafile (a cura di), *Science and Democracy. Controversies and Conflicts*. Amsterdam: Benjamins, pp. 71-86.

Arfini, S., Bertolotti, T., Magnani, L. (2017). Online communities as virtual cognitive niches. *Synthese*, *196*, pp. 377-397.

Birdsell, D.S., Groarke, L. (2007). Outlines of a theory of visual argument. *Argumentation and Advocacy*, *43*, pp. 103-113.

Black, M. (1955). Metaphor. *Proceedings of the Aristotelian Society*, *55*, pp. 273-294.

Block, N. (a cura di) (1981) *Imagery*. Cambridge: MIT Press.

Bowdle, B., Gentner, D. (2005). The career of metaphor. *Psychological Review*, *112*, pp. 193-216.

Bruns, A., Highfield, T. (2012). Blogs, Twitter, and breaking news: The produsage of citizen journalism. In R.A. Lind (a cura di), *Produsing theory in a digital world: The intersection of audiences and production in contemporary theory*. New York: Peter Lang, pp. 15-32.

Bulmer, S., Buchanan-Oliver, M. (2006). Visual rhetoric and global advertising imagery. *Journal of Marketing Communications*, *12*, pp. 9-61.

Burgers, C., Konijn, E.A., Steen, G., Iepsma, M. (2015). Making Ads Less Complex, Yet More Creative and Persuasive. *International Journal of Advertising*, *34*, pp. 515-532.

Caers, R., De Feyter, T., De Couck, M., Stough, T., Vigna, C., Du Bois, C. (2013). Facebook: A literature review. *New Media & Society*, *15*, pp. 982-1002.

Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Carston, R. (2010). Metaphor: Ad hoc concepts, literal meaning and mental images. *Proceedings of the Aristotelian Society*, *110*, pp. 295-321.

Carston, R. (2018). Figurative Language, Mental Imagery, and Pragmatics. *Metaphor & Symbol*, *33*, pp. 198-217.

Carston, R., Wearing, C. (2011). Metaphor, hyperbole and simile: A pragmatic approach. *Language and Cognition*, *3*, pp. 283-312.

Chang, C., Yen, C. (2013). Missing Ingredients in Metaphor Advertising: The Right Formula of Metaphor Type, Product Type, and Need for Cognition. *Journal of Advertising*, *42*, pp. 80-94.

Correia, V. (2011). Biases and fallacies: The role of motivated irrationality in fallacious reasoning. *Cogency*, *3*, pp. 107-126.

Davidson, D. (1978). What Metaphors Mean. *Critical Inquiry*, *5*, pp. 31-47.

De Vries, L., Gensler, S., Leeflang, P.S.H. (2012). Popularity of Brand Posts on Brand Fan Pages. *Journal of Interactive Marketing*, *26*, pp. 83-91.

Dunbar, R.I.M. (2016). Do online social media cut through the constraints that limit the size of offline social networks? *Royal Society Open Science*, *3*, 150292.

Entman, R.M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, *43*, pp. 51-58.

Ervas, F. (2008). *Uguale ma diverso. Il mito dell'equivalenza nella traduzione*. Macerata: Quodlibet.

Ervas, F. (2018). From the "Garrison" to the "Beehive". Metaphors and Framing Strategies in Vaccine Communication. *Notizie di Politeia*, *34*, pp. 28-37.

Ervas, F. (2019). Natura multimodale e creatività del linguaggio poetico. *Rivista di Estetica*, *70*, pp. 75-91.

Ervas, F., Gola, E. (2016). *Che cos'è una metafora*. Roma: Carocci.

Fahlenbrach, K. (a cura di) (2015). *Embodied metaphors in film, television, and video games: Cognitive approaches*. New York: Routledge.

Forceville, C. (1996). *Pictorial metaphor in advertising*. London: Routledge.

Forceville, C. (2008). Metaphor in pictures and multimodal representations. In

R.W. Gibbs (a cura di), *The Cambridge Handbook of Metaphor and Thought*. Cambridge: Cambridge University Press, pp. 462-482.
Forceville, C., Urios-Aparisi, E. (2009). *Multimodal Metaphor*. Berlin: De Gruyter.
Fortin, D.R., Dholakia, R. (2005). Interactivity and Vividness Effects on Social Presence and Involvement with a Web-Based Advertisement. *Journal of Business Research*, *58*, pp. 387-396.
Frixione, M. (2007). *Come ragioniamo*. Roma-Bari: Laterza.
Gellner, E. (1964). *Thought and Change*. Chicago: University of Chicago Press.
Gibbs, R.W. Jr., Colston, H.L. (2012). *Interpreting Figurative Meaning*. New York: Cambridge University Press.
Glucksberg, S., Haught, C. (2006). On the Relation Between Metaphor and Similes: When Comparison fails. *Mind & Language*, *21*, pp. 360-378.
Goldenberg, J., Levav, A., Solomon, S., Mazursky, D. (2009). *Cracking the Ad Code*. New York: Cambridge University Press.
Hesse, M.B. (1974). *The Structure of Scientific Inference*. London: Macmillan.
Indurkhya, B. (1992). *Metaphor and Cognition*. Dordrecht: Kluwer.
Indurkhya, B., Ojha, A. (2013). An empirical study on the role of perceptual similarity in visual metaphors and creativity. *Metaphor & Symbol*, *28*, pp. 233-253.
Indurkhya, B., Ojha, A. (2017). Interpreting Visual Metaphors: Asymmetry and Reversibility. *Poetics Today*, *38*, pp. 93-121.
Jakus, D. (2018). Visual Communication in Public Relations Campaigns. *MINIB*, *27*, pp. 25-36.
Jeong, S. (2008). Visual Metaphor in Advertising: Is the Persuasive Effect Attributable to Visual Argumentation or Metaphorical Rhetoric? *Journal of Marketing Communications*, *14*, pp. 59-73.
Kosslyn, S.M., Thompson, W.L., Ganis, G. (2006). *The Case for Mental Imagery*. New York: Oxford University Press.
Kuhn, T.S. (1979). Metaphor in Science. In A. Ortony (a cura di), *Metaphor and Thought*. Cambridge: Cambridge University Press, pp. 409-419.
Labinaz, P. (2013). *La razionalità*. Roma: Carocci.
Lakoff, G., Johnson, M. (1980). *Metaphors we live by*. Chicago: Chicago University Press.
Lohtia, R., Donthu, N., Hershberger, E.K. (2003). The Impact of Content and Design Elements on Banner Advertising Click-through Rates. *Journal of Advertising Research*, *43*, pp. 410-418.
Macagno, F., Zavatta, B. (2014). Reconstructing metaphorical meaning. *Argumentation*, *28*, pp. 453-488.
Maes, A., Schilperoord, J. (2008). Classifying visual rhetoric: Conceptual and structural heuristics. In E. McQuarrie e B. Phillips (a cura di), *Go figure new directions in Advertising Rhetoric*. Armonk: M.E. Sharpe, pp. 227-253.
Marijn, H.C., Meijers, P.R., Wonneberger, A. (2018). Using Visual Impact Metaphors to Stimulate Environmentally Friendly Behavior. *Environmental Communication*, pp. 1-16.
McQuarrie, E.F., Mick, D.G. (1996). Figures of rhetoric in advertising language. *Journal of Consumer Research*, *22*, pp. 424-438.
McQuarrie, E.F., Phillips, B.J. (2005). Indirect persuasion in advertising: How consumers process metaphors presented in pictures and words. *Journal of Advertising*, *34*, pp. 7-20.

Mercier, H. (2011). What Good is Moral Reasoning? *Mind & Society*, *10*, pp. 131-148.
Mercier, H., Sperber, D. (2017). *The Enigma of Reason. A New Theory of Human Understanding*. Cambridge, MA: Harvard University Press.
Murphy, P. (2012). *The Collective Imagination. The Creative Spirit of Free Societies*. Farnham: Ashgate.
Nickerson, R.S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, *2*, pp. 175-220.
Oeldorf-Hirscha, A., Sundar, S.S. (2015). Posting, commenting, and tagging: Effects of sharing news stories on Facebook. *Computers in Human Behavior*, *44*, pp. 240-249.
Ojha, A. (2015). *Visual Metaphor and Cognition*. Saarbrücken: Lambert Academic Publishing.
Ojha, A., Indurkhya, B. (2016). On the role of perceptual features in metaphor comprehension. In E. Gola e F. Ervas (a cura di), *Metaphor and Communication*. Amsterdam: Benjamins, pp. 147-170.
Ojha, A., Gola, E., Indurkhya, B. (2018). Are hybrid metaphor perceived more strongly than pictorial similes? *Metaphor & Symbol*, *33*, pp. 253-266.
Oswald, S., Rihs, A. (2014). Metaphor as argument: Rhetorical and epistemic advantages of extended metaphors. *Argumentation*, *28*, pp. 133-159.
Paivio, A. (2013). *Imagery and verbal processes*. New York: Psychology Press.
Paivio, A., Walsh, M., Bons, T. (1994). Concreteness Effects on Memory: When and Why? *Journal of Experimental Psychology*, *20*, pp. 1196-1204.
Patel, S. (2016). 85 percent of Facebook video is watched without sound. *Digiday UK*, digiday.com/media/silent-world-facebook-video.
Phillips, B.J., McQuarrie, E.F. (2004). Beyond visual metaphor: A new typology of visual rhetoric in advertising. *Marketing Theory*, *4*, pp. 113-136.
Phillips, B., McQuarrie, E.F. (2009). Impact of Advertising Metaphor on Consumer Belief. *Journal of Advertising*, *38*, pp. 35-48.
Pylyshyn, Z.W. (1973). What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery. *Psychological Bulletin*, *53*, pp. 1-24.
Pylyshyn, Z.W. (2002). Mental Imagery: In Search of a Theory. *Behavioral and Brain Sciences*, *25*, pp. 157-237.
Refaie, E. (2003). Understanding visual metaphor: The example of newspaper cartoons. *Visual communication*, *2*, pp. 75-95.
Rubio Fernández, P. (2007). Suppression in metaphor interpretation: Differences between meaning selection and meaning construction. *Journal of Semantics*, *24*, pp. 345-371.
Sadoski, M., Goetz, E.T., Fritz, J.B. (1993). A Causal Model of Sentence Recall: Effects of Familiarity, Concreteness, Comprehensibility, and Interestingness. *Journal of Reading Behavior*, *25*, pp. 5-16.
Semino, E. (2008). *Metaphor in discourse*. Cambridge: Cambridge University Press.
Semino, E., Demjen, Z., Demmen, J. (2016). An Integrated Approach to Metaphor and Framing in Cognition, Discourse and Practice. *Applied Linguistics*, *5*, pp. 1-22.
Sopory, P., Dillard, J.P. (2002). The persuasive effects of metaphor a meta-analysis. *Human Communication Research*, *28*, pp. 382-419.

Sperber, D., Wilson, D. (1995). *Relevance: Communication and Cognition*, 2nd edn. Oxford: Blackwell.
Stern, J. (2006). Metaphor, Literal, Literalism. *Mind & Language*, *21*, pp. 243-279.
Sundar, S.S. (2008). Self as source: Agency and customization in interactive media. In E.A. Konijn, S. Utz, M. Tanis e S.B. Barnes (a cura di), *Mediated interpersonal communication*. New York: Routledge, pp. 58-74.
Tarzia, F. (2013). Metaphor in the Collective Imaginary. In E. Gola e F. Ervas (a cura di), *Metaphor in focus*. Newcastle: Cambridge Scholar Publishing, pp. 9-20.
Thibodeau, P., Boroditsky, L. (2013). Natural Language Metaphors Covertly Influence Reasoning. *PLoS One*, *8*, e52961.
Tseronis, A., Forceville, C. (2017). *Multimodal Argumentation and Rhetoric in Media Genres*. Amsterdam: Benjamins.
Wagemans, J. (2016). Analysing metaphor in argumentative discourse. *Rivista italiana di filosofia del linguaggio*, *2*, pp. 79-94.
Weiland, H., Bambini, V., Schumacher, P. (2014). The role of literal meaning in figurative language comprehension: Evidence from masked priming ERP. *Frontiers in Human Neuroscience*, *9*, pp. 11-12.
Wilson, D., Carston, R. (2007). A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In N. Burton-Roberts (a cura di), *Advances in Pragmatics*. Basingstoke: Palgrave, pp. 230-260.
Woods, J. (2005). Epistemic Bubbles. In S. Artemov, H. Barringer, A.S. Garcez, L.C. Lamband e J. Woods (a cura di), *We Will Show Them: Essays in Honour of Dov Gabbay*. London: College Publications, pp. 731-774.

Visual Metaphor, Imagined Communities and Deferred Rationality

The paper defines the social network communities as imagined communities and the style in which they are imagined as a «simultaneous collective translation», having a multimodal nature and made possible by a technology which allows members to have a live dialogue. The paper argues that this specific style of social network communities, such as Facebook, has made «liquid» the borders between real, imagined and imaginary communities. Reasoning in anticipation of a disagreement (that might not come) can lead communication via social networks to be «solitary» and collective reasoning to be committed to the same confirmation bias of individualistic reasoning. This mechanism is at the core of «epistemic bubbles» in social networks, which lead others' rationality to be «deferred». The multimodal and especially visual nature of communication via social networks does not help: visual metaphor, its most powerful epistemic and persuasive device, limits (instead of promoting) imagination, anchoring it to perceptual features. By analysing two case studies of visual metaphors translated and reinterpreted in a Facebook group, the paper aims to show that the phenomenon of «deferred rationality» leads to preserve the «epistemic bubbles», producing misinterpretations of visual metaphors.

Keywords: imagination, metaphor, argumentation, social network.

L'autrice ringrazia per il sostegno economico la Fondazione di Sardegna (Progetto di ricerca *Science and its Logics: The Representation's Dilemma*, Cagliari, No. F72F16003220002) e l'Università di Cagliari (Bando MGR "Mobilità Giovani Ricercatori" 2018/2019, Progetto di ricerca: *Conceptual and Perceptual Similarities in Visual Metaphors*, Università di Cagliari/Insitut Jean Nicod, ENS, Paris).

Francesca Ervas, Università di Cagliari, Dipartimento di Pedagogia, Psicologia, Filosofia, via Is Mirrionis 1, 09123 Cagliari, ervas@unica.it

TOMMASO PIAZZA MICHEL CROCE

EPISTEMOLOGIA DELLE *FAKE NEWS*

1. Introduzione

Non vi è dubbio che le nostre vite siano sempre più influenzate dalla nostra interazione con social media quali Facebook, Twitter e Instagram. Non solo le nostre identità sociali – il modo in cui guardiamo noi stessi, e come veniamo percepiti da chi ci sta intorno – ma anche la nostra vita intellettuale – in particolare, il modo in cui acquisiamo informazioni, formiamo credenze e ricerchiamo ragioni in loro supporto – sta diventando sempre più dipendente dalle nostre frequentazioni di ambienti virtuali e social network. Questo fatto pone indubbiamente dei problemi, se pensiamo al recente proliferare di *fake news* sui social media, ovvero alla facilità con cui informazioni inaccurate, false o ingannevoli vengono divulgate, distribuite e prese per vere. Si pensi, ad esempio, al fatto che, come hanno mostrato Vosoughy e colleghi (2018, 1149), i contenuti falsi diffusi su Twitter hanno il 70% di probabilità in più di essere ritwittati dei contenuti veri. La vita delle nostre comunità democratiche viene messa a rischio da questo fenomeno, basti pensare ai presunti report relativi alla cattiva condotta di personaggi politici, alle rappresentazioni fuorvianti circa l'impatto economico di alcune soluzioni politiche, o alle predizioni infondate sulle conseguenze di certi trattamenti medici di larga scala. Inevitabilmente, la nostra propensione ad accettare come vere credenze inaccurate cresce di pari passo con la tendenza ad accumulare informazioni online. Di conseguenza, crescono anche le probabilità che il nostro agire nella sfera pubblica – attraverso dibattiti, campagne, elezioni, etc. – si regga su fondamenta pericolanti.

La problematica in questione riguarda innanzitutto la sfera individuale, laddove affidarsi sistematicamente a stralci di informazione

Questo articolo è il frutto di un lavoro comune dei due autori, che ne hanno condiviso la genesi, la costruzione e la realizzazione finale. Dovendo suddividere la redazione delle varie sezioni nella versione finale, Tommaso Piazza ha redatto i paragrafi 1, 2, e i primi sette capoversi del paragrafo 4; Michel Croce ha redatto i paragrafi 3, 5, e i restanti undici capoversi del paragrafo 4.

condivisi sui social media ci mette in condizione di subire una *deprivazione epistemica*, ossia una contemporanea riduzione di beni epistemici di valore – vale a dire credenze vere o giustificate – e un aumento di beni epistemicamente degenerati – vale a dire credenze false o ingiustificate. Tuttavia, il problema è allo stesso tempo sociale, in quanto le nostre comunità rischiano di condurre le proprie attività pubbliche in un contesto di disinformazione sistematica. Per queste ragioni, il fenomeno delle *fake news* ha recentemente attratto l'interesse di molti esperti in varie discipline ed è ormai quasi unanimemente considerato un male per cui dobbiamo urgentemente trovare una cura – eccezion fatta per Habgood-Coote (2018), secondo cui la malattia dalla quale dobbiamo essere curati è l'uso stesso della nozione di *fake news*.

Questo articolo è principalmente dedicato ad una disamina critica della più recente letteratura filosofica sul tema delle *fake news*, che sta portando l'attenzione su tre questioni fondamentali: primo, che cosa sono le *fake news* e come sarebbe opportuno definirle; secondo, quali sono i meccanismi che favoriscono la proliferazione di *fake news* in quegli ambienti online che sono di maggior interesse per questo numero monografico (ma anche in contesti offline); terzo, chi debba essere ritenuto responsabile e degno di biasimo nel processo sotteso alla generazione, pubblicazione, e diffusione delle *fake news*. L'articolo prende in considerazione queste tre domande ed è organizzato come segue.

Il paragrafo 2 esamina la natura delle *fake news* dal punto di vista delle loro caratteristiche definitorie. La letteratura sul tema si concentra, come vedremo, sulle proprietà che un'asserzione dovrebbe possedere affinché il suo contenuto (letterale o implicato) possa essere considerato una *fake news*. Dopo aver individuato quattro tipi di proprietà rilevanti e aver specificato come le principali teorie sul mercato combinano queste proprietà, offriremo una analisi originale della nozione di *fake news*.

Il paragrafo 3 analizza le modalità tipiche di propagazione delle *fake news*, con particolare attenzione a tre ruoli chiave: il creatore (o assertore) di *fake news*; il propagatore, che vi si imbatte e le condivide sui social media; infine, il consumatore finale. La nostra analisi si concentrerà poi su due categorie di cause della propagazione di disinformazione – li chiameremo fattori individuali e fattori sociali – che spiegano il comportamento degli individui nei ruoli appena menzionati. I fattori individuali fanno riferimento ai pregiudizi o *biases* cognitivi che aumentano la nostra propensione a credere nelle *fake news* e ai problemi legati al livello di attenzione che poniamo ai contenuti sui social media. I fattori sociali, invece, comprendono tre strutture tipiche con cui si organizza la condivisione di contenuti e di valori tra gruppi sui social media: le bolle epistemiche (*epistemic bubbles*), le camere d'eco (*echo chambers*) e la polarizzazione di gruppo (*group polarization*).

Il paragrafo 4 approfondisce il tema della responsabilità e dei (de) meriti *epistemici* dei propagatori e dei consumatori di *fake news*. Dopo

aver illustrato tre concezioni chiave nel dibattito – che spaziano da chi ritiene i consumatori epistemicamente virtuosi e condanna gli ambienti in cui le *fake news* proliferano a chi ritiene il singolo consumatore responsabile ed intellettualmente vizioso – presenteremo la nostra concezione, che intende fare tesoro dei meriti principali delle altre concezioni evitandone i limiti principali, per poi offrire alcune sintetiche osservazioni conclusive nel paragrafo 5.

2. Cosa sono le *fake news*

Gli studiosi che si sono posti il problema di definire che cosa siano le *fake news* concordano sul genere a cui appartengono come specie. Per Rini (2017, E-44) si tratta di un qualche tipo di *informazione*; per Mukerji (2018, 929) le *fake news* sono *qualcosa di asserito*; per Jaster e Lanius (2018, 208) sono un tipo specifico di *notizie*; infine, Gelfert (2018, 103-104) dice esplicitamente che le *fake news* sono un tipo di *informazione*, una forma di *notizia*, qualcosa che viene *presentato come vero*. Tutte queste caratterizzazioni puntano nella stessa direzione: le *fake news* sono un tipo di contenuto proposizionale. Il tipo specifico di contenuto col quale tutti questi autori identificano le *fake news* può essere veicolato in molti modi diversi (da fotografie, video, articoli, etc.). Per evitare ogni volta di differenziare il modo specifico in cui una *fake news* è comunicata, nel resto di questo saggio le *fake news* verranno indicate come contenuto di "asserzioni". Con ciò, intenderemo la nozione di asserzione in senso molto lato, in modo cioè da considerare asserzione ogni atto attraverso il quale è possibile comunicare un contenuto[1].

Entreremo nel vivo della nostra discussione a breve, affrontando la questione di quali siano le asserzioni il cui contenuto può essere identificato con una *fake news*. Per il momento, vale comunque la pena soffermarsi su una questione connessa, relativa al *tipo* di contenuto assertivo col quale si tende a identificare le *fake news*. Tutti i partecipanti al recente dibattito concordano nel sostenere che le *fake news* siano spesso il contenuto *letterale* di asserzioni. In uno degli esempi più ricorrenti tratti dalla recente campagna presidenziale negli Stati Uniti, noto come *pizzagate*, la fake news era che Hillary Clinton fosse coinvolta in una rete

[1] In questo senso lato di asserzione, la fotografia di una tendopoli coperta di neve può essere pubblicata con l'intento di asserire che le vittime di un terremoto sono state abbandonate dai politici (anche se la foto di fatto non ritrae la tendopoli delle vittime), e il video che ritrae un uomo che barcolla e inciampa continuamente può essere pubblicato con l'intenzione di asserire che fosse ubriaco e quindi un etilista (anche se l'uomo era di fatto afflitto da un terribile mal di schiena).

[2] Cfr. https://www.snopes.com/fact-check/pizzagate-conspiracy/ (ultima consultazione effettuata in data 19.07.2019).

di pedofili[2]. Che le cose stessero così era sostenuto in modo esplicito da molti post e ipotetici articoli giornalistici pubblicati e diffusi sui social media. Sembra tuttavia plausibile che una *fake news* non debba necessariamente essere il contenuto letterale di una asserzione, e che possa anche essere il contenuto di una implicatura generata da una asserzione il cui contenuto letterale non è una *fake news*. Jaster e Lanius offrono il seguente esempio immaginario. Un giornale on-line riporta: "Dopo l'arrivo dei rifugiati, ci sono state 47 rapine in città". Immaginiamo che le cose stiano proprio così. Anche in questo caso, quella pubblicata dal giornale on-line può essere considerata una *fake news*. Per quanto non lo dica esplicitamente, infatti, implica che i rifugiati siano responsabili, almeno in parte, delle 47 rapine. Nel caso in cui non fosse così, e 47 fosse il numero di rapine verificatesi indipendentemente dall'arrivo dei rifugiati, sarebbe naturale affermare che l'articolo ha contribuito alla diffusione di una *fake news*[3].

Veniamo ora alla questione principale, relativa a quali proprietà debba avere una asserzione per essere l'asserzione di una *fake news*. Come abbiamo già anticipato, il dibattito si è concentrato su quattro tipi di proprietà: proprietà epistemiche, proprietà intenzionali, proprietà sociologiche e proprietà formali. Vediamo di che cosa si tratta.

L'espressione "epistemico" in "proprietà epistemiche" è da intendersi in senso lato, in modo cioè da includere il *vero* e il *falso*. Il *falso*, in particolare, è solitamente indicato come il principale colpevole, quando si tratta di analizzare le proprietà epistemiche delle asserzioni di *fake news*. Come ricorda Gelfert (2018, 99) Facebook ha proposto di ribattezzare le *fake news* come "false news", ritenendo in questo modo di veicolare in modo più esplicito il fatto che per loro le fake news non sono altro che notizie false camuffate da notizie vere.

Di fatto, l'idea che le *fake news* siano spesso notizie false è ricorrente nella letteratura sull'argomento. Secondo Rini (2017, E-45), quando il creatore di una *fake news* la asserisce, sa che il suo contenuto è "falso in modo significativo"; una *fake news*, per Rini, deve quindi essere falsa in modo significativo, altrimenti il suo creatore non potrebbe sapere che lo è. Altri commentatori hanno un approccio più sfumato. Allcott e Gentzkow (2017), Gelfert (2018) and Jaster e Lanius (2018) concordano sul fatto che la pubblicazione di una *fake news* solitamente genera, come conseguenza, la diffusione di credenze false. Tuttavia, sostengono che asserire una falsità è solo uno dei modi possibili per istillare una credenza falsa. Si può perseguire efficacemente lo stesso fine con un'affermazione *fuorviante*, facendo cioè un'asserzione vera che implica qualcosa di falso. Jaster e Lanius, cercando un'espressione che colga entrambi i casi,

[3] Mukerji (2018, 936) non condivide questa intuizione, e ritiene che casi analoghi a questo andrebbero categorizzati diversamente, come esempi di cattivo giornalismo e non di *fake news*.

hanno sostenuto che le *fake news* sono caratterizzate della *mancanza di verità*, cosa che può "essere determinata dalla falsità di un resoconto giornalistico, o dalla sua natura fuorviante" (2018, 210).

La *mancanza di verità*, nel senso di Jaster e Lanius, non è una condizione sufficiente per le *fake news*. Ci sono almeno due tipi di casi a dimostrazione di questo fatto. Un articolo giornalistico può contenere errori madornali senza per questo essere accusato di veicolare *fake news*, se chi lo ha scritto era in perfetta buona fede. In secondo luogo, un buon divulgatore scientifico di solito è costretto a semplificare la propria materia al punto da raccontare una storia che, pur vera nelle sue linee generali, è piena di piccole falsità (Gelfert 2018, 99). Anche in questo caso, chiamare il testo di divulgazione scientifica *fake news* sembra un'esagerazione.

È molto più controverso se la mancanza di verità, nel senso di Jaster e Lanius, sia una condizione necessaria di una *fake news*. Per valutare adeguatamente ipotetici controesempi, cioè casi di *fake news* veicolate da asserzioni non caratterizzate dalla mancanza di verità in tal senso, è necessario prima discutere le proprietà intenzionali che un'asserzione deve avere perché il suo contenuto sia una *fake news*. Per concludere la discussione sulle proprietà epistemiche, e quindi per capire se la mancanza di verità o qualche altra proprietà epistemica sia necessaria per una *fake news*, dovremo attendere la fine di questo paragrafo.

Per proprietà intenzionali di un'asserzione intendiamo una serie di proprietà psicologiche che possono accompagnare o motivare l'asserzione di un certo contenuto. Nel presente contesto, sono particolarmente rilevanti le *intenzioni* che motivano l'atto di fare una determinata asserzione, e *l'atteggiamento* con il quale tale asserzione può essere fatta. Ad esempio, molti sostengono che per asserire una *fake news* sia necessario avere, con ciò, l'intenzione di ingannare i propri interlocutori (Dentith, 2017, 66; Gelfert, 2018, 108). Formulata in questi termini, però, la tesi è chiaramente falsa. Se infatti si interpreta l'atto di condividere un contenuto su un social media come Facebook o Twitter come asserzione di quel contenuto, è chiaramente possibile asserire una *fake news* senza volere con ciò ingannare nessuno; ad esempio, quando si condivide una *fake news* alla quale ingenuamente si crede[4]. Questo problema può comunque essere facilmente aggirato ricordando che la letteratura sull'argomento tende a definire le *fake news* nei termini delle asserzioni, e delle intenzioni che soggiacciono a tali asserzioni, dei creatori iniziali – e non dei molti propagatori – di contenuti diffusi on-line. Per il resto di questo saggio daremo per scontato che le asserzioni a cui le varie definizioni

[4] Siamo grati ad un anonimo revisore della rivista per aver sollevato questa obiezione. Per una discussione della questione se l'atto di condivisione comporti l'asserzione del contenuto condiviso, si veda *infra*, par. 3 di questo saggio.

di *fake news* in circolazione fanno riferimento sono le asserzioni di chi per la prima volta immette in circolazione un dato contenuto. Chiarita in questo modo, la tesi che l'asserzione di una *fake news* debba essere fatta con l'intenzione di ingannare spiega perché gli errori giornalistici fatti in buona fede non contino: perché appunto, per quanto l'articolo risultante sia falso, non è diffuso con l'intenzione di ingannare. Per la stessa ragione, capiamo perché anche il divulgatore scientifico non possa essere accusato di diffondere *fake news*: le sue asserzioni, lungi dal voler ingannare qualcuno, sono spiegate dal proposito opposto di veicolare in modo efficace qualche importante verità.

Jaster e Lanius (2018, 211) concordano sul fatto che l'intenzione di ingannare possa far sì che l'asserzione di un contenuto fuorviante, o l'asserzione di un contenuto letteralmente falso, sia l'asserzione di una *fake news*. Tuttavia, non credono che le *fake news* siano per forza diffuse con l'intento di ingannare. Nella loro prospettiva, le *fake news* mancano di quella che loro chiamano *veracità* (*truthfulness*). Solitamente, dicendo che un'asserzione è non verace si intende dire che è l'asserzione di un contenuto ritenuto falso da chi parla. Jaster e Lanius, però, intendono apparentemente qualcosa di diverso, ovvero un'asserzione fatta non con l'intenzione di riportare la verità. Naturalmente, un'asserzione fatta con l'intento di ingannare il proprio interlocutore è anche un'asserzione fatta *non* con l'intento di riportare la verità. Tuttavia, un'asserzione può essere fatta non con l'intento di riportare la verità anche se non è fatta con l'intento di ingannare. Perché ciò accada è sufficiente che il parlante o lo scrivente, nell'asserire un dato contenuto, sia *indifferente* alla sua verità. Un esempio di ciò sono le *bait farms* macedoni che producono e disseminano on-line notizie false e sensazionalistiche con il solo intento di generare traffico e, in proporzione, di trarre profitto dalla pubblicità[5]. I produttori e gli iniziali assertori di questi contenuti sono unicamente interessati al proprio profitto; da ciò segue che manchino di veracità solo nel secondo senso di non essere interessati alla verità di ciò che asseriscono, non anche nel senso di voler ingannare i propri destinatari.

Asserire un contenuto disinteressandosi alla sua verità, come Jaster e Lanius ricordano esplicitamente, non è altro che asserire una *stronzata* (*bullshit*) nel senso notoriamente definito da Frankfurt (2005). A quanti dicono stronzate in questo senso, come agli abitanti dei villaggi macedoni, non interessa se quello che dicono è vero o falso: selezionano quello che asseriscono, o lo inventano di sana pianta, perseguendo soltanto i propri obiettivi pratici.

L'analisi delle *fake news* nei termini della nozione frankurtiana di stronzata è al centro della proposta di Mukerji (2018). Anche Mukerji,

[5] Cfr. https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo?utm_term=.ybgwZKQVa#.fcjbDYNOZ (ultima consultazione effettuata in data 17.10.2019).

in particolare, ritiene che chi asserisce una *fake news* debba essere indifferente alla verità di ciò che dice[6]. Mukerji nota che chi asserisce una *fake news*, come chi dice una stronzata nel senso di Frankfurt, deve in un certo senso anche essere animato dal proposito di ingannare il proprio interlocutore. Come abbiamo già visto, ciò non è necessariamente vero relativamente ai fatti intorno a cui verte l'asserzione. Ma deve esserlo intorno al proprio atteggiamento nei confronti della verità: il proposito deve, in altre parole, essere quello di indurre i propri interlocutori a credere di essere interessati alla verità di quello che si dice.

Passiamo adesso all'analisi delle proprietà sociologiche delle asserzioni di *fake news*. Alcuni sostengono che una *fake news*, per essere tale, debba avere successo, cioè raggiungere il fine di una vasta circolazione (Gelfert, 2018). Altri, come Rini (2018), richiedono esclusivamente che le *fake news* siano fatte circolare con l'*obiettivo* di avere vasta circolazione e di ottenere un grande numero di condivisioni. Dal nostro punto di vista, un grande successo nelle condivisioni può essere *sufficiente*, quando anche altre condizioni sono soddisfatte, a rendere l'asserzione di un contenuto l'asserzione di una *fake news*. Tuttavia, respingiamo l'idea che ogni caso in cui questo non avviene debba per ciò *non* essere considerato come caso di *fake news*. A nostro avviso, da questo punto di vista non è tanto rilevante se una notizia abbia o meno una grande diffusione; ciò che è veramente rilevante è se abbia il *potenziale* per diffondersi in tal modo. Riteniamo cioè fondamentale, per stabilire se una certa asserzione sia o meno l'asserzione di una *fake news*, se tale asserzione sia fatta o meno attraverso un canale che sia appropriato, almeno nel senso di non impedire il raggiungimento di una vasta diffusione.

Un esempio proposto da Gelfert può essere utile in questo contesto. Gelfert racconta che, quando era uno studente di fisica, una persona mentalmente disturbata stava accanto alla fotocopiatrice nella biblioteca della sua facoltà e imponeva a chiunque volesse utilizzarla la lettura di un suo *pamphlet*, in cui riteneva di aver dimostrato la falsità della teoria della relatività di Einstein. Gelfert afferma, ragionevolmente, che distribuendo il pamphlet questa persona non avrebbe contribuito alla diffusione di una *fake news* neanche se fosse riuscito a confezionarlo in modo da imitare alla perfezione l'aspetto di un articolo di fisica rispettabile (2018, 102, n. 7). Per Gelfert ciò è dovuto al fatto che il *pamphlet* non è mai riuscito ad avere una grande diffusione. È nostra opinione, tuttavia, che l'esem-

[6] Mukerji prende in considerazione vari sensi in cui si può essere indifferenti alla verità di ciò che si asserisce. Si può essere indifferenti alla verità, ad esempio, quando non si annette alcun valore alla verità di ciò che si asserisce. Per Mukerji, l'assertore di una *fake news* non deve necessariamente essere indifferente alla verità in questo senso. È sufficiente che sia indifferente in un senso più debole, codificato dal fatto che avrebbe fatto la stessa asserzione anche se fosse stata falsa. Questo senso controfattuale è chiaramente più debole del primo: si può infatti annettere valore all'asserire qualcosa di vero pur conservando la disposizione a fare la stessa asserzione anche nel caso in cui sia falsa.

pio illustri un'altra morale, stando alla quale un contenuto, per essere caratterizzato come *fake news*, debba essere asserito in un modo che non comprometta il fine di ottenere una vasta diffusione. Un indizio di questo fatto è che se il pamphlet fosse stato pubblicato on-line, saremmo stati più inclini a caratterizzarlo come una *fake news* indipendentemente dal suo successo, perché la modalità della sua diffusione sarebbe stata per lo meno compatibile con il perseguimento di tale fine.

Come è stato fatto osservare da più parti, l'asserzione di una *fake news* deve anche possedere proprietà formali che la facciano apparire come l'asserzione di una notizia genuina. Un suggerimento del genere, naturalmente, è vago e destinato a lasciare aperte una serie di domande. In questa sede, ci limitiamo alla questione relativa a *dove* una *fake news* debba essere presentata in modo da sembrare una notizia genuina. In risposta a questa domanda, c'è chi dice che una *fake news* debba essere pubblicata on-line (e.g. Klein e Wueller 2017, 6). Altri, ad esempio Rini (2017, E-45), si limitano a registrare l'esistenza di una stretta connessione tra social media e *fake news*. Mukerji (2018, 927) mette in guardia rispetto al pericolo di scambiare una relazione contingente per una connessione concettuale, e offre alcuni esempi di *fake news* diffuse attraverso media tradizionali come giornali e radio. In questo dibattito concordiamo con Mukerji (e probabilmente con Gelfert 2018). Riteniamo in altre parole che le *fake news*, diversamente dall'espressione che oggi utilizziamo per designarle, siano precedenti alla nascita e alla diffusione dei social media. Nonostante ciò, pensiamo che i nuovi media abbiano avuto un ruolo decisivo nell'imprimere alla diffusione di *fake news* una nuova forma, al punto da conferire al fenomeno tratti che non hanno precedenti nella sua storia. Ciò è dovuto a due principali ragioni. Da un lato, come osservato da Gelfert (2018, 102), la nascita del giornalismo on-line ha contribuito a rendere più facile la pubblicazione in rete di contenuti che esibiscano "molti dei tradizionali tratti distintivi del giornalismo serio". Pertanto, se non proprio il fenomeno delle *fake news*, almeno le sue dimensioni sono state fortemente influenzate dall'insorgenza dei nuovi media. D'altra parte, i social media hanno creato opportunità di profitto senza precedenti per tutti coloro che intendano sfruttare la pubblicità *pay-per-click*, mettendo a disposizione in modo immediato un vasto pubblico di potenziali visitatori. Pertanto, oltre a influenzare le dimensioni del fenomeno delle *fake news*, il mezzo attraverso il quale sono principalmente diffuse ha ampliato lo spettro dei possibili *motivi* e *scopi* alle sue spalle.

Adesso che abbiamo concluso la nostra discussione delle proprietà delle asserzioni che possono legittimamente essere considerate asserzioni di *fake news*, possiamo riprendere l'analisi delle proprietà epistemiche di tali asserzioni dove l'avevamo lasciata, ripartendo dalla domanda se la *mancanza di verità* – intesa come falsità tout court o verità letterale accompagnata dall'implicazione di un contenuto falso – sia una proprietà necessaria delle *fake news*. Qualche commentatore ha riconosciuto

che alcuni possibili casi sembrano incoraggiare una risposta negativa a tale domanda (Jaster e Lanius, 2018; Mukerji, 2018). Se tale risposta è corretta, come noi pensiamo che sia, il difetto epistemico delle *fake news* non può consistere nella loro mancanza di verità, ma deve essere identificato altrove. In quanto segue, ci proponiamo di descrivere questo secondo difetto epistemico, e di mostrare che tale difetto è esibito anche dalle più comuni *fake news* afflitte da *mancanza di verità*.

I casi in oggetto riguardano *fake news* che risultano *accidentalmente vere*. Si consideri il seguente esempio (Jaster e Lanius, 2018, 218). *Russia Today* (un sito realmente esistente) pubblica un articolo stando al quale Hillary Clinton ha evaso il fisco. La storia è completamente *inventata*, e viene pubblicata solo per favorire l'ascesa di Donald Trump alla Casa Bianca gettando discredito sulla sua avversaria politica. La storia diventa virale, e milioni di Americani si convincono del fatto che la Clinton non abbia pagato le tasse. Il tratto peculiare di questo esempio è che di fatto gli elettori americani non si sbagliano perché, all'insaputa degli autori dell'articolo, Hillary Clinton ha realmente evaso il fisco nel periodo indicato. La domanda è se, pubblicandola, gli autori abbiano asserito o meno una *fake news*.

In risposta, Jaster e Lanius ammettono che, prendendo in considerazione il caso in oggetto, si possa essere tentati di rispondere che la storia su Hillary Clinton sia una *fake news*. Tuttavia, sostengono che sia una tentazione a cui è necessario resistere[7]. Per farla svanire, sostengono che sia sufficiente distinguere più accuratamente tra la diffusione di una *fake news* e il mero tentativo di diffondere una *fake news*. Facendolo, ci accorgeremo del fatto che la pubblicazione della storia su Hillary Clinton comporta solo il tentativo e non l'effettiva diffusione. Per chiarire il concetto, Jaster e Lanius propongono un'analogia con la menzogna. Secondo i due autori, noi distinguiamo (e puniamo proporzionalmente)

[7] Mukerji persegue una strategia simmetrica: ammette che molti potrebbero avere l'intuizione contraria, secondo cui una notizia accidentalmente vera non sia una *fake news*, e sostiene che *questa* sia una tentazione da contrastare. In particolare, suggerisce che la tentazione di non guardare a *Russia Today* come a una *fake news* potrebbe essere spiegata dal cosiddetto *bias* del senno di poi (*hindsight bias*). Questo bias ci porta ad avere una percezione distorta della prevedibilità di un evento una volta che si è verificato. Nel caso in questione, ci porterebbe inavvertitamente a ricostruire l'esempio in modo diverso da come è stato descritto, e a pensare che l'autore della storia su Hillary Clinton non potesse che sapere, o comunque non potesse che avere ottime ragioni per pensare, che davvero avesse evaso il fisco. Se davvero lo avesse saputo o avesse avuto ragioni per pensarlo, la storia sarebbe stata genuina e la sua pubblicazione non sarebbe contata come l'asserzione di una *fake news*. Ma appunto, per quanto si possa tendere a dimenticarlo, nell'esempio l'autore della storia ignora che la Clinton abbia evaso il fisco, e che il suo articolo sia vero è un mero caso.

[8] L'analogia di Jaster e Lanius si estende anche all'omicidio (e al mero tentativo di commetterne uno). Per contenere la discussione, nel testo ci limitiamo a commentare la più stringente analogia con la menzogna.

il mentire dal mero tentativo di mentire[8]. Quando tentiamo di mentire, asseriamo un contenuto che riteniamo falso con l'intento di ingannare i nostri interlocutori; ma quando ciò che asseriamo di fatto è vero, non abbiamo mentito, abbiamo *soltanto* cercato di farlo. In modo del tutto analogo, quando tentiamo di diffondere una *fake news*, pubblichiamo un contenuto che riteniamo falso con l'intento di ingannare i nostri destinatari; ma quando la storia che abbiamo pubblicato di fatto è vera, non abbiamo diffuso una *fake news*, abbiamo soltanto *cercato* di farlo.

Questo argomento per analogia è facile da smontare. La letteratura sulla menzogna è infatti fondamentalmente unanime nel ritenere che per mentire non sia necessario asserire qualcosa di falso. La definizione standard di menzogna sostiene infatti che una menzogna sia l'asserzione di un contenuto che si ritiene falso con l'intento di fare accettare questo contenuto come vero al proprio interlocutore (Isenberg, 1964; Chisholm e Feehan, 1977; Primoratz, 1984; Williams, 2002; Mahon, 2008; Lackey, 2013). Stando a questa definizione, possiamo stare mentendo anche quando diciamo qualcosa di vero, a condizione di pensare di stare dicendo qualcosa di falso, e di starlo facendo con l'intento di ingannare il nostro interlocutore. Se la menzogna è compatibile con l'asserzione di un contenuto vero, il parallelo tra *fake news* e menzogna non sostiene la conclusione che per diffondere una *fake news* sia necessario diffondere un contenuto falso. Al contrario, incoraggia l'opinione opposta, stando alla quale un caso come quello di *Russia Today* possa coinvolgere una *fake news*.

Un sintomo iniziale del fatto che le cose stiano così è il seguente. Nell'esempio, gli elettori americani formerebbero la credenza vera che Hillary Clinton ha evaso il fisco. Tuttavia, sembra intuitivo concludere che non avrebbero *conoscenza* di questo fatto. Gli elettori americani formerebbero infatti una credenza vera sulla base di una storia che è stata inventata per ingannarli. La loro posizione, in questo caso, assomiglierebbe molto a quella dei protagonisti dei casi Gettier, i quali hanno credenze vere formate in modo troppo accidentale per avere conoscenza (Gettier, 1963). Come i protagonisti dei casi Gettier, gli elettori americani arriverebbero a credere il vero in modo troppo accidentale per avere conoscenza. E questo è esattamente quello che ci aspetteremmo da una *fake news*: che le credenze formate attingendo a una simile fonte non siano, neanche se vere, conoscenze.

Se prendiamo questo suggerimento sul serio, ci resta da affrontare la questione di quale sia la proprietà epistemica, se non la mancanza di verità nel senso di Jaster e Lanius, che rende la storia su Hillary Clinton una *fake news*. A nostro modo di vedere, un candidato molto più promettente è che la storia sull'evasione fiscale non sia basata su evidenza, ovvero su ragioni per pensare che fosse vera, ma sia stata *inventata* di sana pianta[9]. Nello specifico, è naturale affermare che una condizione

[9] È degno di nota che Jaster e Lanius (2018, 220) prendano in considerazione questa

necessaria dell'asserzione di una *fake news* sia non che tale asserzione manchi di verità nel senso di Jaster e Lanius, ma che il suo primo assertore ritenga di non avere evidenza sufficiente a favore della propria asserzione[10]. Questo ci porta alla seguente definizione di *fake news*:

L'asserzione/implicatura che P da parte di S è l'asserzione/implicatura di una *fake news* se e solo se (i) tale asserzione/implicatura è fatta in modo da raggiungere un pubblico sufficientemente esteso[11], (ii) S crede di non avere evidenza sufficiente per P, (iii) l'asserzione/implicatura è fatta con l'intenzione di ingannare i propri destinatari o con un atteggiamento di indifferenza rispetto alla verità di P.

La condizione (i) riassume le conclusioni a cui ci ha portato la nostra discussione delle proprietà sociologiche delle asserzioni di una *fake news*. Tale condizione richiede che per essere una *fake news*, un contenuto, se non raggiunge una vasta diffusione di pubblico, debba essere almeno veicolato in un modo che normalmente porterebbe a tale esito. In altre parole, il soddisfacimento della condizione (i) sembra garantire che la popolarità o la viralità di P, come possibili esiti della asserzione di P da parte di S, siano risultati non accidentali. La condizione (iii) specifica le proprietà intenzionali delle asserzioni di *fake news* in modo disgiuntivo, richiedendo che l'asserzione sia motivata dall'intenzione di ingannare i

risposta alternativa. Tuttavia difendono la scelta di metterla da parte sostenendo che la risultante definizione di *fake news* sarebbe estensionalmente equivalente a quella da loro proposta nei termini della nozione epistemica di mancanza di verità. Questa ragione è a nostro avviso sorprendente, perché le due definizioni chiaramente *non* sarebbero equivalenti. Le due definizioni, in particolare, darebbero risultati divergenti proprio nei casi che stiamo prendendo in considerazione, in cui un contenuto asserito nella forma di una pubblicazione è vero accidentalmente. Mentre l'analisi di Jaster e Lanius implica che una simile asserzione *non* sia l'asserzione di una *fake news*, la definizione basata sulla mancanza di evidenza permette di affermare il contrario.

[10] E è evidenza sufficiente per P se e solo se E è un insieme sufficientemente comprensivo di evidenza sulla base del quale S potrebbe credere giustificatamente che P. È importante notare che S potrebbe non avere sufficiente evidenza per P e credere, ciononostante, di averla. Questo può avvenire in una varietà di casi differenti. Ad esempio, quando S credere di avere evidenza sufficiente per P, di fatto l'evidenza di S sostiene P, ma tale evidenza è *incompleta*, nel senso che facilmente S avrebbe potuto acquisire ulteriore evidenza contraria a P. Quando S afferma P sulla base di evidenza incompleta in questo senso, ma S crede che l'evidenza sia sufficiente per credere P, la sua affermazione può essere criticata come precipitosa o affrettata, ma non come l'asserzione di una *fake news*.

[11] Una preoccupazione legittima, rispetto a questa clausola, è l'innegabile vaghezza di "sufficientemente esteso". Per attenuare tale preoccupazione, ricordiamo che la stessa vaghezza sembra affliggere la non problematica nozione di *news*, e l'estensione del pubblico a cui deve essere indirizzato un contenuto per contare come una notizia giornalistica. Ciò permette di rispondere alla perplessità sopra ricordata che un pubblico sufficientemente esteso è il pubblico a cui è indirizzata quella che non problematicamente considereremmo una comune notizia giornalistica.

propri destinatari, o sia fatta con l'atteggiamento tipico di chi racconta stronzate, ovvero nel pieno disinteresse per la verità e con l'intenzione di ingannare i propri interlocutori circa il proprio atteggiamento.

Queste osservazioni relative alle proprietà intenzionali delle *fake news* ci aiutano a situare questo fenomeno rispetto a fenomeni affini, come le menzogne e, appunto, le stronzate. In sintesi, alla luce della nostra analisi le *fake news* o sono specie del genere menzogna o sono specie del genere stronzata. Una differenza specifica che le accumuna è la modalità della loro diffusione, come richiesto da (i). Oltre a ciò, come richiesto da (ii), sono bugie o stronzate proferite da parte di soggetti che ritengono di non avere evidenza sufficiente a favore di quello che dicono. Come abbiamo già visto, si tende a pensare che chi mente debba credere che quanto asserisce sia falso. Dal momento che di solito chi crede che P sia falso non ritiene di avere evidenza che supporta P, si potrebbe osservare che tutte le menzogne, e non solo le *fake news* che sono una specie del genere menzogna, soddisfino la condizione (ii) della nostra definizione. Questo non è però vero del tipo specifico di stronzate che secondo la nostra analisi sono *fake news*. Chi dice stronzate non vuole di solito ingannare il proprio interlocutore. Come dimostrato dai creatori di *fake news* macedoni, è possibile non avere alcun riguardo per la verità e contemporaneamente nutrire un forte interesse – di solito economico – ad inondare il proprio pubblico di notizie sensazionalistiche. Svolgere questa attività è del tutto compatibile col ritenere di avere ragioni per credere alle notizie che uno diffonde, purché tali notizie soddisfino gli standard richiesti per attrarre molte condivisioni. Quando questo succede, e chi pubblica una notizia sensazionalistica pensa di avere buone ragioni per credere che sia vera, la notizia intuitivamente non è una *fake news*. La condizione (ii) è pertanto fondamentale per distinguere le stronzate che sono *fake news* dalle stronzate che intuitivamente non lo sono[12].

3. Come si propagano le *fake news*: un'analisi descrittiva

Prima di addentrarci nel dibattito sugli aspetti normativi della diffusione di *fake news* sui social media, è opportuno fare chiarezza sulle

[12] Un buon esempio di stronzata che non è una *fake news* è probabilmente il seguente. Supponiamo che una rivista scandalistica pubblichi un articolo che attribuisce a due *vip* una relazione clandestina. L'articolo è pubblicato insieme a una fotografia che ritrae i due ipotetici amanti in una situazione di tenerezza. La foto è talmente esplicita che l'autore dell'articolo, intenzionato a fare vendere copie alla propria rivista, lo avrebbe pubblicato comunque, sia che ritenesse la foto vera sia che la ritenesse contraffatta. Questo sembra indicare che la pubblicazione dell'articolo sia una forma di *bullshitting*. Supponiamo tuttavia che l'autore dell'articolo ritenga la fotografia evidenza sufficiente per credere che i due abbiamo una relazione. In questo caso, non starebbe anche diffondendo una *fake news*.

caratteristiche *descrittive* del fenomeno che stiamo considerando. Nello specifico, le due questioni che analizziamo al fine di comprendere meglio la proliferazione di disinformazione riguardano, l'una, i ruoli che ciascun utente può giocare all'interno del processo di distribuzione di notizie sui social network e, l'altra, i fattori causali che spiegano il comportamento di ciascun utente in base al ruolo che questi svolge.

Procediamo con ordine. A una prima analisi, sembra naturale distinguere due ruoli chiave: quello dei *produttori* di *fake news* – ovvero coloro le cui affermazioni rispettano i requisiti delle *fake news* che abbiamo sviluppato nel paragrafo precedente – e quello dei *consumatori* di *fake news* ovvero coloro che sono di fatto i destinatari dalle asserzioni dei produttori di tali messaggi. Il consumo di *fake news* non deve necessariamente portare alla formazione di credenze nel loro contenuto. Una *fake news* può essere 'consumata' al solo scopo di divagarsi o al fine di smascherarne l'impostura o la falsità. Le varie tipologie di consumatori pongono questioni interessanti che meriterebbero ben più di una categorizzazione sommaria. Tuttavia, in questo lavoro ci limitiamo a considerare la categoria problematica di coloro che consumano *fake news* finendo per accettare come vero il loro contenuto: li chiameremo *ricettori*.

L'analisi bipartita appena offerta applica la distinzione tra parlante e ascoltatore tipica dell'epistemologia della testimonianza all'indagine epistemologica delle *fake news*. A ben vedere, questo modello non è del tutto soddisfacente: nella maggior parte dei casi, i *produttori* di *fake news* – a differenza dei classici testimoni – non trasmettono le proprie credenze ai ricettori in maniera diretta. Nel caso delle *fake news*, i ricettori fanno loro un contenuto che appare nelle loro 'bacheche' (le *news feed*) o in quelle dei loro contatti. Questa dinamica più articolata richiede l'introduzione di una terza figura, perlopiù assente nei tradizionali scambi testimoniali, ovvero i *propagatori*. Il propagatore è un individuo che condivide *fake news* con i propri contatti sui social network ripostando o ritwittando contenuti in cui si imbatte. Questa categoria di utenti si distingue dalle altre per almeno due ragioni fondamentali. A differenza dei ricettori, non è necessario che i propagatori credano nel contenuto delle informazioni che condividono. A differenza dei produttori – cosa ancora più importante – non è necessario che i propagatori condividano pezzi di (dis-)informazione al fine di asserirne i rispettivi contenuti. Un propagatore potrebbe condividere un post o un tweet perché ne ritiene vero il contenuto o più semplicemente perché lo ritiene curioso, interessante, divertente, ridicolo, stravagante, e così via. Come ha giustamente sottolineato Rini (2017, E-48), le persone sono sempre felici di ammettere che credevano in una notizia condivisa sui social quando questa risulta essere vera, ma non lo sono se vengono sollevati dubbi circa l'accuratezza dell'informazione[13].

[13] Secondo Lazer *et al*. (2018, 1095), la diffusione di *fake news* sui social network

Dal punto di vista dei produttori di *fake news*, la presenza dei propagatori gioca un ruolo cruciale per almeno due ragioni: in primo luogo, perché i propagatori contribuiscono alla diffusione di disinformazione sui network; in secondo luogo, perché è probabile che ripostando o ritwittando una *fake news* dal contenuto politicamente, socialmente o moralmente impegnato che non ci aspetteremmo di vedere condivisa da un produttore, i propagatori consentono ai produttori stessi di raggiungere fasce di utenti che non condividono i loro valori e, conseguentemente, di offrire a tali utenti ragioni (*pro tanto*) per credere nel contenuto delle *fake news*. Come risulterà chiaro dal resto di questo paragrafo, molte spiegazioni causali della diffusione di *fake news* chiamano in causa i valori e i pregiudizi cognitivi degli utenti.

Prima di affrontare la seconda questione di questo paragrafo, occorre sottolineare che i tre ruoli degli utenti dei social network sono piuttosto dinamici. Potrebbe benissimo darsi il caso che un utente svolga tutti e tre i ruoli in differenti circostanze o, se può sembrare insolito che un utente qualsiasi produca *fake news*, che molti utenti abbiano svolto il ruolo di propagatori almeno una volta, al di là dei loro intenti più o meno genuini[14]. A prescindere dalla probabilità che queste combinazioni di ruoli si verifichino nelle nostre interazioni ordinarie sui social media, è opportuno che una epistemologia delle *fake news* tenga conto di questa possibilità.

La questione che affrontiamo nel resto di questo paragrafo è *come* tutto ciò possa accadere, ovvero quali meccanismi psicologici e relativi alla struttura dei network assicurino la proliferazione di *fake news*. Distinguiamo tra fattori *individuali*, che innescano la propensione di singoli individui a condividere o accettare come vere delle *fake news*, e fattori *sociali*, che influenzano le reazioni alle *fake news* di gruppi di persone ad un livello collettivo.

Al livello dei fattori individuali, il primo elemento che favorisce la propagazione di *fake news* è senza dubbio il ridotto livello di attenzione che prestiamo ai contenuti dei social media, spiegabile – almeno in parte – come reazione all'enorme mole di dati e informazioni che i social media mettono a nostra disposizione o ci presentano direttamente nelle nostre bacheche. Una recente ricerca di Microsoft Canada ha dimostrato

si fonda su una forma di appoggio implicito (*implicit endorsement*) tipico dell'atto stesso di ripostare o ritwittare un contenuto.

[14] Come riportato da *Statista.com* (cfr. https://www.statista.com/statistics/657111/fake-news-sharing-online/; ultima consultazione effettuata in data 17.10.2019), il 23% dei partecipanti a un sondaggio svolto negli Stati Uniti nel 2016 ha ammesso di avere condiviso una presunta notizia politica poi rivelatasi falsa o una presunta notizia politica che sapeva essere falsa al momento della condivisione. Pertanto, possiamo ovviamente aspettarci una percentuale ancora più alta di individui che abbiano ammesso di aver condiviso almeno una *fake news* dal contenuto politico o che non abbiano realizzato di aver condiviso una *fake news* dal contenuto politico.

che i nostri stili di vita digitale riducono il livello di attenzione sostenuta (*sustained attention*) che prestiamo nelle attività online, diminuendo la nostra capacità di rimanere concentrati per tempi prolungati durante attività ripetitive[15]. Questo tipo di considerazioni sul livello di attenzione che devolviamo ai contenuti sui social media ci permette di spiegare la diffusione delle *fake news* a livello generale. In uno studio recente (2017), Qiu e colleghi hanno dimostrato che "sia il volume di informazioni sia la limitata attenzione conducono ad un potere discriminatorio basso" (p. 5), ovvero alla nostra incapacità di stabilire se i nuovi contenuti che appaiono nelle nostre bacheche offrano informazioni attendibili. In ogni caso, ad oggi non è chiaro se un potere discriminatorio basso conduca direttamente alla viralità della disinformazione: Qiu e colleghi (2019) hanno recentemente modificato la loro analisi originale, ammettendo di non avere raccolto evidenza che l'informazione di bassa qualità abbia la stessa probabilità di diventare virale dell'informazione di alta qualità.

Indipendentemente dal fatto che livelli ridotti di attenzione sostenuta possano contribuire o meno alla diffusione di *fake news*, è indubbio che la diffusione di falsità – almeno su Twitter – sia significativamente più ampia, più rapida e più profonda della diffusione di verità in qualsiasi categoria di informazioni (Vosoughy *et al*. 2018, 1147). Come anticipato nell'introduzione, lo studio di Vosoughy e colleghi mostra che i contenuti falsi hanno il 70% di probabilità in più di essere ritwittati dei contenuti veri (1149). Inoltre, quegli utenti che hanno maggiore probabilità di rimanere intrappolati nelle cosiddette 'cascate di *fake news*' rispondono, in generale, alle seguenti caratteristiche: hanno meno *followers* di coloro che diffondono informazioni affidabili; seguono meno utenti; sono meno attivi sui social media; sono verificati meno spesso; e utilizzano i social media da meno tempo. Da quanto abbiamo potuto verificare, non ci risultano essere studi che offrano una spiegazione complessiva del perché quegli utenti siano più propensi a condividere *fake news* di altri.

Tuttavia, vi è evidenza che ulteriori fattori contribuiscano alla spiegazione della nostra generale propensione a credere e a condividere disinformazione. Uno di questi fattori è rappresentato dai pregiudizi cognitivi (*biases*), in particolare dal *bias* della conferma e dal *bias* della desiderabilità. Il primo conduce l'utente a ricercare informazioni che confermino ciò di cui l'utente è già convinto e a ignorare informazioni che contrastino con la sua opinione. Come Tom Nichols ha sottolineato (2017, 64), il *bias* della conferma è una sorta di strumento di sopravvivenza, che si attiva per evitarci di venire sommersi dall'inarrestabile flusso di informazioni sui social media. Il problema, però, è che questo

[15] Microsoft Attention Span Research Report, Microsoft Canada (cfr. https://www.scribd.com/document/265348695/Microsoft-Attention-Spans-Research-Report; ultima consultazione effettuata in data 17.10.2019).

bias ci rende insensibili all'evidenza contraria alle nostre credenze, pre-comprensioni, e visioni normative, compromettendo così la nostra capacità razionale di valutare nuova informazione. Per esempio, se i negazionisti del cambiamento climatico vengono posti di fronte ad evidenza che il nostro pianeta si sta surriscaldando, essi saranno più propensi a ignorarla o screditarla.

L'altro pregiudizio, ossia il *bias* della desiderabilità, conduce l'utente ad attribuire più credibilità a informazioni che questi desidera ricevere rispetto a informazioni che non desidera ricevere. Così facendo, gli utenti dimostrano una tendenza a tener conto dell'evidenza che conferma le proprie credenze e a trascurare quella contraria (Tappin *et al*., 2017, 1143). Nel caso delle *fake news*, il bias della desiderabilità potrebbe operare come segue: se un individuo è propenso a credere che l'aumento dei fatti criminosi in una città sia dovuto alla presenza di immigrati, egli sarà più propenso a credere alle *fake news* che mostrano e condannano il nuovo arrivo di centinaia di rifugiati[16]. In maniera analoga, se un individuo è già convinto che Hillary Clinton sia una persona moralmente spregevole e pervertita, sarà anche più propenso a credere in una storia falsa relativa al suo coinvolgimento nel traffico di bambini a scopi sessuali.

Come abbiamo già tentato di mostrare, i *bias* cognitivi costituiscono un fattore rilevante nella spiegazione di come le *fake news* si diffondano sui social media, nella misura in cui fanno leva sui limiti della nostra razionalità nella gestione dell'informazione. Altri fattori, potenzialmente più sorprendenti, includono ciò che gli psicologi chiamano *meccanismi basati sulla memoria* ed *effetti di fluidità* (*fluency effects*). A differenza dei precedenti *bias*, questi meccanismi non agiscono impedendoci di riconoscere in quanto tale una *fake news*, ma intervengono in un secondo momento, quando abbiamo già riconosciuto la *fake news* come tale. L'effetto del primo tipo di fattori è quello di farci dimenticare che il contenuto della presunta notizia proviene da una fonte inaffidabile, e nel peggiore dei casi, quello di portarci erroneamente ad attribuire tale contenuto ad una fonte affidabile (Levy 2017, 29). Il secondo fattore, invece, spiega perché l'esposizione ripetuta a *fake news* che l'utente sa essere false o fuorvianti aumenta le possibilità che questi ne dimentichi la fonte, conducendolo ad attribuire all'autore un grado di affidabilità che non possiede e, di conseguenza, a ritenere di possedere ragioni per credere nel contenuto della stessa *fake news* (p. 30). Gli effetti di fluidità rappresentano un fattore causale particolarmente rilevante per spiegare la proliferazione di *fake news* sui social media. Come risulterà chiaro nel resto del paragrafo, diverse caratteristiche strutturali dei social network facilitano la comparsa ripetuta di contenuti nella propria bacheca e

[16] Per una analisi dettagliata delle *fake news* relative ai rifugiati cfr., ad esempio, https://teyit.org/en/how-is-false-information-used-worldwide-to-target-refugees/ (ultima consultazione effettuata in data 17.10.2019).

quindi rendono più probabile che la fluidità con cui l'utente le processa ne venga influenzata.

Consideriamo i tre elementi socialmente rilevanti per rendere conto della diffusione di *fake news* sui social network: le bolle epistemiche, le camere d'eco e la polarizzazione di gruppo. In un recente articolo di Thi Nguyen (2018) emerge chiaramente che sia le bolle epistemiche sia le camere d'eco costituiscono strutture di esclusione, in quanto impediscono a larghi gruppi di utenti dei social media di acquisire consapevolezza di – o di tenere nella dovuta considerazione – alcuni tipi di informazione. Entrambi i fattori contribuiscono a generare e sostenere una sorta di scollegamento ideologico, poiché chi ne fa parte finisce per perdere contatto con l'informazione e le opinioni che sono in contrasto con la visione dei gruppi a cui egli appartiene.

Tuttavia, le bolle epistemiche si creano e funzionano in maniera differente rispetto alle camere d'eco. Le prime sono strutture sociali che impediscono la circolazione di un determinato insieme di informazioni *per omissione*, semplicemente impedendo che la testimonianza proveniente da fonti che sostengono una prospettiva rivale sia immessa nel circuito. Un esempio è la rete dei propri contatti di Facebook. Il meccanismo di selezione in questo caso è dato dal criterio con il quale scegliamo i nostri contatti, ovvero sulla base di affinità ideologiche e normative. In questo modo, la nostra rete finisce inevitabilmente per escludere informazione più probabilmente condivisa da prospettive rivali. Questo procedimento di esclusione è supportato da tendenze ad un livello individuale, quali il *bias* della conferma o l'esposizione selettiva, che possono portare un individuo ad attribuire un eccesivo credito epistemico alle proprie credenze perché la sua comunità è costruita in maniera tale da includere soltanto individui che le condividono. Comunque, è nella natura delle bolle che esse possano essere scoppiate facilmente: nello specifico, le bolle epistemiche spariscono non appena alcuni membri siano esposti ad informazioni che erano state oscurate dalla loro vista.

Le camere d'eco, invece, ostruiscono il consumo di informazione in maniera più attiva: in particolare, l'appartenenza ad una camera d'eco richiede che i suoi membri accettino preventivamente un insieme di meta-credenze che distribuiscono il credito epistemico in maniera asimmetrica tra coloro che fanno parte della camera e coloro che ne sono al di fuori. Di fatto, tali meta-credenze ingigantiscono eccessivamente il grado di fiducia che si ripone nei primi e riducono eccessivamente il grado di fiducia che si ripone nei secondi a prescindere dalla loro effettiva affidabilità. Più precisamente, il discredito epistemico può essere concepito nei termini di inaffidabilità delle fonti, ma anche nei termini dei vizi intellettuali che esse dimostrano di possedere, quali la malizia epistemica, la chiusura mentale e la disonestà intellettuale.

Nel caso delle camere d'eco, si attiva quello che Nguyen chiama meccanismo di rinforzo del disaccordo (*disagreeing-reinforcing mechani-*

sm). Tale meccanismo neutralizza l'effetto della contro-evidenza e delle informazioni contrastanti allertando i membri della camera d'eco che la proposta di contro-evidenza è proprio ciò che dovrebbero aspettarsi da chi non appartiene alla loro comunità ed è epistemicamente corrotto. Per contro, il mero fatto che i membri della comunità condividano le stesse credenze e valutino le informazioni nello stesso modo offre a ciascuno di loro ulteriore evidenza che i loro compagni sono degni di fiducia, incrementando così i livelli interni di credenza nelle camere d'eco. Questo meccanismo rende ancora maggiore la differenza tra le bolle epistemiche e le camere d'eco: paradossalmente, nel secondo caso, a differenza del primo, l'esposizione ad un più ampio spettro di informazioni può avere un effetto contrario alle aspettative, ovvero quello di contribuire al discredito di opinioni contrarie e all'innalzamento dei livelli di fiducia interna ai membri della camera stessa.

Al di là di queste differenze, dovrebbe risultare chiaro che sia le bolle epistemiche sia le camere d'eco favoriscono il proliferare di *fake news* sui social network. Per esempio, alcuni post falsi che contengono affermazioni relative allo sbarco di rifugiati sulle coste meridionali italiane, associati ad immagini completamente indipendenti ritraenti barche colme di migranti, hanno ottime probabilità di diffondersi all'interno di gruppi nazionalistici che condannano le politiche sull'asilo ai rifugiati. Se i membri di tali gruppi non solo ignorassero le fonti di *fact-checking* a proposito della verità dei post in questione, ma fossero anche pronti a smettere di credere ai contenuti di tali post non appena la comunità venga esposta a contro-evidenza, tali bolle epistemiche potrebbero essere scoppiate facilmente. L'unica preoccupazione che rimarrebbe in tale circostanza sarebbe soltanto che la reiterazione di post simili può dar vita agli effetti di fluidità di cui abbiamo parlato in precedenza. Al contrario, se questa storia coinvolge un gruppo in cui le fonti di *fact-checking* sono discreditate in modo aprioristico e squalificate come agenzie di stampa corrotte e controllate da gruppi di interesse rivali, allora i membri del gruppo saranno propensi a resistere all'evidenza contraria, e piuttosto ad interpretarne l'esistenza come conferma della propria visione del mondo e della società. In tal modo, la loro credenza che moltitudini di migranti illegali stanno entrando nei porti italiani verrà ulteriormente rinforzata, come l'analisi delle camere d'eco predice.

L'ultimo fattore che è opportuno considerare in questo paragrafo è la polarizzazione di gruppo, ossia la tendenza di un gruppo a mostrare credenze e atteggiamenti più estremi di quelle che i suoi membri possiedono se presi singolarmente. La polarizzazione di gruppo è la giusta conclusione della nostra analisi dei fattori che spiegano la diffusione di *fake news*, in quanto questa tendenza coinvolge molti – se non tutti – i fattori menzionati in precedenza. In particolare, i bassi livelli di attenzione rivolti ai contenuti dei social media spiegano perché le persone sono riluttanti a cambiare opinione (Riva, 2017; Sunstein, 2017), mentre

il *bias* della conferma e le camere d'eco spiegano perché le dinamiche collettive conducano i membri di un gruppo a polarizzare le proprie opinioni verso visioni sempre più radicali, aumentando il livello di convinzione nelle proprie opinioni e considerando l'ampio (dis)accordo con chi (non) appartiene al proprio gruppo come ulteriore evidenza a supporto di tali opinioni (Nguyen, 2018; Sunstein, 2017)[17]. In altre parole, la polarizzazione di gruppo intensifica gli effetti degli altri fattori analizzati, aumentando la probabilità che gli utenti accettino e condividano *fake news* il cui contenuto supporta le loro visioni e mettano da parte l'informazione che contraddice tali visioni etichettandole come teorie cospiratorie.

A conclusione di questo paragrafo, è opportuno sottolineare che le considerazioni proposte offrono semplicemente un'analisi concisa dei lavori più recenti in un campo che sta crescendo a velocità sostenuta. Da una parte, ciò spiega perché dobbiamo evitare di offrire conclusioni forti circa le cause della diffusione di disinformazione sui social media. Dall'altra, dovrebbe alimentare la speranza che nei prossimi anni saremo in grado di avere risposte conclusive alla questione in oggetto.

4. Questioni normative

In questo paragrafo prendiamo in considerazione la dimensione normativa del fenomeno delle *fake news*. Nel precedente, abbiamo proposto un modello che distingue tre ruoli: il produttore di *fake news*, il propagatore e il ricettore finale. Alla luce di questo modello, è naturale affrontare la dimensione normativa delle *fake news* a partire dalle dimensioni normative dei tre ruoli. Occorre tuttavia essere espliciti su due limitazioni importanti della nostra analisi. Primo, tralasceremo la dimensione morale delle *fake news*, ovvero i problemi morali connessi alla propagazione di contenuti falsi o ingannevoli, per concentrarci esclusivamente sulla dimensione epistemica delle *fake news*. Secondo, anche se la nostra analisi contemplerà qualche considerazione circa l'epistemologia della propagazione delle *fake news*, la nostra attenzione sarà principalmente rivolta alla dimensione epistemica della ricezione di *fake news* e ignorerà gli aspetti epistemici della loro produzione. Questa seconda limitazione della nostra analisi è dovuta al fatto che i produttori di *fake news* perseguono un obiettivo politico o finanziario. Invece, è ragionevole pensare che i propagatori e i ricettori delle *fake news* siano coinvolti in una attività principalmente epistemica, relativa alla ricerca e alla distribuzione di informazioni sul mondo in cui viviamo. Pertanto, riteniamo che la valutazione epistemica di questa pratica costituisca

[17] Per una analisi più dettagliata dell'esposizione a informazioni ideologicamente diverse sui social media, cfr. Bakshy *et al*. (2015).

un'operazione opportuna, che può svolgere una funzione correttiva di eventuali errori epistemici da parte dei soggetti coinvolti. Se il modo in cui i ricettori e i propagatori delle *fake news* svolgono la loro attività risultasse essere epistemicamente difettoso, potrebbe essere ragionevole richiedere loro di modificare la loro condotta.

Alla domanda relativa ai (de)meriti epistemici della ricezione di *fake news* esistono, al momento, tre principali risposte nella letteratura epistemologica. Secondo la risposta offerta da Rini (2017), i ricettori di *fake news* agiscono generalmente in maniera epistemicamente *virtuosa*. Rini non ritiene che la diffusione capillare di *fake news* sia una buona cosa; ritiene, però, che il problema non debba essere collocato al livello dell'individuo, bensì al livello sistemico. La risposta offerta da Nguyen (2018) condivide l'idea che il problema delle *fake news* sia legato alla struttura dei social network, ma considera i ricettori delle *fake news* come non biasimevoli piuttosto che virtuosi dal punto di vista epistemico (*blameless*). La non biasimevolezza epistemica è, in generale, condizione necessaria per la virtù epistemica o per il possesso di altri valori epistemici quali la giustificazione. Pertanto, accettare la prima risposta impegna ad accettare anche la seconda risposta – di fatto, Rini e, probabilmente Nguyen, accettano entrambe. Secondo una terza risposta, proposta da Cassam (2016, 2019), i ricettori delle *fake news* sono invece difettosi dal punto di vista epistemico e gestiscono le loro credenze in maniera epistemicamente viziosa. I sostenitori di quest'ultima risposta credono che il problema delle *fake news* possa essere risolto promuovendo una riforma della condotta epistemica a livello individuale.

Le risposte offerte da Nguyen e Rini presuppongono che il consumo di *fake news* avvenga in ambienti peculiari, sebbene i due filosofi offrano ragioni diverse a supporto della propria concezione. Secondo Rini, la ricezione delle *fake news* sarebbe il frutto di un passaggio di testimonianza particolare, in cui un propagatore solitamente condivide una *fake news* su un social network e il ricettore forma una credenza nel contenuto della news per il fatto di leggerla nella propria bacheca o in quella di un suo contatto. Nella concezione di Rini, un elemento fondamentale delle dinamiche relative al consumo di news sui social media è che i propagatori di news che più facilmente compaiono nella propria bacheca appartengono in genere al proprio network di amici e contatti. Questi amici e contatti, come abbiamo già visto, sono normalmente selezionati in virtù della comune adesione a valori morali e politici fondamentali. Ciò giustificherebbe, secondo Rini, la supposizione che i contenuti da loro condivisi – almeno quando tali contenuti hanno a che fare con l'ambito dei valori – siano degni di essere accettati come veri. In particolare, dato il modo in cui sono stati selezionati, sarebbe razionale presumere che, dal punto di vista dell'utente che ne condivide la prospettiva, i nostri contatti tendano a rispondere correttamente alle questioni che coinvolgono valutazioni normative (2017, E-51). Sarebbe

appunto questa considerazione a spiegare perché, nella prospettiva di Rini, sia ragionevole considerare epistemicamente virtuosa la pratica di accettazione acritica delle news che coinvolgono valutazioni normative condivise dai propri contatti sui social media[18]. La virtù, si noti, non sarebbe da attribuire all'utente soltanto nel caso in cui l'informazione che accetta come vera è affidabile, bensì anche nelle circostanze in cui questi è vittima di *fake news*, se il contenuto dell'informazione coinvolge valori morali o politici.

Due casi specifici consentono di chiarire meglio la posizione di Rini: sebbene la diagnosi da lei offerta possa funzionare nel primo caso, riteniamo che si sbagli nel secondo. La prima tipologia di news che, secondo Rini, sarebbe ragionevole accettare in maniera "partigiana" se proveniente dai nostri contatti sui social media riguarda considerazioni puramente normative, ovvero affermazioni che rendono conto in maniera esplicita dell'appoggio che il propagatore dà a qualche insieme di valori fondamentali – si pensi, ad esempio, a news riguardanti la liceità morale di una determinata decisione politica o la sua immoralità. Non è difficile schierarsi dalla parte di Rini a questo proposito: nella misura in cui abbiamo ragione di credere che i nostri contatti condividano i nostri valori, *ceteris paribus* siamo giustificati a fidarci delle loro considerazioni in campo normativo più di quanto siamo giustificati a fidarci nelle considerazioni normative offerte da chi non appartiene al nostro network partigiano. Si noti che questo argomento non esclude la possibilità che uno sia ingannevolmente condotto a credere in una *fake news* se un membro della stessa comunità ne condivide uno. Al contrario, l'argomento spiega perché potrebbe non esserci nulla di sbagliato da un punto di vista epistemico se il destinatario accettasse come vera una *fake news* riguardante considerazioni normative propagata da un membro della stessa comunità.

Tuttavia, è raro che le *fake news* abbiano la forma di un vero e proprio giudizio di valore su particolari decisioni socialmente, moralmente o politicamente rilevanti. È alquanto più probabile che una *fake news* riguardi quelle che Rini chiama "considerazioni normativamente rilevanti" (*normatively relevant claims*), ovvero affermazioni *descrittive* che hanno una chiara rilevanza per qualche questione normativa. Considerazioni di questo genere comprendono, ad esempio, storie riguardanti le azioni di un politico, le conseguenze più probabili di un determinato programma politico, ecc. Rini sostiene che, anche in questo caso, è epistemicamente virtuoso fidarsi partigianamente dei membri del gruppo a cui si appartiene.

La posizione di Rini in questo secondo caso è difficilmente sostenibile. Supponiamo che uno dei nostri contatti condivida un articolo secondo cui P è vera. Quando P è normativamente rilevante, secondo Rini, il fatto che il nostro contatto abbia deciso di condividere l'articolo

[18] Rini parla a questo proposito della virtù della *partigianeria epistemica* (2017).

costituisce una ragione epistemica per accettare P come vera, in quanto tale decisione è di fatto una decisione *normativa* che rispecchia la valutazione da parte del nostro contatto circa l'importanza della questione relativa alla verità di P (2017, E-52). Condividiamo l'analisi secondo cui il fatto che l'articolo sia stato condiviso da un nostro contatto renda ragionevole aspettarsi che esso sarà importante anche per noi. Tuttavia, la questione in gioco non è se sia razionale impiegare il nostro tempo a leggere l'articolo; la questione è piuttosto se sia razionale accettare ciò che l'articolo sostiene, ovvero P. Il fatto che condividiamo una scala di valori con il nostro contatto ci consente di rispondere affermativamente alla prima domanda, ma sembra non aver niente da dire in risposta alla seconda. Infatti, se l'importanza che attribuiamo all'articolo dipende dalla scala di valori che adottiamo, e dal grado del loro accordo con i valori dei nostri contatti, la stessa cosa non vale per la sua accuratezza, che dipende dal grado di accordo del contenuto dell'articolo con la realtà dei fatti.

Qualcuno potrebbe obiettare che l'essere amici di qualcuno ci autorizza razionalmente a presumere che questi sia moralmente integro. Se questo è vero, si potrebbe sostenere che essere amici di qualcuno ci autorizzi anche ad accettarne acriticamente la testimonianza. L'argomento, in particolare, potrebbe essere il seguente[19]. Supponiamo che l'amico A dica che P. Se P è falso, A ha fatto qualcosa di moralmente criticabile nell'affermare che P. La mia amicizia con A, tuttavia, mi autorizza razionalmente ad escludere che A abbia fatto qualcosa di moralmente criticabile; e quindi, *a fortiori*, mi autorizza a pensare che non possa aver fatto niente di male nell'affermare P. Ne segue, per *modus tollens*, che affermando P, A ha affermato qualcosa di vero. Questa risposta all'obiezione che abbiamo mosso a Rini, seppur interessante, non sembra in grado di disinnescarne il potenziale. Dire qualcosa di falso non equivale a mentire; e se mentire è moralmente deprecabile, almeno *prima facie*, non è affatto chiaro che asserire il falso sia altrettanto deprecabile da un punto di vista morale. Ad esempio, il fatto che un amico condivida sulla propria bacheca un contenuto falso che egli del tutto in buona fede ritiene essere vero non sembra essere moralmente criticabile. Se asserire qualcosa di falso non è moralmente biasimevole, la presunzione razionale a favore dell'integrità morale dei nostri amici è del tutto irrilevante rispetto al fatto che essi abbiano asserito qualcosa di falso e, pertanto, non è in grado di generare alcuna ragione aggiuntiva per credere ciò che essi riportano, con buona pace del sostenitore della risposta in questione.

Come abbiamo già rilevato, Nguyen (2018) affronta la questione relativa alla responsabilità epistemica del consumo e della accettazione di *fake news* in relazione alla sua analisi degli ambienti che ne favoriscono la diffusione. Per Nguyen questi ambienti esibiscono le caratteristiche

[19] Ringraziamo un revisore anonimo per aver suggerito questa obiezione.

tipiche delle camere d'eco, che impediscono agli utenti il libero accesso all'informazione non solo omettendo o oscurando alcune particolari fonti, ma anche discreditandole attivamente. Questa analisi è fondamentale al fine di valutare la responsabilità epistemica che possiamo attribuire ai membri delle camere d'eco per il modo in cui gestiscono le loro credenze. Nguyen sostiene esplicitamente che, sebbene i membri di una camera d'eco conformino il proprio comportamento alla irrealistica distribuzione del credito epistemico imposto dalle meta-credenze su cui le camere d'eco si fondano, possono agire come farebbe un agente epistemicamente ragionevole (p. 15). Il fatto di essere in una certa misura intrappolati all'interno di una camera d'eco, quindi, non esclude necessariamente che i membri stiano agendo in modo virtuoso sul piano epistemico.

Per capire meglio la posizione di Nguyen e valutarne le conseguenze nel caso della ricezione di *fake news*, può essere utile considerare l'esempio seguente[20]. Supponiamo che Oliver sia cresciuto in una camera d'eco all'interno della quale il crollo delle torri gemelle è considerato un crimine commesso da qualche organizzazione governativa americana. I suoi genitori e le autorità epistemiche della comunità di cui fa parte hanno da sempre sostenuto questa tesi, esibendo evidenza in suo favore e fornendo ulteriore evidenza a favore delle altre credenze della camera d'eco. Tra le altre cose, queste credenze sostengono che la versione ufficiale del disastro delle torri gemelle è falsa, e che è stata creata appositamente, e diffusa su tutti gli organi di informazione, nel tentativo di tenere l'opinione pubblica all'oscuro delle reali cause del crollo delle torri. Immaginiamo che un sito internet X che Oliver considera una fonte affidabile[21] pubblichi una storia, chiamiamola P, che racconta ulteriori dettagli sul modo in cui le torri gemelle sarebbero state abbattute da esponenti dello stesso governo americano. Supponiamo poi che P soddisfi tutte le condizioni per essere una *fake news* offerte in § 2. Oliver riflette su come dovrebbe comportarsi rispetto alla storia che ha appena letto, non trova ragioni per dubitare della sua affidabilità, e si convince che P è vera. Come dovremmo valutare la sua condotta epistemica?

Innanzitutto, si noti che Oliver sembra aver acquisito le credenze centrali della camera d'eco in cui è intrappolato in maniera ragionevole, cioè fidandosi delle autorità epistemiche della sua comunità. Se questo è vero, non potendolo biasimare per il fatto di essere stato cresciuto all'interno di una camera d'eco, nemmeno possiamo biasimarlo per il fatto di essere arrivato a condividerne le credenze centrali. Possiamo anche immaginare che Oliver abbia esercitato svariate virtù epistemiche nell'arrivare a credere in P: ha cercato attivamente nuove fonti di

[20] L'esempio, nella fattispecie, è la versione riveduta da Nguyen di un caso proposto originariamente in Cassam (2014).

[21] Possiamo assumere per ipotesi che la credenza di Oliver circa l'affidabilità del sito X costituisca una delle meta-credenze che egli eredita dalla camera d'eco di cui fa parte.

informazione prima di imbattersi nella storia pubblicata dal sito X e ha condotto un'indagine accurata sulla credibilità di X in base alle meta-credenze che ha ereditato dalla camera d'eco. Infine, accettando P ha finito per formare una credenza perfettamente in linea con la valutazione dell'evidenza a sua disposizione. Secondo l'analisi di Nguyen, tutto ciò spiegherebbe perché Oliver non sia colpevole dal punto di vista epistemico: piuttosto, è la comunità epistemica a cui appartiene ad essere epistemicamente viziosa e questo spiegherebbe perché, nonostante la ragionevolezza della sua condotta, Oliver non riesca ad assicurarsi il raggiungimento di un bene epistemico[22].

A questo punto, qualcuno potrebbe sospettare che l'analisi di Nguyen sia perfettamente compatibile con la diagnosi della condotta dei ricettori di *fake news* offerta da Rini. Tuttavia, ad uno sguardo più attento, due dettagli importanti ci permettono di mostrare che la posizione di Nguyen sia più ragionevole di quella di Rini. In primo luogo, la tesi di Nguyen è più debole. Secondo il primo, un individuo può essere epistemicamente virtuoso *sebbene* sia il ricettore di una *fake news*, nella misura in cui questi si impegni nell'attività di ricerca di ragioni a supporto dell'informazione ricevuta esercitando le proprie virtù epistemiche. Al contrario, Rini afferma che un individuo può essere epistemicamente virtuoso *in quanto* ricettore di una *fake news*, nella misura in cui dimostra di possedere la virtù della partigianeria epistemica, attribuendo cioè maggiore credibilità ad un testimone quale il propagatore di una *fake news* in virtù del fatto che questi condivide gli stessi valori sul piano normativo. In secondo luogo, la tesi di Nguyen è più ristretta, nel senso che si limita a garantire la possibilità di attribuire virtù epistemiche ai ricettori di *fake news* che non sono colpevoli in quanto cresciuti all'interno di camere d'eco. Al contrario, Rini sostiene che la partigianeria epistemica sia una virtù a prescindere dalle caratteristiche socio-epistemiche della comunità a cui si appartiene.

Finora abbiamo offerto ragioni per rigettare la tesi forte secondo cui i ricettori di *fake news* che intrattengono valutazioni epistemicamente partigiane possono essere virtuosi e, al contempo, per accettare la tesi più debole secondo cui i ricettori di *fake news* cresciuti all'interno di camere d'eco possono essere non biasimevoli per il fatto di credere a determinate *fake news*, e addirittura virtuosi per il modo in cui adottano i metodi di acquisizione dell'evidenza che hanno appreso dalle autorità della propria comunità. Se tutti i ricettori di *fake news* fossero di default membri di camere d'eco, questa tesi debole sarebbe in grado di spiegare

[22] Questa valutazione della condotta epistemica di Oliver presuppone che essere epistemicamente virtuosi sia una caratteristica relativa all'attività mentale dell'individuo. È opportuno menzionare, per dovere di completezza, che in una prospettiva alternativa, di stampo esternalista, Oliver non sarebbe colpevole dal punto di vista epistemico ma, al contempo, non potrebbe agire in maniera virtuosa perché si trova ad agire in un ambiente epistemicamente vizioso.

il fenomeno nella sua completezza. Tuttavia, sembra plausibile che le *fake news* si propaghino anche all'interno di bolle epistemiche e di comunità strutturalmente meno viziose. Tali scenari possono essere analizzati adeguatamente facendo riferimento alle considerazioni di Cassam sulla responsabilità individuale dei ricettori delle *fake news*.

Consideriamo una versione modificata del caso di Oliver, in cui il protagonista non è un membro di una camera d'eco, bensì un individuo che ha semplicemente letto una *fake news* secondo cui il crollo delle torri gemelle sarebbe stato dovuto ad esplosivi inseriti nell'edificio da agenti governativi. Questa presunta notizia ha suscitato l'interesse di Oliver per l'ipotesi del complotto, e lo ha motivato a fare ulteriori ricerche. Le nuove indagini lo portano a imbattersi in nuove fonti – siti internet, forum on-line, libri – che univocamente supportano l'ipotesi complottistica. Alla luce dei nuovi dati in suo possesso, Oliver finisce per convincersi del fatto che la spiegazione ufficiale è stata creata per coprire le vere cause dell'evento, e che le torri gemelle sono state abbattute dallo stesso governo americano.

Secondo Cassam (2016), Oliver è epistemicamente biasimevole, in quanto non si è attenuto alle norme di un'indagine responsabile, che richiedono quantomeno una certa dose di sensibilità alle situazioni in cui si rischia di essere ingannati (p. 163). I difetti fondamentali della condotta epistemica di Oliver possono essere individuati facendo riferimento ai vizi intellettuali. Oliver è affetto da: chiusura mentale, poiché non tiene in dovuta considerazione alcuna evidenza contraria alla sua teoria; "disfunzione pregiudiziale" (*prejudicial dysfunction*, Fricker 2012, 340), poiché ripone erroneamente fiducia in altre persone attribuendo una eccessiva dose di credibilità a fonti che non la meritano e negando la dovuta dose di credibilità a fonti legittime; e mancanza di accuratezza, poiché non riconosce che la presunta evidenza a supporto della teoria del complotto non sopravvivrebbe ad un'indagine approfondita.

Si noti che l'insieme di vizi intellettuali che giocano un ruolo nell'epistemologia delle *fake news* è decisamente più ampio dell'insieme di elementi coinvolti nel caso di Oliver. Per esempio, un soggetto epistemico che si comporta nel modo descritto da Rini – attribuendo eccessiva credibilità ai membri del proprio gruppo in situazioni in cui sarebbe opportuno cercare evidenza fattuale – mostra di essere vittima di un vizio epistemico peculiare, che potremmo chiamare di *credulità partigiana*. Inoltre, la nostra analisi si è concentrata sugli atteggiamenti epistemici dei ricettori delle *fake news*, ma simili considerazioni potrebbero – e, di fatto, dovrebbero – essere sviluppate a proposito dei propagatori, che probabilmente possiedono altri vizi intellettuali[23].

[23] Un riferimento all'epistemologia delle virtù come strategia rilevante per rendere conto della diffusione di *fake news* è stato proposto recentemente da Heersmink (2018) e Smart (2018).

Si noti anche che svolgere indagini epistemiche in maniera responsabile – quindi prive di vizi intellettuali – non è soltanto fondamentale per la nostra fioritura intellettuale in quanto utenti dei social network, bensì anche per gli obiettivi epistemici delle nostre comunità. Come giustamente dimostra la letteratura sulla proliferazione delle *fake news*, le credenze che formiamo attraverso i social media e le nostre attività in quanto utenti hanno ottime probabilità di influenzare non solo le nostre valutazioni epistemiche, ma anche quelle degli altri membri delle nostre e di altre comunità (Garrett, Weeks e Neo, 2016; Vosoughi *et al*., 2018).

Qualcuno potrebbe sospettare che l'analisi che abbiamo offerto degli aspetti normativi della diffusione delle *fake news* lasci aperte due strade alternative, l'una che solleva i singoli ricettori dalla responsabilità di accettare e diffondere *fake news* e l'altra che li condanna per i loro atteggiamenti epistemicamente irresponsabili e viziosi. A ben vedere, però, le due opzioni sono compatibili perché riguardano due scenari differenti: uno in cui la diffusione di disinformazione attraverso un network è dovuta alle caratteristiche particolari della struttura sociale ed epistemica del network stesso (ad esempio, la presenza di camere d'eco); l'altro in cui i difetti epistemici dei singoli utenti contribuiscono in maniera decisiva alla pervasività delle *fake news*.

Allo stato attuale dei fatti, sembra ragionevole guardare con sospetto a qualsiasi tentativo di offrire un correttivo alla diffusione di *fake news* che non tenga in considerazione entrambi i fattori della responsabilità individuale e di quella del sistema all'interno del quale si diffondono[24]. Pertanto, ci auguriamo che la nostra analisi contribuisca a motivare il bisogno di soluzioni complessive ad un problema che minaccia di compromettere il progresso epistemico dell'intero villaggio globale.

5. Conclusione

Questo articolo ha offerto una ricostruzione dell'odierna letteratura filosofica sul tema delle *fake news* al fine di rispondere a tre questioni fondamentali per una epistemologia delle *fake news*. In primo luogo, abbiamo analizzato il dibattito relativo a che cosa le *fake news* siano e abbiamo articolato una concezione originale che evita alcune obiezioni sollevate contro altre teorie disponibili. In secondo luogo, abbiamo indagato le cause descrittive della proliferazione di *fake news*, soffermandoci, in particolare, sulle caratteristiche psicologiche degli utenti dei social media, sulle proprietà strutturali dei network e le loro dinamiche di condivisione, che favoriscono la propagazione di disinformazione. Infine,

[24] Si noti che le soluzioni proposte da Rini (2017) e Nguyen (2018) sembrano incorrere in questa limitazione. Da questo punto di vista, la nostra diagnosi si allinea con quella offerta in Lazer *et al*. (2018).

abbiamo discusso alcune questioni normative legate alla ricezione di *fake news*: in particolare, abbiamo offerto argomenti per rigettare qualsiasi visione che offra un verdetto generale di condanna o di assoluzione nei confronti dei ricettori di *fake news* sui social media. Abbiamo concluso difendendo la tesi che i singoli utenti possano essere ritenuti colpevoli per credere nelle *fake news*, a meno che non siano cresciuti all'interno di camere d'eco e non abbiano mai avuto l'opportunità di rendersi conto della parzialità epistemica delle loro comunità[25].

Riferimenti bibliografici

Allcott, H., Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *National Bureau of Economic Research*, Working Paper 23089.

Bakshy, E., Messing, S., Adamic, L.A. (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. *Science*, *348*, 6239, pp. 1130-1132.

Cassam, Q. (2014). *Self-Knowedge for Humans*. Oxford: Oxford University Press.

Cassam, Q. (2016). Vice Epistemology. *The Monist*, *99*, 2, pp. 159-180.

Cassam, Q. (2019). *Vices of the Mind*. Oxford: Oxford University Press.

Chisholm, R.M., Feehan, T.D. (1977). The Intent to Deceive. *Journal of Philosophy*, *74*, pp. 143-159.

Dentith, M.R.X. (2017). The Problem of Fake News. *Public Reason*, *8*, 1-2, pp. 65-79.

Frankfurt, H.G. (2005). *On Bullshit*. Princeton: Princeton University Press.

Garrett, R.K., Weeks, B.E., Neo, R.L. (2016). Driving a Wedge Between Evidence and Beliefs: How Online Ideological News Exposure Promotes Political Misperceptions. *Journal of Computer-Mediated Communication*, *21*, pp. 331-348.

Gelfert, A. (2018). Fake News: A Definition. *Informal Logic*, *38*, 1, pp. 84-117.

Gettier, E. (1963). Is Justified True Belief Knowledge?. *Analysis*, *23*, 6, pp. 121-123.

Fricker, M. (2012). Epistemic Injustice and Role for Virtue in the Politics of Knowing. In J. Greco e J. Turri (a cura di), *Virtue Epistemology: Contemporary Readings*. Cambridge (MA): MIT Press, pp. 329-350.

Habgood-Coote, J. (2018). Stop Talking About Fake News. *Inquiry*, doi: 10.1080/0020174X.2018.1508363.

Heersmink, R. (2018). A Virtue Epistemology of the Internet: Search Engines, Intellectual Virtues, and Education. *Social Epistemology*, *32*, 1, pp. 1-12.

Isenberg, A. (1973). Deontology and the Ethics of Lying. Aesthetics and Theory of Criticism: Selected Essays of Arnold Isenberg. Chicago: University of Chicago Press, pp. 245-264.

[25] Gli autori desiderano ringraziare i curatori di questo numero speciale per aver voluto includere questo contributo nel loro progetto e due revisori anonimi per i preziosi suggerimenti che hanno permesso di migliorare una versione precedente dell'articolo. Michel Croce ha lavorato a questo contributo grazie al supporto di *Horizon 2020 Research & Innovation Programme* (Grant n. 675415).

Jaster, R., Lanius, D. (2018). What is Fake News?. *Versus*, *2*, 127, pp. 207-227.
Klein, D., Wueller, J. (2017). Fake News: A Legal Perspective. *Journal of Internet Law*, *20*, 10, pp. 5-13.
Lackey, J. (2013). Lies and Deception: An Unhappy Divorce. *Analysis*, *73*, pp. 236-248.
Lazer, D.M., Baum, M.A., Benkler, Y., Berinski, A.J., Greenhill, K.M., Menczer, F., ... Schudson, M. (2018). The Science of Fake News. *Science*, *359*, 6380, pp. 1094-1096.
Levy, N. (2017). The Bad News about Fake News. *Social Epistemology Review and Reply Collective*, *6*, 8, pp. 20-36.
Mahon, J.E. (2008). Two Definitions of Lying. *International Journal of Applied Philosophy*, *22*, pp. 211-230.
Mukerji, N.S. (2018). What is Fake News?. *Ergo: An Open Access Journal of Philosophy*, *5*, pp. 923-946.
Nichols, T. (2018) *La conoscenza e i suoi nemici*. Roma: Luiss University Press.
Nguyen C.T. (2018). Echo Chambers and Epistemic Bubbles. *Episteme*, doi: 10.1017/epi.2018.32.
Primoratz, I. (1984). Lying and the "Methods of Ethics". *International Studies in Philosophy*, *16*, pp. 35-57.
Qiu, X., Oliveira, D.F., Shirazi, A.S., Flammini, A., and Menczer, F. (2017). Limited Individual Attention and Online Virality of Low-Quality Information. *Nature Human Behaviour*, *1*, 7, p. 0132.
Qiu, X., Oliveira, D.F., Shirazi, A.S., Flammini, A., Menczer, F. (2019). Retraction Note: Limited Individual Attention and Online Virality of Low-Quality Information. *Nature Human Behaviour*, doi: 10.1038/s41562-018-0507-0.
Rini, R. (2017). Fake News and Partisan Epistemology. *Kennedy Institute of Ethics Journal 27*, S2, pp. E43E64.
Riva G. (2018). *Fake News*. Bologna: Il Mulino.
Smart, P. (2018). (Fake?) News Alert: Intellectual Virtues Required for Online Knowledge! *Social Epistemology Review and Reply Collective*, *7*, 2, pp. 45-55.
Sunstein, C.R. (2017). *#republic. La democrazia nell'era dei social media*. Bologna: Il Mulino.
Tappin, B.M., van der Leer, L., McKay, R.T. (2017). The Heart Trumps the Head: Desirability Bias in Political Belief Revision. *Journal of Experimental Psychology: General*, *146*, 8, p. 1143.
Vosoughi, S., Roy, D., Aral, S. (2018). The Spread of True and False News Online. *Science*, *359*, 6380, pp. 1146-1151.
Williams, B. (2002). *Truth and Truthfulness: An Essay in Genealogy*. Princeton: Princeton University Press.

Fake News Epistemology

This paper addresses the proliferation of fake new from a philosophical – mainly epistemological – point of view. We devote special attention to three questions: how to define fake news, which mechanisms cause the proliferation of fake news on social media; who is to be blamed epistemically in the process through which fake news are generated, published and distributed. Starting from the extant

literature on the topic, we endeavor to: offer a novel definition of fake news immune from the main difficulties which afflict alternative accounts available in the literature (§ 2); describe the most salient causal factors underwriting the phenomenon of fake news (§ 3); offer a novel account of the normative dimension of this phenomenon (§ 4).

Keywords: fake news, epistemology, epistemic responsibility, epistemic vices and virtues.

Tommaso Piazza, Università di Pavia, Dipartimento di Studi Umanistici, Università di Pavia, Piazza Botta 6, 27100, Pavia, tommaso.piazza@unipv.it; LanCog, Centro de Filosofia, University of Lisbon, Alameda da Universidade, 1600-214, Lisbon, Portugal

Michel Croce, University of Edinburgh, Department of Philosophy, Dugald Stewart Building, 3 Charles Street, Edinburgh, UK, EH8 9AD, michel.croce@ed.ac.uk; LanCog, Centro de Filosofia, University of Lisbon, Alameda da Universidade, 1600-214, Lisbon, Portugal

TERESA NUMERICO

SOCIAL NETWORK E ALGORITMI DI *MACHINE LEARNING*: PROBLEMI COGNITIVI E PROPAGAZIONE DEI PREGIUDIZI

1. Introduzione

L'obiettivo dell'articolo è un'analisi epistemologica critica dei possibili effetti discriminatori e pregiudiziali della diffusione di tecniche di interpretazione algoritmica dei comportamenti umani, basate su metodi di apprendimento non supervisionato (*machine learning*) addestrati su basi di dati prodotti dagli utenti dei *social network*.

La generale questione epistemologica sollevata da questi algoritmi verrà introdotta da una discussione della performance di due tra i principali algoritmi per il trattamento dei testi, GloVe e Word2vec, che rappresentano i metodi attualmente più avanzati nell'ambito della comprensione testuale e adottano la tecnica del *word embedding*. Si può dimostrare che tali algoritmi, addestrati su basi di dati testuali di pubblico dominio, tendono ad associare le parole replicando stereotipi di genere e pregiudizi etnici, poiché inferiscono i collegamenti tra termini secondo la probabilità della loro distanza nei *training set*. Il modello di significato da cui sono ispirati li conduce a restituire la rappresentazione stereotipata e pregiudiziale di categorie sociali, di genere o etniche che sono intessute nella base dati su cui sono stati addestrati (Caliskan, Bryson e Narayanan, 2017; Bolukbasi *et al*., 2016).

La discussione critica sull'uso di questi strumenti nei *social network* non si riferisce alla validità della semantica distributiva, un affermato settore della ricerca linguistica (Lenci, 2008) o alla *latent semantic analysis* (Landauer *et al*., 1998). Si tratta piuttosto di un caso esemplare di come, dal punto di vista epistemologico, gli algoritmi addestrati su basi dati umane costituiscano delle rappresentazioni categoriali che replicano, anche con maggiore rigidità, le credenze e i pregiudizi degli umani, proprio in conseguenza della correttezza dell'ipotesi del valore contestuale del significato.

Il caso degli algoritmi per il trattamento dei contenuti testuali è solo un esempio dell'attività più generale condotta nei *social network* al fine di profilare le persone, categorizzarle per anticiparne le azioni o per suggerire loro preferenze, sulla base di comportamenti esibiti in passato o sull'appartenenza a gruppi, cluster o categorie di segmentazione della popolazione.

L'utilizzo degli algoritmi di *machine learning* nei *social network* si avvale di un grande apparato tecnico messo a disposizione da alcune delle principali aziende Internet, ma soprattutto si basa su due premesse equivoche. La prima è che la quantità di dati permetta all'uso degli strumenti algoritmici di agire in una situazione di abbondanza nella quale l'effetto distorsivo di eventuali errori sarebbe riassorbito dall'ampiezza del campione e non intaccherebbe la loro capacità previsionale. La seconda premessa ipotizza che i sistemi algoritmici interpretino in maniera più efficiente e precisa degli esseri umani il significato dei dati disponibili sugli utenti e quindi ne catturino il valore cognitivo ai fini di una categorizzazione affidabile, univoca e capace di previsione. Nessuna di queste premesse risulta essere dimostrata, ma solo assunta come punto di partenza dagli attori principali della scienza orientata ai dati (*data-driven*).

L'articolo suggerisce che non ci siano prove definitive per sostenere la tesi che le inferenze tratte dall'analisi di una grande quantità di dati, basandosi su algoritmi di apprendimento, garantisca di per sé un maggiore equilibrio nei giudizi rispetto a quelle effettuate dagli esseri umani senza supporto delle macchine. Inoltre, la presenza di dati sporchi o non controllati, la rigidità della capacità inferenziale degli algoritmi di apprendimento che si addestrano sui dati e l'orientamento utilitario del design sperimentale dei metodi adottati potrebbero produrre esiti peggiorativi e socialmente pericolosi sulla capacità interpretativa, modellistica e previsionale delle scienze sociali.

In alcuni casi, come nell'esempio degli algoritmi per il trattamento dei testi, dove è possibile testare le procedure che stabiliscono associazioni tra parole, o nel caso di Google Flu Trends[1] (Lazer *et al*., 2014), dove era possibile controllare le previsioni del picco di influenza con degli *small data* sullo stesso fenomeno, si può dimostrare l'inaffidabilità degli algoritmi, soprattutto analizzando gli output delle conclusioni, valutandone l'efficacia anticipatoria o la capacità di spiegazione. Ma più spesso, a causa della segretezza dei metodi adottati, e dell'eccesso di informazioni disponibili, risulta impossibile esercitare un controllo (Pasquale, 2016).

Tali procedure di calcolo producono spesso problemi epistemologici e conseguenze politiche difficili da affrontare, soprattutto quando si riferiscono a conoscenze e modelli che riguardano la categorizzazione delle persone, oltre a sollevare questioni di carattere etico e sociale, come è evidente per esempio nel caso ormai famoso di Cambridge Analytica (Sumpter, 2018, cap. 5).

[1] Google Flu Trends era un sistema di Google per monitorare l'andamento dell'influenza negli Stati Uniti e in molti altri paesi controllando la diffusione nelle ricerche di alcuni termini considerati cruciali per l'autovalutazione della salute da parte degli utenti. Il sistema non è più in funzione dopo aver fallito le previsioni.

Il suggerimento che consegue dalla presente analisi epistemologica dei metodi algoritmici per la profilazione degli utenti dei *social network* è che sia necessaria una migliore comprensione di regole e criteri progettati per implementare questi dispositivi al fine di controllare la *fairness* e l'affidabilità dei metodi usati. Pertanto è importante che questi non siano protetti dal segreto, perché non è sufficiente valutare gli algoritmi solo dai loro output, ma è necessario organizzare un sistema analitico che ne dimostri l'efficienza e l'affidabilità, comparandola con decisioni prese dalle persone. Sarebbe inoltre auspicabile che fosse possibile regolamentare quali siano gli ambiti in cui è sensato fare affidamento su procedure di calcolo di natura probabilistica, senza incorrere in potenziali effetti pregiudiziali per chi ne sia oggetto.

2. La datificazione e le sue conseguenze epistemologiche

Con l'introduzione della digitalizzazione assistiamo a una grande disponibilità di dati relativi ai comportamenti umani che sono accessibili alle piattaforme che offrono un'ampia disponibilità di spazi per la conservazione di contenuti generati dall'utente. La prospettiva degli studi sociali che vanno dalla sociologia al marketing sembra quindi essere riorganizzata dall'analisi di tutte queste fonti che si presentano già in forma datificata e che non vengono considerate come un campione della popolazione o una sezione della testualità, ma si propongono di ricostruire l'intero ambito dei comportamenti sociali, non solo una loro semplice rappresentazione.

Tale disponibilità, quindi, promette un cambiamento radicale dei metodi di studio e della capacità di analisi e previsione dei contesti sociali che precedentemente appartenevano a una cultura scientifica diversa da quella strettamente definita dalla quantificazione e misurazione dell'oggetto di ricerca (Mayer-Schönberger e Cukier, 2013; Kitchin, 2014; Nielsen, 2012).

Sabina Leonelli (2018) solleva una serie di problemi epistemologici intorno alla costruzione dei dati e alla loro portabilità. L'autrice sostiene la tesi che i dati si possano considerare come entità relazionali tra il mondo reale e chi lo studia e, quindi, come elementi di una mediazione. La definizione delle loro caratteristiche è sempre frutto di una negoziazione che deve essere tenuta in conto quando si vogliano mettere a valore i dati per la costruzione di una conoscenza affidabile, sia essa delle scienze biologiche o sociali. Per questo è inadeguato pensare ai dati come a oggetti puri che debbano solo essere sapientemente manipolati dalle regole di raffinati algoritmi che ne sappiano organizzare il senso, indipendentemente da dove sono stati raccolti, dalle modalità e dagli obiettivi della loro acquisizione.

Come si costruisce il rapporto tra modelli e dati, cioè tra i dati, e come questi vengono correlati per essere interpretati? Un metodo che potrebbe aiutarci in questo lavoro è quello proposto da Longo (2009), cioè una critica della ragione informatica. Osservare la struttura storica e genealogica della nascita e dello sviluppo del computer, e delle relative tecniche di programmazione, ci supporta nel mantenere uno sguardo critico sulla dimensione epistemologica di queste tecniche (Numerico, 2017; 2019). Calude e Longo hanno, inoltre, ottenuto un notevole risultato tecnico in merito alle correlazioni spurie. Hanno dimostrato che più aumenta la base dati, più si ampliano le evidenze di correlazioni casuali che non sono generate da una reciprocità nel comportamento delle serie di dati delle variabili che apparentemente sono correlate:

> [...] we show that this 'philosophy'[Big Data philosophy against science] is wrong. For example, we prove that very large databases have to contain arbitrary correlations. These correlations appear only due to the size, not the nature, of data. They can be found in 'randomly' generated, large enough databases, which – as we will prove – implies that most correlations are spurious. Too much information tends to behave like very little information (Calude e Longo, 2017, 595).

Il risultato di Calude e Longo manda in frantumi la tesi secondo la quale all'aumento dei dati corrisponde una neutralizzazione del rumore. Si dimostra invece che, aumentando i dati, il potenziale rumore rappresentato da serie correlate solo casualmente è destinato a esplodere indefinitamente, rendendo quindi difficile distinguere le correlazioni dotate di senso da quelle che non lo sono. Questo è tanto più valido quando si riferisce alle scienze sociali e umane.

Già Wiener aveva mostrato il suo scetticismo per il progetto di applicare le tecniche della cibernetica in questo ambito. Affermava, infatti:

> for a good statistic of society, we need long runs *under essentially constant conditions*. [...] Thus the human sciences are very poor tersting-grounds for a new mathematical techniques. [...] Moreover, in the absence of reasonably safe routine numerical techniques, the element of the judgement of the expert in determining the estimates to be made of sociological, antropological, economic quantities is so great that it is no field for a newcomer who has not yet had the bulk of experience which goes to make up the expert" (1948/1961, 25).

Wiener, quindi, intravedeva già il rischio che il giudizio soggettivo di un non esperto fosse indebitamente incluso nel fornire le stime di valutazione dalle quali trarre conclusioni statistiche prive di fondamento. Sottolineava, inoltre, come la mancanza di stabilità nelle osservazioni rendesse i dati troppo sporchi per trarre delle previsioni affidabili (sulla questione degli algoritmi come attività interpretativa, cfr. § 6). Nel libro *the human use of human beings* (1950/1954), Wiener rincarava la dose

affermando che il computer elettronico non fosse in grado di catturare la probabilità che caratterizzava la complessità umana, e che il rischio di dominio da parte della macchina sull'uomo fosse confinato a una sola possibilità. "The dominance of the machine presupposes a society in the last stages of increasing entropy, where probability is negligible and where the statistical differences among individuals are nil" (Wiener, 1950-1954, 181). Il dominio della macchina presupponeva, cioè, che l'umanità fosse diventata completamente prevedibile senza manifestare la variabilità e la molteplicità che le sono proprie.

Nel prossimo paragrafo tratterò alcuni risultati che dimostrano come i metodi per il *pattern recognition* che utilizzano alcuni algoritmi di *machine learning* applicati ai testi scritti possono riprodurre pregiudizi semantici e cognitivi simili a quelli degli umani che hanno prodotto le basi dati linguistiche di addestramento. Tale risultato è presentato come un esempio del più generale problema dell'utilizzo degli algoritmi sulle basi dati dei social network per fare previsioni che riguardino la valutazione delle persone.

3. Algoritmi per il Natural Language Processing e potenziali pregiudizi

Negli ultimi anni gli strumenti per il *Natural Language Processing* (NLP) hanno fatto un salto di qualità. In particolare la diffusione di Word2vec inventato da un gruppo di ricercatori che all'epoca lavoravano a Google (Mikolov *et al*., 2013a; 2013b; 2013c) è stato una specie di rivoluzione nell'ambito del *parsing* del linguaggio e della possibilità di 'comprendere' i testi.

Nel caso di Word2vec c'era anche un'analogia semplice e immediata proposta come elemento di affidabilità del progetto: "man is to woman as king is to x. It is impressive that one can just download Word2vec and discover that x is queen" (Church, 2017, 156). Il carattere intuitivo del rapporto faceva passare in secondo piano altri problemi dell'algoritmo che poi saranno riscontrati, come la difficoltà a riconoscere analogie nelle quali la relazione tra concetti risulti più complessa rispetto a quella dell'inferenza pilota, e la tendenza ad avere una risposta pregiudiziale in alcuni contesti, come vedremo fra poco.

Lo *Stanford Natural Language Processing Group* guidato da Jeffrey Pennington ha trovato un metodo per insegnare a un algoritmo a imparare il significato delle parole e le loro relazioni attraverso la lettura di un gran numero di pagine web, migliorando alcune delle prestazioni dell'algoritmo Word2vec, che rappresentava la punta più avanzata della ricerca. Il dataset di *training* dell'algoritmo GioVe è composto dalle pagine di Wikipedia oltre che da *Annotated English Gigaword* (*Fifth edition*), un database di pagine di notizie provenienti da diverse fonti

informative relative ad agenzie di stampa e altri database di testi, come Common Crawl, che indicizza ogni mese circa tre miliardi di pagine web e rende i dati accessibili al pubblico dominio. Questo database di diversi corpora testuali contiene più di 10 milioni di documenti e più di 4 miliardi di parole, di cui vengono tokenizzate 400.000, considerate importanti per convogliare il significato.

Il criterio usato nell'algoritmo è la misurazione della distanza tra le parole come indice di probabilità della loro correlazione, e in particolare si misurano quanto spesso coppie di parole si presentano in frasi che contengono parole simili. Tuttavia il metodo di calcolo e l'organizzazione dei dati permettono di migliorare le prestazioni computazionali rispetto a Word2vec. L'inferenza pilota di GloVe è del tipo "Trump sta agli US come Merkel sta a?", e il team di sviluppo ha dimostrato che dopo l'addestramento, l'algoritmo era in grado di sciogliere la variabile inserendo Germania al posto della domanda.

Un altro gruppo di ricercatori che ha testato i risultati usando l'algoritmo già pre-addestrato dal team di sviluppo, ha mostrato che il sistema di *word embedding* di GloVe tende a codificare al proprio interno non solo stereotipi di genere ma anche altre conoscenze, come la viscerale piacevolezza dei fiori e la distribuzione per genere delle occupazioni (Caliskan, Bryson e Narayanan, 2017). Gli autori dello studio suggeriscono che questo ultimo risultato supporti l'ipotesi distributiva in linguistica, cioè il fatto che misurare statisticamente la probabilità della distanza tra le parole catturi molto di quello che intendiamo per significato (Sahlgren, 2008; Lenci, 2008). L'evidenza degli studi sociali sugli esseri umani mostra che, anche senza un'esplicita volontà di atteggiamenti discriminatori, questi tendono ad adottare diversi tipi di comportamenti linguistici che si prestano a riprodurre stereotipi su persone di diversa origine etnica (tipicamente gli afro-americani, o la popolazione di origine latina), oppure riservati alle donne o ad altre minoranze sociali (Greenwald *et al*., 1998). Il problema non è quindi la *distributional hypothesis* in linguistica, ma il corpus di addestramento che è prodotto da esseri umani che riversano nei corpora i loro impliciti *bias*, poi catturati dall'algoritmo. Alcuni studi di psicologia sperimentale mostrano un aumento della probabilità di interesse in CV identici relativi a possibili candidati al posto di lavoro, solo cambiando l'origine etnica del nome di battesimo dei candidati (Lavergne e Mullainathan, 2004; Bertrand *et al*., 2005). Secondo Gillsespie (2014), il modo in cui è costruito l'insieme di addestramento ha impatto su come l'algoritmo reagirà quando eserciterà le proprie conoscenze acquisite, ripresentando gli stereotipi coi quali è stato addestrato, anche con una dose di irrigidimento causata dal meccanismo inferenziale iscritto nella procedura.

Our work has implication for AI and machine learning because of the concern that these technologies may perpetuate cultural stereotypes. Our findings suggest

that if we build an intelligent system that learns enough about the properties of language to be able to understand and reproduce it, in the process it will also acquire historical cultural associations, some of which can be objectionable" (Caliskan *et al*., 2017, 185).

Il caso di GloVe non è isolato. Altri studi (Bolukbasi *et al*., 2016) mostrano che anche cambiando l'algoritmo il risultato non cambia. Il gruppo di ricercatori, infatti, ha testato le prestazioni di Word2vec, addestrato su una base dati più controllata come quella delle news, ottenendo risultati del tutto analoghi a quelli precedentemente illustrati.

> The primary embedding studied in this paper is the popular publicly-available word2vec [...] embedding trained on a corpus of Google News texts consisting of 3 million English words and terms into 300 dimensions, which we refer to here as the w2vNEWS. One might have hoped that the Google News embedding would exhibit little gender bias because many of its authors are professional journalists. We also analyze other publicly available embeddings trained via other algorithms and find similar biases (Bolukbasi *et al*., 2016, 3).

Tutti e due gli algoritmi di maggior successo nell'ambito della comprensione dei testi, molto usati nella traduzione automatica, nella ricerca sul web e in altri contesti manifestano, quindi, gli stessi problemi, cioè prediligono l'attribuzione di lavori intellettuali ai maschi, mentre associano attività tipicamente legate alla casa e alla gestione delle esigenze familiari e di cura al femminile.

La tesi del gruppo di Bolukbasi è che sia possibile creare dei meccanismi che consentano all'algoritmo usato di diminuire il livello di pregiudizio di questi strumenti. Tali metodi di *debiasing* (ridurre la quantità dell'incidenza dei discorsi stereotipici e pregiudiziali) sono presentati così: "The first step, called Identify gender subspace, is to identify a direction (or, more generally, a subspace) of the embedding that captures the bias. For the second step, we define two options: Neutralize and Equalize or Soften" (Bolukbasi *et al*., 2016, 11).

Queste azioni di identificazione prima e successivamente di intervento descritte come neutralizzazione o addolcimento delle differenze portano alla luce alcune difficoltà. Innanzi tutto come individuare lo spazio parziale delle parole che sono erroneamente interpretate e associate dall'algoritmo? Subentra una dimensione di indeterminatezza che riguarda la soggettività del programmatore e che presenta altrettante oscurità, se tale decisione viene presa e implementata senza esplicitare preventivamente criteri e metodi adottati. L'algoritmo prevale sull'essere umano per il suo carattere di precisione e affidabilità. Se i suoi risultati sono interpolati e maneggiati o corretti a mano dai ricercatori questo configura anche la possibilità di peggiorare le cose, o comunque di inserire nell'algoritmo l'arbitrio umano e allora dove sarebbe la sua presunta superiorità? Si

tratterebbe solo di scegliere tra umani che programmano la macchina, i cui pregiudizi sono accolti, e umani che subiscono le conseguenze della programmazione.

Inoltre il passaggio relativo alla decisione di una eventuale neutralizzazione della differenza riscontrata o a un suo addolcimento rende ancora più difficile valutare il risultato. Se non si neutralizza la differenza, in che senso si addolcisce? Tale soluzione potrebbe includere ancora un pregiudizio a questo punto però più pericoloso perché ufficialmente espunto proprio dai ricercatori che lo hanno 'debiased' e che decidono quale sia la percentuale giusta di differenza tra uomini e donne circa le attività che svolgono. E più in generale pensare di intervenire sui risultati dell'algoritmo ci fa mettere in discussione l'utilità di utilizzarlo.

Riportando un confronto con Joanna Bryson – coautrice dello studio (Caliskan *et al*., 2017) – David Sumpter (2018, 194) suggerisce che ci sarebbero due livelli di interazione con il linguaggio degli umani. Il primo sarebbe esclusivamente un "sistema per acquisire informazioni" e sarebbe subalterno a un altro sistema che funziona attraverso la memoria esplicita e, secondo Bryson, ci permette di "negoziare con gli altri individui e costruire una nuova realtà".

"Mathematical models of relationships between words, like Word2vec and GloVe, only capture the first level. These systems find relationships between words, but don't reflect how we reason and think about the world" (Sumpter, 2018, 194). Questo sarebbe il motivo, secondo Joanna Bryson, dell'incapacità di questi algoritmi di cogliere nel segno quando cercano di comprendere il linguaggio contando e misurando le distanze tra le parole e le loro somiglianze e dissimiglianze. Tuttavia Sumpter è convinto che, mentre gli algoritmi possano essere posti sotto controllo da parte degli esseri umani che sono in grado di applicare dei correttivi al loro funzionamento, come nel caso di Bolukbasi *et al*. (2016), gli esseri umani siano invece completamente fuori controllo quando misurano le associazioni immediate e incontrollate. Il famoso Implicit Association Test (IAT)[2] esibisce la presenza di una serie di gradienti impliciti che tendono ad associare concetti in maniera tacitamente governata dai pregiudizi sui gruppi sociali. La risposta sperimentale umana a questi test mostra che in media esistono implicite preferenze per i bianchi verso

[2] L'implicit-association test (IAT) è una misura adottata in psicologia sociale che serve a valutare la forza delle associazioni implicite delle persone tra le rappresentazioni mentali degli oggetti nella memoria. Si applica di solito per misurare il livello di accettazione implicita degli stereotipi valutando se persone associno in maniera inconscia ai nomi afroamericani degli stereotipi negativi razziali. Il test si presta a misurare diversi tipi di pregiudizi verso gruppi etnici, il genere, la sessualità, la religione, l'età ecc. È possibile eseguire il test online per autovalutarsi usando la pagina https://implicit.harvard.edu/implicit/takeatest.html. Il contesto del test è il progetto *Implicit social attitude* dell'università di Harvard che si propone di misurare il livello inconscio di pregiudizio sociali delle persone.

i neri, per i giovani verso gli anziani, oltre ad altre risposte frutto di stereotipi come associare i maschi alla scienza e alla carriera mentre le femmine alle arti liberali e alla famiglia (Nosek *et al*., 2002). Tuttavia sebbene questi stereotipi esistano sul piano tacito (a patto di affidarsi alla dinamica complessa dei test adottati), non vuol dire che poi le persone li seguano quando agiscono, o che si facciano poi condizionare dal punto di vista del giudizio e della categorizzazione. Mentre pare evidente che, per come sono configurati, gli algoritmi che costruiscono somiglianze tra i vettori che misurano la probabilità della distanza tra le parole perseguono proprio la costruzione di categorie e giudizi pregiudiziali.

Inoltre le distanze scelte per valutare la correlazione sono troppo ridotte per contribuire efficacemente a catturare il senso di una parola nel contesto. Ci sono poi parole che restano sottorappresentate nei corpora e per le quali esistono pochi esempi, che rendono la probabilità delle associazioni a distanza meno affidabili. La normalizzazione statistica, eseguita dall'uso dei corpora e dai meccanismi di calcolo adottati, costruisce inevitabili appiattimenti e fraintendimenti dei termini. Inoltre i ricercatori stessi che hanno inventato GloVe dichiarano: "The model produces a word vector space with meaningful substructure, as evidenced by its state-of-the-art performance of 75% accuracy on the word analogy dataset" (Pennington *et al*. 2014, 1532). Sebbene l'errore di classificazione possa riguardare qualsiasi elemento, l'affidabilità dell'algoritmo resta limitata e potrebbe causare casi di falsi positivi in grado di peggiorarne la performance anche relativamente alla classificazione pregiudiziale dei concetti. GloVe è allo stato dell'arte l'algoritmo che funziona meglio in questo compito rispetto agli altri. Tale performance dell'algoritmo collegata con i corpora prodotti da esseri umani, pur funzionando egregiamente sotto il profilo della *distributional hypothesis* in linguistica, restituisce risultati che rispecchiano le relazioni che storicamente si danno tra le parole e può finire per rafforzare gli stereotipi che sono già iscritti nel linguaggio. Possiamo quindi affermare che il problema nell'uso di questi algoritmi si pone non tanto per la loro efficienza nell'analisi testuale, quanto nella potenziale natura discriminatoria dei corpora sui quali si addestrano e per gli obiettivi classificatori che in alcuni casi si propongono.

4. La fairness negli algoritmi: tra diritto e cognizione

I risultati dell'addestramento di algoritmi come GloVe sui corpora testuali disponibili rimandano a una riflessione più generale sul ruolo degli algoritmi nei *social network* e sulla loro capacità di supportare la previsione dei comportamenti degli esseri umani analizzando le loro tracce testuali e non lasciate online sulle piattaforme, che è l'oggetto principale del presente articolo. Questo campo di ricerche, in grande espansione,

riguarda gli studi sull'equità dei risultati sui quali si basa la presa di decisione di impostazione algoritmica (*fairness in machine learning*). In una recente sintesi della letteratura sull'argomento, Chouldechova e Roth (2018) passano in rassegna le diverse questioni che sono ormai sul tavolo della ricerca di *machine learning* dedicata a questi temi. Si può notare un aumento progressivo del numero degli articoli dedicati alla possibilità di costruire algoritmi in grado di funzionare preservando l'equità e l'assenza di valutazioni pregiudiziali[3]. Non ci sono, quindi, solo obiezioni che pongono la questione dell'efficienza dei metodi algoritmici sotto il profilo della segretezza delle loro pratiche (Pasquale, 2015) o sulla tendenza a esibire risultati che finiscono per danneggiare la rappresentazione delle minoranze etniche come nel caso di Noble (2018), ma anche lavori che si occupano di segnalare la possibilità di *bias* impliciti e involontari negli algoritmi di classificazione. In particolare possiamo considerare quelli dedicati al *pattern recognition* rispetto alla profilazione degli utenti e alle pratiche di *collective filtering* e *recommendation system* come quelli di Amazon, ma anche di Tripadvisor e di altre piattaforme di raccomandazione.

Il problema del rischio di decisioni pregiudiziali è riassunto molto dettagliatamente in un articolo sul concetto di *disparate impact* relativo ai big data (Barocas e Selbst, 2016). Il *disparate impact*[4] è una nozione giuridica della costituzione americana che segnala quando un sistema per la presa di decisione rischia di fornire risultati pregiudiziali relativamente a sottogruppi, o gruppi di popolazione che sono sottoposti a una condizione di fragilità che può essere accresciuta da stereotipi e giudizi superficiali. Tale rischio si manifesta anche in assenza di una esplicita volontà discriminatoria.

Advocates of algorithmic techniques like data mining argue that these techniques eliminate human biases from the decision-making process. But an algorithm is only as good as the data it works with. Data is frequently imperfect in ways that allow these algorithms to inherit the prejudices of prior decision makers. In other cases, data may simply reflect the widespread biases that persist in society at large. In still others, data mining can discover surprisingly useful regularities that are really just preexisting patterns of exclusion and inequality. Unthinking reliance on data mining can deny historically disadvantaged and vulnerable groups full participation in society. Worse still, because the resulting discrimination is almost always an unintentional emergent property of the algorithm's use rather than a conscious choice by its programmers, it can be unusually hard to identify the source of the problem or to explain it to a court (Barocas e Selbst, 2016, 671).

[3] Al tema della fairness e transparency degli algoritmi di machine e deep learning è dedicata da qualche tempo una conferenza interdisciplinare annuale: ACM Conference on Fairness, Accountability, and Transparency (ACM FAT*).

[4] Il concetto di disparate impact può essere tradotto con un termine giuridico italiano

Il problema che Barocas e Selbst pongono è la possibilità che questa disparità di trattamento non sia esplicitamente classificata e deliberatamente perseguita dai programmatori, ma che sia il frutto di una loro implicita adesione a regole e decisioni precedenti che finiscono per danneggiare le minoranze senza che ci sia un dolo volontario in tal senso. Secondo Binns (2017) la questione della *fairness* (equità, giustizia sociale) non può essere risolta a livello tecnico perché richiede una definizione di tipo politico. È possibile, infatti, avere diverse definizioni di cosa si debba intendere per *fairness* relativamente alla prospettiva di chi deve prendere le decisioni.

Sebbene il tema della possibile discriminazione delle minoranze abbia un'applicazione più generale di quella nei *social network*, tali piattaforme costituiscono il luogo privilegiato in cui raccogliere, categorizzare, valutare i comportamenti delle persone, le loro abitudini e soprattutto le loro relazioni – uno degli elementi chiave dei meccanismi di clusterizzazione alla base della possibile valutazione pregiudiziale e discriminatoria.

La potenziale influenza negativa esercitata dalle procedure automatiche di decisione è vietata anche nel regolamento approvato dal parlamento dell'Unione Europea nel 2016, entrato in vigore a maggio del 2018 in Europa, il General Data Protection Regulation (GDPR). Nell'articolo 22 del regolamento, dal titolo *Automated individual decision making, including profiling* si dichiara al comma 1:

> The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

Questo articolo, quindi, protegge i cittadini europei da trattamenti automatici che possano significativamente avere effetti su di loro, trattamenti ai quali i cittadini potrebbero essere sottoposti senza un esplicito consenso da parte del soggetto di dati. Il soggetto dei dati è colui che è portatore dei dati e al quale i dati, su cui si basa il trattamento, appartengono. Tale tutela vale anche nel caso in cui il trattamento automatico non abbia l'esplicita intenzione di danneggiare o discriminare il soggetto. Inoltre il Regolamento Europeo (2016) chiarisce e sancisce il diritto del soggetto a una spiegazione "right to explanation" rispetto a come sono stati ottenuti e poi interpretati i suoi dati.

Secondo Goodman e Flaxman (2016):

> Articles 13 and 14 [GDPR] state that, when profiling takes place, a data subject has the right to "meaningful information about the logic involved." This requirement prompts the question: what does it mean, and what is required, to explain an algorithm's decision?" (p. 6).

'disparità di trattamento'. Tuttavia la letteratura giuridica americana è particolarmente sensibile su questo tema a causa della lunga storia di segregazione raziale nel paese.

I presenti articoli del regolamento segnalano, quindi, la necessità di spiegare la logica implicata dalla decisione e come il modello funzioni, sia relativamente alla base dati sulla quale è stato addestrato, sia rispetto ai metodi di astrazione adottati per effettuare la categorizzazione. Secondo Burrell (2016), la parte più complicata del fornire una spiegazione del funzionamento degli algoritmi è rappresentata da un "mismatch between the mathematical optimization in high-dimensionality characteristic of machine learning and the demands of human-scale reasoning and styles of interpretation". Il problema sarebbe quindi che i meccanismi utilizzati dall'algoritmo non sono facilmente comprensibili per gli esseri umani, pur avendo effetto su di loro.

5. Esempi di algoritmi discriminatori sui dati raccolti dai social network

È evidente che nel caso della presa di decisione sull'attribuzione o meno di un prestito o sull'accesso o meno all'ambito lavorativo l'algoritmo deve rendere comprensibili i suoi criteri di giudizio. Ma anche per i dati raccolti nei *social network* che promuovono la profilazione e la valutazione dei contenuti generati dagli utenti al fine di una loro categorizzazione devono avvenire secondo regole certe. Gli esempi sulla capacità discriminatoria dei meccanismi di funzionamento dei *social network* sono, infatti, già molto eloquenti. Ci sono prove che dimostrano che è possibile discriminare sull'origine etnica per chi vuole affittare la casa (Hynes, 2016; Silva e Kenney, 2018), o sull'età per chi sta mandando un'inserzione con un'offerta di lavoro e sembra anche che Facebook abbia recentemente aiutato aziende in cerca di lavoratori a discriminare sulla base del sesso di coloro che avrebbero visualizzato l'annuncio (Scheiber, 2018).

Si può dimostrare che molti algoritmi adottati in campi differenti non riescono a rendere conto dei risultati delle proprie capacità di previsione e di organizzazione della base dati. Mi concentrerò soprattutto, ma non esclusivamente, su esempi relativi al trattamento testuale e a come tali dati siano utilizzati per fornire raccomandazioni alla presa di decisione, suggerimenti di categorizzazione, spinte a orientare comportamenti e opinioni.

Un esempio è rappresentato dal meccanismo delle *query suggestion*, quella lista di espressioni che Google suggerisce quando immettiamo nella finestra di dialogo del motore alcune parole o *keyword*. Il caso di questo strumento così utile a raffinare e precisare i termini della ricerca è emblematico della delicatezza della questione. I suggerimenti hanno lo scopo di anticipare e indovinare il desiderio o l'intenzione sottesa dietro la formulazione della *query* di ricerca per aiutare l'utente a centrare me-

glio l'argomento sul quale vuole ottenere informazioni. I suggerimenti dipendono anche dal modo in cui le richieste sono state formulate in passato dagli altri utenti e da come si è comportato ogni utente nelle ricerche passate. Tuttavia la responsabilità di come vengono presentati i consigli è tutta dei provider di informazioni, in particolare dei motori di ricerca. Dall'analisi di Miller e Record (2017) si evince che:

> Search providers have the bulk of required features and technological possibilities to identify and prevent harms. They therefore bear the bulk of responsibility to mitigate them and should eliminate autosuggestions that result from organized attacks to game the system, those that perpetuate damaging stereotypes about socially disadvantaged groups, and those that associate negative characteristics with specific individuals (Miller e Record, 2017, 1959).

La responsabilità etica, ma soprattutto epistemica, circa le scelte dei suggerimenti, secondo i due studiosi, deve essere attribuita ai motori di ricerca, che hanno gli strumenti, le capacità e le competenze per fare scelte consapevoli, evitando di supportare e sostener stereotipi circa gruppi svantaggiati socialmente. Possiamo notare come in molti casi questa responsabilità non sia sostenuta o sia considerata scarsamente rilevante. Il problema nel caso degli algoritmi usati per la *query suggestion* è legato al controverso obiettivo del motore di ricerca. Posto che Google dichiara che la propria mission sia: "to organize the world's information and make it universally accessible and useful", il motore deve decidere se si tratta di rappresentare lo Zeitgeist o se invece il ruolo di *gatekeeper* dell'informazione abbia un vincolo epistemologico e gli imponga una responsabilità sociale maggiore. Si tratta di un'ambivalenza che il motore di ricerca non accenna a sciogliere (Vaidhyanathan, 2011). Google deve scegliere se i suoi algoritmi sono organizzati per favorire l'accesso all'informazione disponibile, controllando di non sostenere la diffusione di informazioni false, o se sono tarati sul massimizzare le richieste più popolari, che producono anche il ritorno pubblicitario più importante. Se il motore si limita a mettere in relazione domanda e offerta di informazione come se si trattasse di un qualsiasi mercato di commodity, come può essere considerato un agente informativamente affidabile, a cui tutti fanno riferimento?

La diffusione di contenuto stereotipato e pregiudiziale relativamente a identità etniche e di genere tramite Google è l'oggetto di *Algorithms of oppression* di Safiya Noble (2018). In questo libro, l'autrice affronta il tema della rappresentazione che risulta dai link proposti come prime risposte alle query che hanno per oggetto gruppi etnici femminili come "black girls" o "asian girls" o "ragazze italiane" o ancora "ragazze nigeriane". Se si utilizzano queste chiavi di ricerca, troviamo nei risultati del motore ancora il prosperare degli stereotipi e la rappresentazione del femminile soprattutto nella forma dell'oggetto del desiderio dello sguardo maschile. Google sembra scegliere di concentrare gli esiti

delle ricerche intorno alla richiesta che viene formulata dallo sguardo maschile desiderante, in particolare nelle immagini e nei video, nei tag per categorizzare le immagini e pure nei risultati generali del motore. È interessante osservare che dal momento della ricerca condotta da Noble (2018), in conseguenza delle pressioni del movimento *Black Lives Matter* e del cattivo ritorno di immagine causato dal dibattito sui risultati di ricerca discriminatori, i link per quella *query* sono stati modificati e ora includono principalmente informazioni corrette politicamente riguardo a quel gruppo etnico femminile. Si mostra quindi che l'opinione pubblica è in grado di intervenire e cambiare l'esito delle ricerche algoritmiche associate a una certa interrogazione e che la capacità di ottimizzazione dei risultati di ricerca è, per Google, sensibile all'approvazione sociale.

Proseguendo nella carrellata dei malfunzionamenti algoritmici basati sull'analisi testuale, non possiamo dimenticare il *chatbot* di Microsoft Tay che era stato sperimentato su Twitter per ingaggiare e intrattenere i giovani tra i 18 e i 24 anni e che dopo poche ore ha cominciato ad avere comportamenti razzisti e a esprimere posizioni estreme (Hunt, 2016). Il *chatbot*, silenziato dopo sole sedici ore di attività, era stato programmato per interagire imparando e addestrandosi seguendo i comportamenti linguistici nei quali si imbatteva. Il risultato piuttosto deludente e anche un po' spaventoso ha portato Microsoft a interrompere bruscamente l'esperimento (Lee, 2016). Insieme con la tendenza di Linkedin a correggere i nomi femminili, suggerendo a chi fa ricerca di usare lo stesso nome al maschile (Day, 2016), si tratta di casi nei quali il funzionamento degli algoritmi nel maneggiare la testualità pare ancora più carico di pregiudizi e stereotipi rispetto alle interazioni umane.

Infine segnalo la discussione circa il ruolo dei social network nell'accesso all'informazione e nell'evidente funzione di *gatekeeping* svolta dagli algoritmi di ricerca come Edgerank che presiede alla personalizzazione della visibilità delle informazioni nel News Feed. In un recente studio che analizza il funzionamento di Facebook durante le elezioni politiche italiane del 2018, si mostra che il meccanismo di filtraggio basato sulle preferenze pregresse degli utenti ha avuto un impatto rilevante sulla personalizzazione della visibilità delle pagine, introducendo quello che viene chiamato in letteratura *ranking bias* che pesa soprattutto sui primi posti nella visualizzazione (Hargreaves *et al*., 2018). Il risultato ottenuto dopo aver costruito un framework di misurazione riproducibile del News Feed è che:

> We were able to conclude that the algorithm tends to reinforce the orientation indicated by users about the pages they "like", by filtering posts and creating biases among the set of followed publishers. The effects of filtering are stronger on the topmost position where only a fraction of the set of publishers followed by the users was represented (Hargreaves *et al*., 2018, 7).

6. I problemi epistemici degli algoritmi di profilazione e filtraggio

Come abbiamo visto finora esistono alcune aree critiche che suggeriscono di mantenere alta la guardia epistemologica intorno al funzionamento degli algoritmi che si propongono di anticipare il futuro dei comportamenti umani. La prima è la questione storicamente nota come il problema dell'induzione introdotto da David Hume (*A treatise of human nature*, 1739). Come si fa partendo dalle passate esperienze a dimostrare che quelle stesse esperienze si ripresenteranno anche in futuro?

> Our foregoing method of reasoning will easily convince us, that there can be no *demonstrative* arguments to prove, that *those instances, of which we have had no experience, resemble those, of which we have had experience* (Hume, 1888, VI, 89).

Questo problema classico dell'epistemologia dell'inferenza induttiva aleggia irrisolto anche nella ricerca di generalizzazione effettuata dagli algoritmi. Anche loro utilizzano una base dati di *training* per estrarre caratteristiche, idee, categorie sulle quali eseguire delle inferenze che hanno come premesse quello che hanno "sperimentato" in passato, e come conclusione quello che prevedono accadrà in futuro. Il meccanismo dell'induzione, come sostiene Hume nel Settecento, non può essere considerato un'inferenza affidabile in qualsiasi condizione, perché non sempre ci sono garanzie che quello che è accaduto in passato verrà replicato nel futuro. Eppure il grande successo degli algoritmi di apprendimento è basato soprattutto sulla capacità di anticipare comportamenti, come nel caso di quelli che abbiamo visto relativi al NLP, i quali cercano di anticipare le relazioni tra le parole misurando la loro distanza probabilistica. La possibilità di previsione e anticipazione è centrale anche negli algoritmi di profilazione, che servono per modellare i comportamenti degli utenti a scopo pubblicitario nei social network. Gli obiettivi degli algoritmi possono essere molteplici, per esempio: anticipare gli interessi degli utenti nelle ricerche sul web, attribuire un punteggio alle persone per classificarle, come nel caso dei *credit scoring*, algoritmi per valutare l'affidabilità delle persone come pagatori di debiti (Citron e Pasquale, 2014), creare sistemi di raccomandazione basati sulle preferenze passate degli utenti (Muchnik *et al*., 2013), valutare il rischio di recrudescenza del condannato (*risk assessment algorithm* nel campo del *recidivism*) per supportare la presa di decisione del giudice intorno alla pena da infliggere a un condannato (Angwin *et al*., 2016).

Tutti questi metodi sono basati sull'assunzione che analizzare la lista dei dati di comportamenti pregressi di persone o oggetti simili permetterà di anticipare i comportamenti futuri di altre persone o altre cose. Il funzionamento degli algoritmi, quindi, si concentra su due principi fondamentali: quello che è accaduto in passato si ripeterà e quello che

fanno persone categorizzate come simili ad altre tenderà a estendersi a tutti gli altri membri dello stesso *cluster*. Le critiche a queste ipotesi nei diversi casi possono essere molteplici, a prescindere dal problema dell'induzione che vale anche per l'analoga tendenza umana a inferire da una serie di casi sperimentali sparsi la generalizzazione di una regolarità di funzionamento. Ci sono, però, alcuni problemi che sono specificamente relativi all'uso degli algoritmi.

In un recente risultato pubblicato su *Nature machine intelligence* (Ben-David *et al*., 2019) e commentato da Castelvecchi (2019) si mostra come il problema della capacità di apprendimento, o *learnability*, da parte di un algoritmo su una certa base dati si scontra con un grande dilemma della logica. Gli studiosi hanno dimostrato che il problema della *learnability* è equivalente all'indecidibilità del problema del continuo, dimostrata da Gödel (1940) e Cohen (1963,1964). In particolare si mostra che decidere come scegliere il sottoinsieme giusto dei dati che costituiranno la base di addestramento di un algoritmo è un problema indecidibile. Si tratta di un risultato che segnala i limiti della nostra conoscenza (Castelvecchi, 2019). Nell'articolo si mostra che comprimere adeguatamente l'insieme di conoscenze che sarebbero necessarie per l'addestramento nel più piccolo insieme di dati che sia adeguato all'addestramento è un problema che si dimostra equivalente a quello del continuo. Il concetto di *learnability* viene definito come l'abilità di fare previsioni su un insieme di dati, scegliendo un campione rappresentativo di questi dati. Il risultato consiste nel mostrare che solo se l'ipotesi del continuo di Cantor fosse vera sarebbe possibile fare l'estrapolazione da un numero finito di dati, ma se l'ipotesi fosse falsa, nessun sottoinsieme potrebbe mai essere sufficiente. La questione della *learnability*, perciò, resta nel limbo proprio come quella dell'ipotesi del continuo (Ben-David *et al*., 2019).

Anche lasciando da parte questo, pur notevole, risultato teorico, abbiamo la difficoltà che raccogliere dei dati per rappresentare in maniera digitalmente efficiente un quesito richiede la capacità di formalizzare la rappresentazione del problema. Questo vincolo permette di rendere conto solo della porzione che si presti a una misurabilità delle caratteristiche stesse dell'incognita o della situazione da analizzare.

Gillespie (2014) segnala come il modo in cui sono raccolti i dati orienta le capacità degli algoritmi di costruire un modello. Inoltre, per essere usati, i dati devono essere normalizzati e formalizzati per essere oggetto dell'analisi della macchina, il che significa presentarli come un sottoinsieme misurabile delle tracce dei fenomeni da analizzare. La possibilità dell'errore e del malfunzionamento della raccolta non è remota, anzi è molto probabile. Secondo Hito Steyerl (2019): "Dirty data thus become, so to speak, a reminder of reality in systems that are pegged to ideal models, averages, and Platonic assumptions, inspired by an ideal fictional world" (p. 7).

I dati sporchi, quindi, sono inevitabili. Il problema, segnalato da Steyerl, è la tendenza alla loro cancellazione quando questi non corrispondono al modello prestabilito. La propensione a uniformare la base dati per poterla trattare attraverso gli strumenti offerti dagli algoritmi di *machine learning* e *deep learning* è l'altro grande rischio di questo processo di normalizzazione. Gli algoritmi, infatti, oltre a spingere verso un'omologazione probabilistica della casistica presente nei dati, sono sempre più incontrollabili dagli stessi programmatori che li costruiscono. Non è possibile entrare nei dettagli della spiegazione che gli algoritmi forniscono dei dati. Eppure il *pattern recognition*, la funzione di reperire somiglianze e correlazioni tra i dati, riguarda sempre l'interpretazione, sebbene il meccanismo analitico promosso dagli algoritmi computazionali dichiari che il contesto e il valore dei dati è soltanto quantificato e probabilisticamente calcolato (Cramer, 2019):

> Yet it could be argued that data is always qualitative, even when its processing is quantitative: this is why algorithms and analytics discriminate, in the literal as well as in the broader sense of the word, whenever they recognize patterns (Cramer, 2019, 24-25).

Ogni volta che si costruisce una categorizzazione, che si può anche definire come una discriminazione o un giudizio, si esercita l'ermeneutica. Anche gli algoritmi esercitano un'interpretazione dei dati, ma tale attività è occultata dai sistemi di segretezza delle aziende, che li proteggono attraverso la proprietà intellettuale, considerandoli il loro vantaggio competitivo (Citron e Pasquale, 2014; Pasquale, 2015). I criteri usati classificare e normalizzare i dati sono spesso impliciti e non possono essere discussi e controllati. Se l'interpretazione è in azione, lo è cancellando ogni parvenza di soggettività.

Come segnala John MacCormick (2013), l'attività di *pattern recognition* – quella più importante usata nell'ambito dei *social network* per classificare comportamenti e anticiparli profilando le preferenze e le attività degli utenti delle piattaforme – è molto più familiare agli esseri umani di quanto non lo sia alle macchine. Questo non vale per molti altri algoritmi che vengono usati per esempio nell'ambito della crittografia, della compressione dei dati, nella firma digitale. La capacità di riconoscere regolarità anche laddove sono nascoste e implicite è una prerogativa umana.

> It turns out that computers are vastly inferior to humans at such tasks. [...] The standard approach here is to view pattern recognition as a classification problem. [...] The basic strategy is to give the computer a large amount of labeled data: samples that have already been classified[...]. Because each sample comes with a label (i.e., its class), the computer can use various analytical tricks to extract characteristics of each class. When it is later presented with an unlabeled

sample, the computer can guess its class by choosing the one whose characteristics are most similar to the unlabeled sample (MacCormick, 2012, 81-82).

Può accadere però è che, a causa dell'opacità del processo di addestramento e di quello di attribuzione della classificazione, si possano attivare meccanismi implicitamente discriminatori quando si tratta di profilare gli utenti delle piattaforme come appartenenti a cluster precisi. Inoltre è possibile che l'attività della raccomandazione dei *social network* che tende a presentare con maggiore visibilità contenuti simili a quelli che sono stati visti o commentati in passato, possa condurre alla *filter bubble* (Pariser, 2011). La bolla del filtraggio impedisce alle persone di entrare in contatto con opinioni discordanti dalla propria, perché si favoriscono contenuti più affini a quelli ciascuno ha mostrato di gradire di più.

Wendy Chun (2019) suggerisce che la tendenza di questo processo di classificazione, espulsione, discriminazione presuppone la ricerca dell'omofilia nella costruzione delle classi identificate. Le parole che si trovano spesso vicino alle stesse parole sono simili, le persone che abitano in una stessa zona sono simili, le persone che prediligono certi consumi sono simili, ecc.: come se nella rappresentazione di questa somiglianza e nell'amore per l'omologo, non fosse già in azione, una struttura dell'interpretazione, una certa performatività, come mostrano gli algoritmi per il *credit scoring*, o il supporto algoritmico alle attività della polizia (Predpol, Predictive Policing). Secondo Chun:

> Crucially, these algorithms perpetuate the discrimination they "find." They are not simply descriptive but also prescriptive and performative in all senses of that word. Capture systems, as Phil Agre theorized in 1994, reshape the activities they model or "discover." Through a metaphor of human activity as language, they impose a normative "grammar of action" as they move from analyzing captured data to building an epistemological model of the captured activity (Chun, 2019, 66).

Tale rischio di performatività e prescrittività degli algoritmi di *machine* e *deep learning* è segnalato anche da un report della Rand Corporation dal titolo significativo: *An intelligence in our image* (Osoba e Welser, 2017). La Rand è un'importante agenzia di ricerca a supporto della presa di decisione pubblica, che ha avuto un ruolo notevole nello sviluppo dell'informatica. Il risultato di questo report di monitoraggio e analisi dell'uso degli strumenti di Intelligenza artificiale per prendere decisioni pubbliche e politiche è piuttosto inquietante. La tesi del report è che:

> the opacity of algorithms makes it harder to judge correctness, evaluate risk, and assess fairness in social applications. It can also obscure the causal understanding behind decisions. These issues might be harmless if algorithms were (near) infallible. But most algorithms have only probabilistic guarantees of accuracy. And this is in the best possible scenarios, in which the right models

and algorithms are applied appropriately, with the best intention to "perfect" data. Algorithm designers and users rarely have the luxury of such perfect scenarios. They must rely on assumptions that can fail and lead to unexpected results (Osoba e Welser, 2017, 3).

7. Osservazioni finali: un dialogo tra data science e critical algorithmic studies

Il *think tank* politico della Rand condivide le stesse perplessità degli studiosi critici di *algorithmic studies* a proposito dei rischi di meccanismi di classificazione inavvertiti. Del resto il problema della definizione di modelli basati sul software richiama alla memoria anche le obiezioni di Weizenbaum (1976) all'uso del software per costruire modelli. Nel suo testo *Computer power and human reason*, Weizenbaum pose in discussione l'uso del software al posto dei modelli matematici per valutare una teoria. La sua perplessità era fondata su una questione molto concreta legata al funzionamento del software in un contesto di opacità.

Come inventore del famoso *chatbot* Eliza, capace di interpretare in una conversazione una psicoanalista rogersiana, Weizenbaum sapeva bene quanto fossero controversi e complessi i rapporti tra teoria e modelli. Per lui i modelli testati attraverso il computer svolgevano l'attività di interpretazione dei dati che dovevano essere spiegati. Ma come fa un programma che governa l'esecuzione di un compito a costituirsi come una spiegazione, senza incorrere in possibili errori di interpretazione?

Le perplessità sull'uso talvolta distorto di tecniche di *machine learning* non vengono solo dall'ambito dei *critical algorithmic studies* e da riviste scientifiche come *Big Data and Society*, ma anche da voci autorevoli interne al campo di studi. Secondo Zachary Lipton e Jacob Steinhardt (2018) ci sarebbero alcune cattive abitudini epistemiche nel settore che sarebbe bene correggere:

Innanzi tutto: "failure to distinguish between explanation and speculation". In un contesto nel quale inequivocabilmente si maneggia la capacità di interpretare i dati, non sempre sarebbe possibile distinguere le speculazioni dalle dimostrazioni vere e proprie. Un altro possibile errore sarebbe la tendenza all'offuscamento dei risultati o con l'eccesso di inutile matematizzazione (confondendo concetti tecnici e non tecnici) e il cattivo uso del linguaggio: "by choosing terms of art with colloquial connotations or by overloading established technical terms" (Lipton e Steinhardt, 2018). Tutti rischi che minano la capacità dei metodi di costituire dei buoni modelli per il fenomeno da spiegare.

Secondo Anja Bechmann e Geoffrey Bowker (2019) dobbiamo analizzare anche il ruolo dei progettisti umani di AI quando si applicano a creare modelli di *machine learning* per risolvere problemi di classificazione. Abbiamo visto nell'articolo che tutti i metodi di *pattern reco-*

gnition possono, a certe condizioni, essere spiegati come meccanismi di classificazione. Il modo in cui sono costruite le etichette delle classi e sono identificate le caratteristiche e gli attributi di ciascuna classe è deciso dal programmatore. È questa dinamica completamente umana che può condurre a selezionare i dati in maniera pregiudiziale o stereotipica e concludersi quindi con un esito ingiusto e irresponsabile dell'algoritmo che viene costruito. Bechmann e Bowker (2019) costruiscono un frame per valutare l'impatto della progettazione umana non avvertita e delle possibili conseguenze implicitamente discriminatorie degli esiti della categorizzazione. Tra le attività che vengono esercitate e che possono nascondere insidie classificatorie segnalano: *data collection* (il modo in cui si raccolgono e si organizzano i dati); *data cleaning* (il modo in cui i dati vengono puliti per essere trattati dall'algoritmo); *model training* (la scelta dei dati che costituiscono il campione di addestramento del modello previsto).

Gli autori mostrano un caso di studio che usa il metodo Latent Dirichlet Allocation (LDA) Text2Vec (text3vec.org) come modello non supervisionato per fare analisi semantica nell'ambito del News Feed di Facebook, al fine di comprendere come costruire messaggi in modo che diventino virali in una certa popolazione. Nell'esempio si osserva come la selezione dei risultati da ottenere impatta sul modo in cui sono costruiti i dati, su come viene definito il modello per identificare i cluster rilevanti, sulla preparazione dei dati e infine sul campione di addestramento. Ognuno di questi passaggi, se non è svolto con consapevolezza epistemica, può avere effetti discriminatori, marginalizzando popolazioni non interessanti e/o silenziando la loro voce.

> Doing so means trying to think computationally and to understand why the model choses to cluster these particular words together and at the same time we tend only to use words that are actually meaningful, otherwise we return to the data preparation phase. The political power that lies in interpreting the probability scores and labeling the clusters is enormous and as a consequence is setting the agenda for what is actually inferred from the data (Bechmann e Bowker, 2019, 6).

La citazione si riferisce alla preparazione di un modello di classificazione che clusterizza e associa i topic del discorso nella News Feed di Facebook e che ci illustra i potenziali elementi di discriminazione esercitati nella preparazione del modello da parte dei progettisti. Il punto debole non è rappresentato dal modello linguistico adottato, ma dall'organizzazione dei dati e dal criterio di addestramento funzionale al risultato desiderato.

Secondo il Report della Rand corporation sui rischi di pregiudizi introdotti dall'uso di algoritmi per il *pattern recognition* che agiscono in maniera poco chiara o superficiale, le azioni possibili per limitare i danni

sono di tre tipi: "avoiding algorithms altogether, making the underlying algorithms transparent, or auditing the output of algorithms" (Osoba e Welser, 2017). È improbabile evitare l'uso degli algoritmi ed è molto difficile rendere il processo di funzionamento degli algoritmi trasparente, sia a causa della mancanza di cultura informatica del largo pubblico, sia a causa della impossibilità di comprenderne il meccanismo, soprattutto quando le tecniche usate sono quelle del *deep learning*, sebbene alcuni ricercatori propongano la presenza di esperti indipendenti come valutatori (Shneiderman, 2016). Resterebbe, quindi, solo la possibilità di valutare i risultati forniti come output. Ma anche questa opzione sembra difficile da realizzare, perché quasi mai si controllano i risultati dei costosi algoritmi acquistati per usarli nell'ambito della presa di decisione nei più diversi contesti. Per esempio, pure se in molte città americane si stanno sperimentando i programmi di *predictive policing*, non ci sono studi che dimostrano che i programmi funzionano correttamente e tuttavia ciò non impedisce di adottarli (Haskins, 2019). Sembra che la frenesia della misurabilità, che ha contagiato tutti gli ambiti delle scienze sociali non, riguardi anche l'affidabilità degli algoritmi.

Eppure, come suggerisce lo studio di David Moats e Nick Severs (2019) "Experimenting in the 'divide' between data science and critical algorithm studies", valutare l'efficacia di questi algoritmi potrebbe essere utile sia per i *data scientist*, per evitare di incorrere negli errori segnalati da altri colleghi (Lipton e Steinhardt, 2018), sia per gli studiosi critici degli algoritmi, per imparare a uscire dalla torre d'avorio ed entrare a contatto con la dimensione sperimentale e i vincoli concreti della scienza dei dati.

Riferimenti bibliografici

Agre, P. (1994). Surveillance and Capture: Two Models of Privacy. *The Information Society*, *100*, pp. 101-127.

Angwin, J., Larson, L., Mattu, S., Kirchner, L. (2016). Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks. *ProPublica*, May 23, 2016, https://www.propublica.org/article/ machine-bias-risk-assessments-in-criminal-sentencing (ultima consultazione 10/08/2019).

Barocas, S., Selbst, A.D. (2016). Big data's disparate impact. *Cal. L. Rev.*, *104*, pp. 671-732.

Bechmann, A., Bowker, G.C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society*, *6*, 1, 2053951718819569.

Ben-David, S., Hrubeš, P., Moran, S., Shpilka, A., Yehudayoff, A. (2019). Learnability can be undecidable. *Nature Machine Intelligence*, *1*, 1, pp. 44-48.

Bertrand, M., Chugh, D., Mullainathan, S. (2005). Implicit discrimination. *American Economic Review*, *95*, 2, pp. 94-98.

Binns, R. (2017). Fairness in machine learning: Lessons from political philosophy. *arXiv preprint arXiv:1712.03586*.

Bolukbasi, T., Chang, K.W., Zou, J., Saligrama, V., Kalai, A. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. *Advances in neural information processing systems*, pp. 4349-4357.

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *BigData & Society*, *3*, 1, 2053951715622512.

Caliskan, A., Bryson, J., Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356, pp. 183-186.

Calude, C.S., Longo, G. (2017). The deluge of spurious correlations in big data. *Foundations of science*, *22*, 3, pp. 595-612.

Castelvecchi, D. (2019). Machine learning leads mathematicians to unsolvable problem. *Nature*, *565*, 7739, p. 277.

Chouldechova, A., & Roth, A. (2018). The frontiers of fairness in machine learning. *arXiv preprint arXiv:1810.08810*.

Chun, W.H. (2019). Queerying Homophily. In C. Apprich, W.H. Chun, F. Cramer e H. Steyerl, *Pattern Discrimination*. Minneapolis: University of Minnesota Press and Meson Press, pp. 59-97.

Church, K.W. (2017). Word2Vec. *Natural Language Engineering*, *23*, 1, pp. 155-162.

Citron, D.K., Pasquale, F. (2014). The scored society: due process for automated predictions. *Wash. L. Rev.*, *89*, 1.

Cohen, P.J. (1963). The independence of the continuum hypothesis. *Proc. Natl Acad. Sci. USA*, *50*, pp. 1143-1148.

Cohen, P.J. (1964). The independence of the continuum hypothesis, II. *Proc. Natl Acad. Sci.* USA, *51*, pp. 105-110.

Cramer, F. (2019). Crapularity hermeneutics: interpretation as the blind spot of analytics, artificial intelligence, and other algorithmic producers of the postapocalyptic peresent. In C. Apprich, W.H. Chun, F. Cramer e H. Steyerl, *Pattern Discrimination*. Minneapolis: University of Minnesota Press and Meson Press, pp. 23-58.

Day, M. (2016). How LinkedIn's search engine may reflect a gender bias. *The Seattle Times*, https://www.seattletimes.com/business/microsoft/how-linkedins-search-engine-may-reflect-a-bias/ (ultima consultazione 10/08/2019).

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, K. Foot (a cura di), *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, Mass: MIT Press, pp. 167-194.

Gödel, K. (1940). *The Consistency of the Continuum Hypothesis*. Princeton: Princeton University Press.

Goodman, B., Flaxman, S. (2016). European Union regulations on algorithmic decision-making and a "right to explanation". *arXiv preprint arXiv:1606.08813*.

Greenwald, A.G., McGhee, D.E., Schwartz, J.L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, *74*, 6, pp. 1464-1480.

Hargreaves, E., Agosti, C., Menasché, D., Neglia, G., Reiffers-Masson, A., Altman, E. (2018, August). Biases in the Facebook News Feed: A Case Study on the Italian Elections. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 806-812.

Haskins, C. (2019). Dozens of Cities Have Secretly Experimented With Predictive

Policing Software. *Motherboard* 6 febr. 2019, https://motherboard.vice.com/en_us/article/d3m7jq/dozens-of-cities-have-secretly-experimented-with-predictive-policing-software (ultima consultazione 10/08/2019).
Hume, D. (1888). *A Treatise of human nature* [1739]. Ed. by L.A. Selby-Bigge. Oxford: Clarendon Press.
Hunt, E. (2016). Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. *The Guardian*, 24/3/2016, https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter (ultima consultazione 10/08/2019).
Hynes, C. (2016). The Airbnb bias row shows prejudice matters – and not just to its victims. *The Guardian*, 1/11/2016, https://www.theguardian.com/commentisfree/2016/nov/01/airbnb-bias-prejudice-sharing-sites-discrimination-black-lgbt-travellers (ultima consultazione 10/08/2019).
Kitchin, R. (2014*). The data revolution: Big data, open data, data infrastructures and their consequences*. New York: Sage.
Landauer, T.K., Foltz, P.W., Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 2-3, pp. 259-284.
Lavergne, M. Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American economic review*, *94*, 4, pp. 991-1013.
Lazer, D., Kennedy, R., King, G., Vespignani, A. (2014). The parable of Google Flu: Traps in 796 Big data analysis. *Science*, *343*, 14 March, pp. 1203-1205.
Lee, P. (2016). Learning from Tay's introduction. Blog, Microsoft, March 25, 2016: http://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/#sm.0001 v8vtz3qddejwq702cv2annzcz (ultima consultazione 10/08/2019).
Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian journal of linguistics*, *20*, 1, pp. 1-31.
Leonelli, S. (2018). *La ricerca scientifica nell'era dei big data*. Milano: Meltemi.
Lipton, Z.C., Steinhardt, J. (2018). Troubling trends in machine learning scholarship. *arXiv* preprint arXiv:1807.03341.
Longo, G. (2009). Critique of Computational Reason in the Natural Sciences. In E. Gelenbe e J.-P. Kahane (a cura di), *Fundamental Concepts in Computer Science*. London: Imperial College Press, pp. 43-70.
MacCormick, J. (2012), *Nine Algorithms that Changed the Future: The Ingenious Ideas that Drive Today's Computers*, Princeton: Princeton University Press.
Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013*a*). Efficient estimation of word representations in vector space. *arXiv* preprint arXiv:1301.3781.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J. (2013*b*). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pp. 3111-3119.
Mikolov, T., Yih, W.T., Zweig, G. (2013c). Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 746-751.
Miller, B., Record, I. (2017). Responsible epistemic technologies: A social-epistemological analysis of autocompleted web search. *New Media & Society*, *19*, 12, pp. 1945-1963.
Moats, D., Seaver, N. (2019). "You Social Scientists Love Mind Games":

Experimenting in the "divide" between data science and critical algorithm studies. *Big Data & Society*, *6*, 1, 2053951719833404.
Muchnik, L., Aral, S., Taylor, S.J. (2013). Social influence bias: A randomized experiment. *Science*, *341*, 6146, pp. 647-651.
Noble, S.U. (2018). *Algorithms of Oppression: How search engines reinforce racism*. New York: NYU Press.
Nosek, B.A., Banaji, M.R., Greenwald, A.G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*, 1, p. 101.
Numerico, T. (2017). La memoria e la rete. In A. Bertollini e R. Finelli (a cura di), *Soglie del linguaggio. Corpo, mondi, società*, Roma: RomaTre-Press, pp. 81-102.
Numerico, T. (2019). Politics and epistemology of big data: a critical assessment. In M.V. D'Alfonso e D. Berkich (a cura di), *On the cognitive, ethical and scientific dimension of artificial intelligence*. Dordrecht: Springer, pp. 147-166.
O'Neil, C. (2016). *Weapons of math destruction*. London: Penguin.
Osoba, O.A., Welser IV, W. (2017). *An intelligence in our image: The risks of bias and errors in artificial intelligence*. Santa Monica, CA: RAND Corporation: https://www.rand.org/pubs/research_reports/RR1744.html.
Pariser, E. (2011). *The filter bubble*, New York: Penguin.
Parliament and Council of the European Union (2016). *General Data Protection Regulation*.
Pasquale, F. (2015). *Black box society*. Cambridge (MA): Harvard University Press.
Pennington, J., Socher, R., Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532-1543.
Sahlgren, M. (2008). The distributional hypothesis. *Rivista di Linguistica*, *20*, 1, pp. 33-53.
Scheiber, N. (2018). Facebook Accused of Allowing Bias Against Women in Job Ads. *New York Times*, https://www.nytimes.com/2018/09/18/business/economy/facebook-job-ads.html (ultima consultazione 10/08/2019).
Shneiderman, B. (2016). Opinion: The dangers of faulty, biased, or malicious algorithms requires independent oversight. *Proceedings of the National Academy of Sciences*, *113*, 48, pp. 13538-13540.
Silva, S., Kenney, M. (2018). Algorithms, Platforms, and Ethnic Bias: An Integrative Essay. *Phylon*, *55*, 1-2, pp. 9-37.
Steyerl, H. (2019). A Sea of Data: Pattern Recognition and Corporate Animism (Forked Version). In C. Apprich, W.H. Chun, F. Cramer e H. Steyerl, *Pattern Discrimination*. Minneapolis: University of Minnesota Press and Meson Press, pp 1-22.
Sumpter, D. (2018). *Outnumbered*. London: Bloomsbury.
Weizenbaum, J. (1976). *Computer power and human reason*. New York: Penguin Books.
Wiener, N. (1948/1961). *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge (MA): MIT Press.
Wiener, N. (1950/1954). *The human use of human beings*. Boston: Houghton Mifflin.

Social networks and machine learning algorithms: Cognitive problems and propagation of biases

The objective of this paper is to critically analyze the potential discriminatory and biased effects of the spread of algorithmic techniques for the interpretation of human behaviors, based on unsupervised machine learning methods, trained on uncontrolled data, produced by social networks users. I will introduce the epistemological general issue starting from the exam of the performance of two algorithms for textual analysis, such as GloVe and Word2vec. They are, at present, two of the most successful tools for textual analysis based on word embedding. Such algorithms, trained on public domain textual databases, tend to associate words by replicating gender and ethnic stereotypes, because they infer connections between words following the probability of their distances in the training sets. The meaning model that informs their judgments pushes them to rely on a stereotyped and biased representation of social, gender and ethnic categories, that are intertwined in the databases on which they are trained (Caliskan, Bryson and Narayanan 2017; Bolukbasi *et al*. 2016). The Natural Language Processing (NLP) techniques are just examples of the possible use of algorithms for profiling users of social networks with the aim of predicting their behaviors, or of suggesting their preferences, or of nudging them toward believing or desiring something predetermined by the system. The process is based on the analysis of past behaviors, preferences, or actions in order to create groups, clusters or categories. This complex system is founded on two equivocal premises. The first is that the huge increase of data allows the algorithms to work with such a quantity of information that prevents the distortive effects and potential mistakes to become relevant in terms of prevision. The second hypothesis is that algorithmic methods interpret more efficiently than human beings the meaning of available data on users, and consequently are capable of capturing its cognitive value for the scope of a trustworthy, univocal categorization and for the prediction of the probability of future events. Both hypotheses are not demonstrated or validated; they are just assessed and rhetorically supported by the major agents of the successful Big Data field. Moreover the combination of dirty, old or non-controlled data, of rigidity in the inferential capacity of learning algorithms and of the utilitarian orientations of their experimental design, could produce socially dangerous outputs in terms of interpretations and predictions in social sciences. It is desirabile to better understand rules and criteria adopted in machine learning algorithms used for socially sensitive analysis, with special regards to data acquired in social networks, in order to guarantee the respect of fairness and minority protections in judgments and predictions. It is also relevant to avoid secret methods in decisions taking processes that impact on people life and social justice, because the evaluation of algorithms only from their outputs is not enough, due to the asymmetry of power in terms of knowledge distribution between those who are and those who are not in control of data.

Keywords: algorithmic fairness, Big Data and biased decisions, disparate impact, NLP algorithms, social networks.

Teresa Numerico, Università degli studi Roma Tre, Dipartimento di Filosofia, Comunicazione e Spettacolo, Via Ostiense 234, 00146 Roma, teresa.numerico@uniroma3.it

FRANCESCA PANZERI

STAI SCHERZANDO? (NON) RICONOSCERE L'IRONIA NEI SOCIAL NETWORK

1. Introduzione

Uno stesso commento, per esempio "Salvini è un grande politico", può essere inteso in senso letterale, e quindi come un attestato di stima nei confronti del Ministro dell'Interno del primo governo Conte, oppure in senso ironico, ossia come una valutazione negativa del suo operato. Per mezzo dell'ironia, infatti, chi parla ha l'intenzione di manifestare il proprio dissenso verso l'interpretazione letterale dell'enunciato proferito, arrivando quindi a comunicare al suo interlocutore un significato opposto rispetto a quanto detto. L'ironia è una forma di linguaggio figurato, come la metafora, l'iperbole o l'attenuazione, e per ricostruire il significato realmente inteso dal parlante l'ascoltatore deve andare oltre il significato letterale e inferire l'intenzione comunicativa di chi parla. Come vedremo, il parlante ironico può servirsi di particolari segnali meta-comunicativi che facilitano il suo uditorio nel riconoscimento del significato inteso, ed evitare quindi fraintendimenti. Nel momento però in cui un commento compaia in forma scritta, alcuni di questi marcatori dell'ironia (a livello di intonazione e di espressione) sono necessariamente assenti.

Scopo di questo contributo è quello di analizzare l'espressione, e soprattutto il riconoscimento, dell'ironia nei social network, con l'obiettivo di valutare quali siano i marcatori dell'ironia, e se il rischio di fraintendimenti risulti maggiore rispetto alle interazioni faccia a faccia. La questione è rilevante, non solo per l'elevata e sempre crescente diffusione di questi mezzi di comunicazione (a gennaio 2019 risultano quasi tre miliardi e mezzo di utenti attivi nei social media nel mondo, e 35 milioni in Italia, ossia il 59% della popolazione, con un tempo medio di quasi due ore al giorno passato sui social[1]), ma soprattutto perché le caratteristiche della comunicazione attraverso i social media

[1] Dati tratti dal report annuale DIGITAL 2019: ITALY, a cura di Hootsuite & We Are Social, pubblicato il 31 gennaio 2019, https://datareportal.com/reports/digital-2019-italy?rq=italy.

pongono questioni interessanti. In primo luogo, infatti, la forma scritta e a distanza non permette la codifica di informazioni prosodiche (è assente l'intonazione con cui una frase verrebbe pronunciata) e di tipo espressivo-gestuale (non è accessibile il comportamento corporeo che accompagnerebbe l'enunciazione del commento), indizi che facilitano il riconoscimento dell'intento sarcastico. Inoltre, il tipo di platea cui sono rivolti i commenti scritti sui social differisce dal classico uditorio delle interazioni faccia a faccia, non solo in termini di numerosità, ma anche, soprattutto, per il fatto che chi legge un posto non condivide le stesse informazioni di chi lo ha scritto (sia perché non si trova nello stesso luogo e tempo, sia perché può non conoscere le opinioni personali di chi scrive), e la mancanza del cosiddetto terreno comune tra mittente e ricevente può ostacolare la comprensione del significato globale che voleva essere comunicato.

Per poter quindi affrontare la questione di come avvenga il riconoscimento dell'ironia nei social media, cominceremo introducendo la nozione di ironia verbale (§ 2), e dei segnali che facilitano il riconoscimento dell'ironia nelle interazioni faccia a faccia (§ 3), per poi passare brevemente in rassegna degli studi volti a trovare un algoritmo in grado di riconoscere in maniera automatica l'ironia presente nei social network (§ 4). Passeremo poi ad analizzare alcuni casi recenti di commenti su social network che non sono stati riconosciuti come ironici (§ 5), e quindi trarre le conclusioni (§ 6) sulla questione del (non) riconoscimento dell'ironia nei social network.

2. L'ironia

L'interpretazione ironica di una frase si fonda sulla constatazione di un'incongruenza tra ciò che è stato detto e la situazione o comunque in generale le aspettative contestuali (Attardo, 2000): in una gelida giornata sferzata da pioggia battente, l'enunciazione di frasi come "È una bellissima giornata" piuttosto che "Adoro le giornate di sole" farà scattare una loro interpretazione ironica perché risultano semplicemente false (la prima) o violano le aspettative di commenti pertinenti alla situazione (la seconda). Approcci tradizionali nell'ambito della retorica considerano l'ironia una forma di linguaggio figurato, in cui ciò che il commento ironico di fatto significa è l'opposto della sua interpretazione letterale. Come ci ricorda Ervas (2011), Quintiliano, portavoce della retorica classica, nell'*Institutio oratoria* definisce l'ironia come quel tropo in cui si deve intendere il contrario di ciò che si dice letteralmente (*contrarium quod dicitur intelligendum est*). Anche Grice (1975) propone che nell'ironia il significato del parlante – ciò che vuole realmente comunicare – si sostituisca al significato letterale della frase, ma nel suo approccio la sostituzione avverrebbe a livello pragmatico, non semantico: una frase

come "È una bellissima giornata" manterrebbe il suo significato semantico, ma se qualcuno la pronuncia in una giornata di pioggia battente, l'interlocutore si interrogherà sul motivo che ha portato il parlante a dire qualcosa di manifestamente falso (a violare apertamente la Massima di Qualità secondo la terminologia griceana), e per "salvare" la cooperatività del parlante a un livello più profondo l'ascoltatore deriverebbe come significato implicito (implicatura conversazionale) che il parlante di fatto intendeva comunicare che la giornata non è affatto bella.

Sperber e Wilson (1981) muovono una serie di obiezioni a questi approcci: in primo luogo, la plateale falsità di un enunciato non è né necessaria né sufficiente ad assicurarne l'interpretazione ironica, visto che esistono enunciati ironici che risultano comunque veri (come "Adoro le giornate di sole"), e non tutte le palesi falsità risultano ironiche, come già ammesso da Grice (1978). Ma soprattutto contestano l'idea che un commento ironico *significhi* (a livello semantico o a livello pragmatico) l'opposto di ciò che viene detto. Nella proposta di Sperber e Wilson, quando una persona dice "È una bellissima giornata" in senso ironico non sta *usando* questa frase (per descrivere la situazione), ma la sta *menzionando*, con l'intento di fare eco a quanto detto, o anche solo supposto, da qualcuno. Nel fare riferimento a tale enunciato, l'intento dell'ironista sarebbe quello di manifestare il proprio atteggiamento di scherno e ludibrio nei confronti del pensiero espresso da tale frase, e nel momento in cui l'interlocutore riconosce questo comportamento denigratorio può quindi inferire che l'ironista non crede nella verità di quanto sta letteralmente affermando. L'idea che il commento ironico non assolva alla normale funzione di descrizione della realtà è accettata anche da Clark e Gerrig (1984), che propongono che l'ironia, coerentemente alla sua etimologia (dal greco εἰρωνεία, dissimulazione), sia una forma di pretesa, di fare finta (*pretense*): dicendo "È una bellissima giornata" sotto il diluvio, l'ironista starebbe facendo finta di essere una persona così stupida da pensare una cosa del genere, e quindi starebbe ridicolizzando non solo quel pensiero (che sia una bella giornata), ma anche la persona che potrebbe dirlo, e gli altri che potrebbero crederci.

Negli ultimi trent'anni si è sviluppato un acceso dibattito tra sostenitori della teoria ecoica e della teoria della pretesa, in particolare sulla capacità di coprire adeguatamente il maggior numero possibile di forme di ironia[2]. Per il momento, ci interessa sottolineare come sia nella teoria ecoica di Sperber e Wilson sia nella teoria della pretesa di Clark e Gerrig

[2] Sperber e Wilson hanno chiarito, e parzialmente modificato, la loro teoria ecoica in una serie di scritti (Sperber, 1984; Wilson, 2006; Wilson e Sperber, 1992; 2012); teorie che combinano l'aspetto attributivo tipico della prospettiva ecoica con l'idea che l'ironista stia facendo finta di essere una persona così ottusa da enunciare quel commento sono state difese, tra gli altri, da Kumon-Nakamura, Glucksberg e Brown (1995), Recanati (2005) e Currie (2006).

si assuma che il commento ironico mantenga il suo significato letterale, e che ciò che l'ironista intenda davvero comunicare sia il suo atteggiamento derogatorio e di scherno – nei confronti del pensiero espresso dal commento nella teoria ecoica, o delle persone che farebbero un tale commento o vi acconsentirebbero nella teoria della pretesa.

Le teorie discusse si sono focalizzate sul tema di quale fosse il significato comunicato dall'ironista. Ribaltando la prospettiva, e mettendoci nei panni dell'interlocutore, la questione riguarda il come questi possa *riconoscere*, accorgersi, che un determinato commento è da intendersi in senso ironico. Diversamente da altre forme di linguaggio non-letterale, come la metafora, l'enunciato ironico può infatti ricevere una interpretazione letterale plausibile: se mio fratello mi telefona da Londra e mi chiede com'è il tempo, e io rispondo "È una bellissima giornata", se lui non vede come me la finestra rigata da rigagnoli di gelida acqua piovana, può non cogliere la vena ironica del mio commento. Se invece gli rispondessi semplicemente "Le strade sono fiumi", l'interpretazione letterale di quanto dico porterebbe a una immediata anomalia semantica, che attiverebbe una lettura metaforica, senza rischi di fraintendimenti. Per individuare l'intento ironico di un commento, l'ascoltatore tipicamente si basa sulla incongruenza di quanto è stato detto rispetto alla situazione: la palese falsità, o comunque inappropriatezza, del commento interpretato letteralmente è un forte indizio di ironia (Wilson, 2006: 1724). Si noti però che quando i commenti esprimono valutazioni personali e soggettive, come il "Salvini è un grande politico" da cui eravamo partiti, per poter capire se sono in linea o in contrasto con la situazione, ossia se sono detti in maniera sincera o ironica, non è sufficiente essere immersi nella stessa situazione comunicativa, ma bisogna anche conoscere le opinioni personali di chi parla. In altre parole, per riuscire a comprendere l'inappropriatezza di un commento ironico, l'ironista e il suo uditorio dovrebbero condividere le stesse informazioni (relative alla situazione, o alle credenze di chi parla) necessarie per individuare l'inconguenza del commento ironico, il cosiddetto terreno comune (*common ground*, Stalnaker, 2002).

Ma visto che non sempre si può avere la certezza che l'uditorio condivida le informazioni rilevanti, per evitare malintesi e fallimenti comunicativi, chi ricorre all'ironia dovrebbe seguire quello che Kreuz (1996) chiama *Principle of Inferability*, ossia sincerarsi che la sua reale intenzione comunicativa possa essere riconosciuta.

3. Riconoscere l'ironia nelle interazioni faccia a faccia

Per evitare che un suo commento venga frainteso, il parlante ironico può servirsi dei cosiddetti *marcatori dell'ironia*, una serie di indizi meta-comunicativi che aiutano l'interlocutore a riconoscere l'intento

ironico, e quindi a inferire correttamente il significato comunicato. Nelle lingue parlate, e nelle interazioni faccia a faccia, questi indizi sono di tipo fonologico, lessicale, sintattico, e gestuale. A livello fonologico, si può individuare una particolare intonazione prosodica delle frasi, il cosiddetto tono di voce ironico (Anolli, Ciceri e Infantino, 2002; ma si veda anche Bryant e Fox Tree, 2005 per una critica), un ritmo di parlato più lento, con allungamento di alcune sillabe. A livello lessicale, invece, si nota la presenza di aggettivi e avverbi cosiddetti "estremi" (Kreuz e Roberts, 1995): se in una giornata di pioggerella autunnale dico "È una bella giornata!", il mio interlocutore potrebbe pensare che la mia natura malinconica mi porti ad apprezzare delle giornate uggiose; se però esclamo "È una giornata spettacolare!", la polarizzazione estrema della valutazione segnala più agevolmente l'incongruenza del commento, e quindi il mio intento ironico. A livello sintattico, la presenza di costruzioni particolari, come le esclamazioni ("Bella giornata!"), le cosiddette *tag questions* ("bella giornata, no?"), la topicalizzazione ("una bella camicia ti sei messo") avrebbe la funzione di attirare l'attenzione del destinatario, e facilitarlo nel riconoscimento dell'ironia (Seto, 1998). Se ci concentriamo poi sulle espressioni e sul comportamento gestuale di chi parla, l'ironia può essere segnalata dalla posizione delle sopracciglia (sollevate o corrugate), degli occhi (spalancati, strizzati, fatti roteare), da movimenti del capo (l'annuire) e dal sorriso (Attardo, Eisterhold, Hay e Poggi, 2003; Burgers e van Mulken, 2017).

Come già accennato, diversamente dai fattori dell'ironia (incongruenza e atteggiamento valutativo), che sono necessari per poter considerare un commento come ironico, i marcatori dell'ironia non sono essenziali: un commento ironico rimane tale anche in assenza di qualsiasi marcatore di ironia – potrebbe solo risultare più difficile da riconoscere, e quindi essere frainteso. Mantovan, Giustolisi e Panzeri (2019), analizzando gli indizi paracomunicativi di ironia nella lingua dei segni italiana, sostengono che i marcatori svolgano diverse funzioni: alcuni di questi sono segnali di allerta (*warning*): rallentare il ritmo del parlato (o del segnato), così come particolari espressioni facciali come gli occhi sgranati e le sopracciglia inarcuate, indurrebbero nell'uditorio il sospetto che quanto detto/segnato non è da intendersi in senso letterale. Altri segnali svolgono una funzione di sovra-enfatizzazione: la forma iperbolica, realizzata attraverso la scelta di vocaboli estremi e costruzioni superlative, servirebbe per amplificare ulteriormente la discrepanza tra l'interpretazione letterale di quanto detto/segnato e la situazione, e quindi facilitare il riconoscimento della incongruenza del commento. Infine, ulteriori marcatori avrebbero uno scopo correttivo, segnalando esplicitamente che l'interpretazione intesa deve essere ribaltata rispetto a quella letterale, come ad esempio il "no?" aggiunto in coda ad un commento.

Gli studi sopra citati si sono focalizzati sull'espressione dell'ironia nelle interazioni faccia a faccia, attribuendo un ruolo a marcatori espressi a livello fonologico, e quindi tramite una specifica intonazione nelle lingue parlate, e grazie alla presenza di marcatori non manuali nelle lingue dei segni, e anche a particolari comportamenti gestuali e corporei. Che cosa succede invece se il commento non è pronunciato in presenza di un uditorio fisicamente presente, ma è scritto? Su quali indizi si possono basare i lettori per inferirne l'intento ironico, e non scambiarlo per un commento sincero?

4. Riconoscere automaticamente l'ironia nei social network

Si consideri la differenza tra l'affermare che, e il domandare se, Leo è partito. Se siamo in una conversazione *vis à vis*, in italiano è l'intonazione a distinguere questi due atti linguistici: pronunciare "Leo è partito" con un tono finale discendente lo rende una constatazione, se invece il tono finale è ascendente otteniamo una domanda. Se invece lo stesso enunciato si trova scritto, è la punteggiatura a permetterci di individuare la funzione dell'enunciato: l'affermazione è seguita dal punto fermo, la domanda dal punto interrogativo. Come raccontato in Houston (2013), già dal 1600 vennero proposti dei segni di punteggiatura che avevano proprio la funzione di segnalare che la frase era da intendersi in senso ironico, e che quindi doveva essere interpretata con significato opposto a quello letterale: si trattava spesso di variazioni del punto interrogativo, ribaltato secondo l'asse orizzontale o verticale, ma questi segni di punteggiatura non ebbero successo, almeno fino alla invenzione delle emoticon. La prima emoticon viene fatta risalire a un messaggio scritto nel 1982 da Scott Elliot Falham, all'interno di un *Bullettin Board* universitario. Falham, attualmente professore emerito alla Carnegie Mellon University, racconta[3] che nel forum venivano postati messaggi di vario tipo, con informazioni e richieste serie, ma anche con commenti che erano (o cercavano di essere) umoristici. In molti casi, però, il sarcasmo del messaggio non veniva colto, e questo generava incomprensioni, con conseguenti lunghe diatribe, e polemiche che montavano col passare del tempo. Falham propose allora di utilizzare la sequenza di particolari segni di punteggiatura, ossia :-), come *joke marker*, marcatori dello scherzo. Le emoticon hanno poi avuto una diffusione estremamente ampia, non solo per marcare le frasi da intendere in maniera ironica, ma per raffigurare le emozioni e gli stati d'animo connessi alla espressione di contenuti, e alcuni contenuti stessi. Rimane però aperta la questione se – oltre alla

[3] Il post "Smiley Lore :-)" che ricostruisce la vicenda, scritto nel 2002 in occasione del ventennale della invenzione dell'emoticon, è accessibile dalla *home page* del professor Scott Falham: http://www.cs.cmu.edu/~sef/sefSmiley.htm.

eventuale presenza di emoticon – il passaggio dalla conversazione in presenza alla scrittura ha comportato altri cambiamenti, e quindi se è possibile individuare specifici marcatori di ironia nei testi scritti.

Diversi studi si sono posti l'obiettivo di sviluppare un algoritmo che fosse in grado di riconoscere in maniera automatica i commenti che erano da intendere come ironici. Lo scopo è, ovviamente, pratico ed economico: un'azienda che abbia appena messo in commercio un nuovo modello di sneakers vuole ad esempio sapere se il commento di un utente "Queste scarpe sono davvero fantastiche" sia da intendersi come sincero – e quindi come un complimento – oppure come smaccatamente sarcastico – e quindi estremamente negativo.

Per poter quindi individuare degli indizi che permettano di discriminare i messaggi sinceri da quelli ironici, l'idea alla base di questi approcci è quella di selezionare una serie di commenti che siano chiaramente ironici (per mezzo di annotazioni manuali, oppure basandosi sulla presenza dell'hashtag *irony*) che fungono da *gold standard*, ossia punto di riferimento, e quindi identificarne le caratteristiche salienti che li differenzino da altri commenti. Oltre a notare, anche nei testi scritti, la presenza di *intensificatori*, ossia aggettivi ed avverbi estremamente polarizzati (Barbieri e Saggion, 2014), sono interessanti le caratteristiche che sono peculiari del canale – testi scritti su social network. Reyes, Rosso e Veale (2013), ad esempio, individuano la caratteristica della *pointedness*, che si articola in una punteggiatura particolare (maggiore presenza di puntini di sospensione, punti esclamativi e di domanda, e virgolette), nell'uso di lettere maiuscole su alcune parole, e nel ricorso a emoticon. Simili alle emoticon sono le "espressioni di risata", come l'acronomimo LOL (*laugh out loud*), e le espressioni onomatopeiche come *ah*, *eh* e *ih* (Carvalho, Sarmento, Silva e De Oliveira, 2009)[4]. Si noti come questi stratagemmi possano essere visti come una trasposizione in forma scritta dei marcatori prosodici e gestuali/espressivi identificati per la conversazione faccia a faccia: l'uso delle maiuscole corrisponde all'enfasi derivata dall'allungamento di alcune parole, la punteggiatura segnala un'intonazione marcata, le emoticon e le espressioni di risata corrispondono a comportamenti gestuali come il fare l'occhiolino o la risata stessa.

È infine interessante osservare come si sia anche cercato di trovare degli indici che individuassero i cosiddetti fattori della ironia: l'incongruenza tra il messaggio e il contesto, e l'atteggiamento valutativo. Per quanto concerne il secondo fattore, studi all'interno del filone della *Sen-*

[4] In una ricerca pubblicata sul blog di Facebook (https://research.fb.com/the-not-so-universal-language-of-laughter/), Weinsberg, Adamic e Develin hanno analizzato i cosiddetti e-laughter, i marcatori di "risata elettronica", utilizzati dagli utenti di Facebook in una settimana del maggio 2015. Ben il 15% delle persone hanno inserito una e-risata nei loro post, con la netta maggioranza di trascrizioni di risate (51% di haha, e 13% di hehe), un terzo di emoticon, e solo il 2% di ricorso all'acronimo LOL.

timent analysis si sono focalizzati su commenti che contenessero parole valutative, connotate positivamente o negativamente. Per individuare invece la contraddizione che è alla base della interpretazione ironica, è stato proposto di prendere in considerazione recensioni sulla piattaforma di Amazon che attribuissero un voto basso (una o due stelline) a un prodotto, ma che al loro interno contenessero invece termini valutativi molto positivi (Davidov, Tsur e Rappoport, 2010); Bharti, Vachha, Pradhan, Babu e Jena (2016) propongono invece di creare un database che elenchi i fatti universalmente e contingentemente veri (prendendoli da enciclopedie e da titoli di giornali), nonché l'insieme di preferenze di uno specifico individuo – basandosi sui suoi *likes* e *dislikes*: i commenti postati da quell'individuo possono poi essere confrontati con questo insieme di fatti e di giudizi; quelli che risultano in palese contraddizione (ad es., "Sono andate proprio bene le elezioni al PD", piuttosto che "Salvini è un grande politico", se scritto da chi aveva messo *likes* a dichiarazioni di Saviano, e *dislikes* a quelle di Salvini stesso) verranno interpretati come sarcastici.

Abbiamo passato in rassegna una serie di studi volti a individuare le caratteristiche dei commenti ironici scritti su social network; il loro obiettivo è quello di trovare un algoritmo che riesca a riconoscere in maniera automatica i commenti sarcastici, e quindi comprendere se una recensione sia da intendere come positiva piuttosto che negativa, o anche etichettare le opinioni personali di chi scrive. Il nostro interesse verte invece sui meccanismi inferenziali che permettono a chi legge un commento sui social network di coglierne il senso inteso, e quindi comprendere – nel caso si tratti di un commento ironico – che il significato veicolato non è quello letterale: attraverso il riconoscimento dell'atteggiamento denigratorio espresso dall'ironista nei confronti di quel messaggio, l'interlocutore deve inferire che di fatto chi scrive pensa l'opposto. Dagli studi sopra elencati possiamo ricavare indicazioni utili sulla presenza di indizi che facilitano il riconoscimento dell'intento ironico: se si sa come stanno realmente le cose, o quali sono le opinioni personali di chi scrive, è possibile cogliere l'incongruenza di commenti ironici. Per quanto riguarda invece i marcatori dell'ironia, oltre alla presenza di indizi lessicali e sintattici che sono rintracciabili anche nelle lingue parlate, come l'iperbole o le domande retoriche, alcuni accorgimenti tipografici permettono di codificare informazioni analoghe a quelle di tipo prosodico (lettere maiuscole, punteggiatura) e corporeo-gestuale (emoticon).

Sebbene sia quindi possibile individuare dei segnali che facilitino il riconoscimento del tono sarcastico dei commenti scritti, tuttavia questi non sono sempre riconosciuti come tali. Nel prossimo paragrafo, analizzeremo alcuni casi di fallimento comunicativo, ossia dei casi in cui l'autore di un commento intendeva essere ironico, ma la sua intenzione non è stata colta, causando quindi un fraintendimento, e una interpreta-

zione letterale di quanto scritto. L'obiettivo di questa disamina è quello di riflettere sui motivi soggiacenti a questi fallimenti, per trarre delle considerazioni generali sulle motivazioni che spingono le persone a scegliere questa forma indiretta di comunicazione del loro pensiero. Concluderemo infine provando ad analizzare questi casi nell'ottica delle teorie ecoiche e della pretesa introdotte precedentemente.

5. Non riconoscere l'ironia nei social network

A marzo 2018 si sono svolte le elezioni politiche in Italia, che hanno segnato un successo elettorale del Movimento 5 Stelle e della Lega, e un crollo dei voti per il Partito Democratico. La disfatta è stata particolarmente pesante in alcune delle cosiddette roccaforti della sinistra, tra cui Pesaro, in cui il PD aveva schierato, per il collegio uninominale, l'uscente Ministro dell'Interno, Marco Minniti, il quale si è piazzato al terzo posto, dopo un candidato del Movimento 5 Stelle coinvolto in uno scandalo di rimborsi elettorali che viene quindi sospeso dal suo partito appena dopo l'elezione, e dopo la candidata del centrodestra. Poche ore dopo la diffusione dei risultati, i Giovani Democratici di Pesaro pubblicano sulla loro pagina Facebook quella che definiscono la loro analisi della sconfitta:

> Il PD pesarese alla camera durante le politiche del 2013 prese (fonte portale Eligendo) 18.763 voti.
> Ieri ha preso 14.875 voti.
> 3888 voti in meno. Che potrebbe sembrare un grosso problema se non considerassimo che il tasso di mortalità a Pesaro è 10 su 1000, cioè circa 950 persone l'anno.
> Non è che gli elettori di sinistra non c'hanno votato, è che sono morti.

Questo post attira subito grande attenzione, con molte persone che ne colgono l'intento ironico: non si tratta infatti di una "vera" analisi della sconfitta, nel senso che chi scrive non crede davvero che tutti i voti persi dal PD siano effettivamente da imputare alla morte dei precedenti elettori. Tuttavia, un alto numero di utenti sembra interpretarlo in senso letterale, con interventi che confermano, o obiettano alla plausibilità della dichiarazione che tutti gli elettori persi dal PD sono morti, affermazione che comunque viene accettata come una vera analisi della sconfitta. Il dibattito suscitato dal post viene anche ripreso da testate giornalistiche, e la "inaspettata visibilità" fa sì che i Giovani Democratici pesaresi decidano di aggiungere una nota alla fine del post originario:

> ++ Vista l'inaspettata visibilità forse è il caso di sottolineare l'ironia del post che voleva far riflettere su un problema generazionale di cui ora più che mai è necessario parlare. Ci scusiamo per non aver messo la faccina che ride. ++

Questo caso è interessante per almeno due motivi. In primo luogo, illustra la complessità del processo interpretativo soggiacente a casi di linguaggio non letterale: per "far riflettere su un problema generazionale", ossia che il PD non si rivolge ai giovani, gli autori del post enunciano una iperbole, tutti gli elettori persi dal PD sono morti, che non può essere vera nella sua interpretazione letterale, ma che comunque intende sottolineare che gli interlocutori a cui si era rivolto il PD sono anziani, e quindi almeno alcuni saranno morti rispetto alle elezioni precedenti. Ancora più interessante, ai fini della trattazione di questo elaborato, è però l'ultima frase della nota aggiunta al post originario, ossia "Ci scusiamo per non aver messo la faccina che ride". Questa affermazione è, a sua volta, ironica, come confermato da un post successivo dei Giovani Democratici di Pesaro, in cui hanno sentenziato che "Le emoticon sono la morte della comicità, crediamo fermamente in questa cosa e non ci piegheremo al loro utilizzo".

Come abbiamo discusso nella sezione precedente, la presenza di emoticon è stata individuata come una delle strategie che permettono il riconoscimento dell'intento ironico di quanto scritto, un equivalente di particolari comportamenti gestuali (il sorriso, lo strizzare l'occhio) rintracciabili nel corso delle conversazioni faccia a faccia. Il post originario dei Giovani Democratici di Pesaro non conteneva alcun chiaro indicatore di umorismo (anzi, la presenza di numeri precisi, con indicata la fonte ufficiale da cui erano tratti, sembrava confermare la serietà dell'analisi), ed effettivamente molti utenti lo hanno interpretato letteralmente. Questo fatto da un lato può avvalorare le preoccupazioni espresse da Kreuz, riassunte nel suo *Principle of Inferability*, che raccomanda a chi vuole fare ironia di segnalare, per mezzo dei marcatori dell'ironia, il suo intento sarcastico per evitare fraintendimenti e fallimenti comunicativi. Dall'altro lato, però, il fatto che gli autori del post frainteso rivendichino la loro scelta di non utilizzare segnali espliciti, come la faccina che ride, per guidare l'interpretazione del lettore, fa riflettere sulla funzione dell'ironia. Prima però di trarre delle conclusioni generali, prendiamo in considerazione un altro caso di ironia non colta.

Ad agosto 2017, Luca Bottura, giornalista e autore di satira, crea un meme, con la foto di Samuel Jackson e Magic Johnson, attore ed ex cestista, seduti su una panchina a Forte dei Marmi, dopo aver fatto shopping di lusso (si intravede una borsa di Louis Vitton), e il seguente testo:

RISORSE BOLDRINIANE A FORTE DEI MARMI FANNO SHOPPING DA PRADA COI 35 EURO CONDIVIDI SE SEI INDIGNATO!!!

La discrepanza tra il testo ("risorse boldriniane" fa riferimento ai migranti, per il sostentamento dei quali lo Stato pagava 35 euro al giorno) e l'immagine (due personaggi estremamente famosi, con vestiti di marca, a Forte dei Marmi) è evidente – e questo dovrebbe essere sufficiente

per cogliere l'intento ironico del post. E in effetti molti ne colgono il sarcasmo. Tuttavia alcuni lo condividono "indignati" (pensando quindi che si trattasse davvero di due migranti che utilizzavano in maniera fraudolenta i soldi messi a disposizione dallo Stato), o comunque ritengono che l'autore del meme davvero non avesse riconosciuto i due VIP. La situazione degenera ulteriormente quando Nina Moric, personaggio televisivo, rilancia il meme sulla sua pagina Facebook: la soubrette ne aveva colto l'intento sarcastico, ma dato il suo apprezzamento per le proposte di Casa Pound e la sua partecipazione al corteo contro lo *Ius soli*, una gran parte dei suoi followers ha interpretato il post come un vero messaggio di denuncia dello spreco delle risorse pubbliche. Anche in questo caso, il fraintendimento dell'ironia del messaggio aveva avuto grande eco, anche in testate internazionali.

Diversamente dal caso precedente, in cui la comprensione dell'intento umoristico richiedeva il realizzare che i Giovani Democratici non credevano davvero che gli elettori persi dal PD fossero semplicemente morti (e quindi il fare inferenze sulle credenze di chi scrive), in questo caso l'incongruenza tra l'interpretazione letterale del messaggio (questi sono migranti, e fanno shopping coi 35 euro dati dallo Stato) e il contesto (due VIP ben vestiti a Forte dei Marmi) avrebbe dovuto essere talmente chiara da non permettere il fraintendimento dell'intento ironico. Si noti come il testo che accompagnava l'immagine era scritto completamente in maiuscolo, e la frase sotto l'immagine ("condividi se sei indignato!!!") ricorre molto spesso in meme o post di denuncia di scandali o malefatte, che i "poteri forti" vorrebbero tenere nascosti. Lo stile, quindi, ricalcava quello di messaggi che circolano in determinati ambienti complottistici.

La breve disamina di questi due casi di fraintendimenti di post ironici nei social ha permesso di individuare due caratteristiche (la voluta assenza di emoticon come marcatori espliciti di ironia, e uno stile che scimmiottava veri post di denuncia) che possono fungere da chiave interpretativa di una questione fondamentale che riguarda il ricorso allo stile ironico, e che finora non abbiamo ancora affrontato, ossia la sua funzione, il *perché* un individuo dovrebbe volere comunicare il suo messaggio ricorrendo alla ironia.

6. Fraintendimenti fortuiti o intenzionali?

Come già sottolineato, diversamente da altre forme di linguaggio figurato, come la metafora, un commento ironico può avere una plausibile interpretazione letterale, e quindi essere frainteso. Inoltre, il riconoscimento del messaggio che si vuole comunicare per mezzo del sarcasmo è difficile, perché l'interlocutore deve non solo capire che il parlante (o lo scrivente) non crede nella verità di ciò che afferma, ma anche che sa che l'interlocutore è a conoscenza di questo fatto. Sarebbe

infatti necessaria una inferenza di secondo livello sugli stati mentali di chi parla (o scrive) per comprenderne l'intento ironico, e distinguere il sarcasmo dalla bugia (Sullivan, Winner e Hopfield, 1995) – e questo spiegherebbe perché la comprensione dell'ironia è un'acquisizione tardiva per i bambini a sviluppo tipico (Filippova e Astington, 2010), e risulta compromessa in popolazioni atipiche che mostrano ritardi nello sviluppo di abilità connesse alla Teoria della mente (Saban-Bezalel, Dolfin, Laor e Mashal, 2019; Winner, Brownell, Happé, Blum e Pincus, 1998). A questo punto ci si potrebbe anche chiedere perché mai una persona dovrebbe ricorrere all'ironia invece che enunciare direttamente ciò che intende dire.

La teoria ecoica di Wilson e Sperber (1992; 2012), e la teoria della pretesa di Clark e Gerrig (1984), come abbiamo visto, intendono rispondere proprio a questa domanda. Nel fare un commento ironico, l'ironista sta comunicando il suo atteggiamento derogatorio, di disapprovazione, nei confronti del pensiero espresso dal commento, o della persona che farebbe quel commento. La differenza principale tra questi due approcci consiste proprio nell'identificazione dell'oggetto dello scherno: il pensiero o la frase (di contenuto simile, non necessariamente uguale) che viene tacitamente attribuito a qualcuno per la teoria ecoica, e invece la persona che farebbe tale commento, e quelle che vi acconsentirebbero, per la teoria della pretesa.

Al fine di avvalorare la loro ipotesi, Clark e Gerrig riprendono la definizione di ironia presente nel Dizionario sull'uso dell'inglese moderno di Fowler (scritto inizialmente nel 1926, e poi rivisto più volte). Per Fowler, l'ironia postulerebbe un doppio uditorio: una parte di ascoltatori "non iniziati" ascolterebbe il commento senza rendersi conto che è ironico, e quindi interpretandolo letteralmente; l'altra parte, di "iniziati", invece sarebbe consapevole non solo della finzione dell'ironista, ma anche della mancata comprensione da parte degli altri ascoltatori, e il piacere nascerebbe proprio dal senso di intimità che si crea tra l'ironista e chi comprende il suo humour. Come poi riassunto da Clark e Gerrig, vittima dell'ironia diventa non solo la persona che si fa finta di essere dicendo qualcosa di assolutamente inappropriato rispetto alla situazione, ma anche chi non capisca che si tratta di una finzione, e creda nella verità del commento ironico. Il caso discusso da Winner (1997) è quello di una persona che dica "Non c'è assolutamente niente di male nel tagliare i fondi destinati ai più deboli per investire maggiormente nella difesa", avendo come interlocutori un conservatore e un progressista. Chi parla potrebbe volere che il progressista, di cui condivide le idee, riconosca il suo intento ironico, ma anche far sì che il conservatore non colga il sarcasmo, e interpreti il suo commento come sincero. In questo caso, il conservatore sarebbe doppiamente sbeffeggiato, sia perché le sue idee vengono presentate come sbagliate, sia perché non capirebbe neanche di essere preso in giro.

I casi di fallimenti comunicativi analizzati in precedenza possono quindi essere riletti alla luce di queste considerazioni, ipotizzando che gli autori dei post si stessero di fatto rivolgendo a un doppio uditorio: quello degli iniziati, che sono in grado di cogliere l'intento ironico, con i quali l'ironista solidarizza, e quello dei non-iniziati, talmente ottusi da interpretare letteralmente quanto scritto, senza capire di essere presi in giro, dai quali si distanzia. I Giovani Democratici di Pesaro volevano muovere una critica alle politiche del partito, che a loro avviso erano rivolte troppo verso la cosiddetta vecchia guardia di elettori, e ignoravano le istanze dei più giovani. Ed è stata probabilmente proprio la vecchia guardia a scandalizzarsi all'idea che l'analisi della sconfitta del PD si limitasse ad asserire che i voti persi erano di elettori morti. Il post di Bottura voleva prendere in giro chi si diceva indignato per lo spreco di risorse statali destinate al mantenimento di profughi e migranti, e proprio queste persone sono state quelle che lo hanno rilanciato credendolo un vero post di denuncia.

In altre parole, non si tratterebbe di veri casi di fallimento comunicativo, dovuto a un genuino fraintendimento, ma di provocazioni intenzionali, ossia di persone che hanno volutamente lanciato un messaggio che potesse essere colto come ironico solo dal loro gruppo di pari (l'in-group), e frainteso dal gruppo target dell'ironia (l'out-group). Indizi in questo senso sono da un lato il rivendicare la scelta di non avvalersi di emoticon per segnalare l'ironia, e dall'altro lato la scelta di utilizzare stili che sono caratteristici dei post tipicamente condivisi dal gruppo vittima dell'ironia (scritte in carattere maiuscolo, l'espressione "condividi se sei indignato!!!"). La teoria della pretesa di Clark e Gerrig riesce quindi a rendere adeguatamente conto di questi casi di ironia fatta per essere capita da alcuni, e fraintesa da altri, visto che si postula che l'oggetto dello scherno sia non solo chi è talmente ottuso da fare quel tipo di commento, ma anche l'uditorio che si trova d'accordo. La teoria ecoica, d'altro canto, asserisce che per mezzo dell'ironia si esprime un atteggiamento denigratorio nei confronti del pensiero espresso dal commento, e solo indirettamente della persona che formulerebbe quel pensiero, e soprattutto non prevede che l'ironista possa voler mascherare il suo intento ironico ai non iniziati.

Al tempo stesso, però, Sperber (1984) obietta a Clark e Gerrig che le caratteristiche di cui ci siamo qui serviti per interpretare i casi di (voluti) fraintendimenti non siano di fatto vere proprietà dell'ironia verbale *tout court*. In particolare, viene criticata l'idea che l'ironia abbia sempre delle vittime, e postuli un doppio uditorio: nel momento in cui esco in bicicletta per andare al lavoro, e a metà strada si buca una gomma, e inoltre comincia a piovere, io posso esclamare "Ma è proprio una splendida giornata!" senza avere in mente una vittima particolare, e senza voler sbeffeggiare qualcuno che crederebbe a quanto dico. Inoltre, Sperber ritiene che almeno alcuni dei casi analizzati da Clark

e Gerrig non si configurino come genuini casi di ironia, ma piuttosto di *parodia*. Secondo Sperber, pur avendo diversi elementi in comune (in entrambi i casi si fa riferimento a quanto detto, o potenzialmente pensato, da qualcuno), la parodia e la vera ironia sarebbero differenti: con la parodia si starebbe scimmiottando il proprio bersaglio, facendone un'imitazione caricaturale, mentre con l'ironia si starebbe esprimendo la propria disapprovazione verso un pensiero. Diventa interessante notare come, secondo Sperber, a queste due forme corrispondano due intonazioni diverse: quella parodistica imita, esasperandolo, il modo di parlare del proprio bersaglio, mentre quella ironica manifesterebbe più semplicemente scherno. Il post di Bottura effettivamente riprendeva alcuni elementi stilistici (l'abuso delle maiuscole e dei punti esclamativi, che abbiamo argomentato essere controparti scritte di elementi prosodici) che sono spesso utilizzati dalle persone che scrivono post di stampo razzista, le vittime che Bottura intende prendere di mira, e verrebbe quindi probabilmente letto da Sperber come scimmiottamento parodistico, e non ironia. Non ci interessa in questa sede disquisire se i casi discussi si configurino come casi di ironia piuttosto che di parodia (sicuramente non nel senso originario del termine), quanto piuttosto trarre delle conclusioni generali sulla questione se tali post, in cui chi scrive non stava sinceramente esprimendo le sue opinioni, siano stati equivocati da alcuni per questioni legate al mezzo (scrittura invece che conversazione faccia a faccia) che non permette di utilizzare alcuni marcatori prosodici e corporei dell'ironia, o se invece, come stiamo qui sostenendo, il fraintendimento sia stato cercato al fine di sbeffeggiare doppiamente le vittime dell'ironia. Riteniamo che la teoria della menzione di Clark e Gerrig possa render conto di questo meccanismo, in cui l'ironista decide di non rendere chiaramente manifesto il suo intento sarcastico (senza avvalersi quindi di marcatori espliciti di ironia) al fine di trarre in inganno alcuni, gli non-iniziati, rinsaldando inoltre un legame di intimità con gli iniziati che colgono la finzione.

7. Conclusioni

Abbiamo sostenuto che i casi analizzati non si configurino come genuini malintesi in cui l'intento ironico di chi scrive non è stato colto, ma che costituiscano invece delle provocazioni un cui l'ironista decide di rivolgersi a un doppio uditorio, al fine di creare un legame con il suo in-group, e di mostrare l'ottusità dell'out-group che casca nel tranello tesogli. Sebbene questa strategia possa rivelarsi effettivamente molto divertente per l'ironista, che riesce a beffare doppiamente il suo target, potrebbe essere anche pericolosa, in un certo senso, perché la vittima potrebbe anche non accorgersi del tutto di essere presa in giro, e rimanere convinta che la sua interpretazione letterale del commento ironico sia

corretta. Questo rischio è stato efficacemente illustrato nella Legge di Poe, uno degli assiomi di internet[5], che afferma che "Senza la faccina che ride o un altro evidente segnale di humor, è impossibile fare una parodia di un gruppo di fondamentalisti senza essere fraintesi ed essere interpretati letteralmente". In altre parole, è possibile che le persone che hanno interpretato il post dei Giovani Democratici di Pesaro come un'analisi seria della sconfitta elettorale, e quelli che hanno ri-condiviso il post con Samuel Jackson e Magic Johnson, indignati per lo spreco di risorse pubbliche, non si siano mai resi conto che si trattava di ironia, e siano quindi rimasti ancorati alla loro interpretazione erronea, consolidando le loro credenze.

Un ultimo caso può illustrare la situazione: ad agosto 2018, la nave Aquarius era rimasta al largo per diversi giorni, con le autorità maltesi e italiane che negavano lo sbarco nei loro porti. Nel pieno del dibattito sulla situazione dei 141 migranti a bordo, molti dei quali in fuga da guerre, Gian Marco Saolini ha pubblicato un video in cui si presentava come un marinaio licenziato dall'Aquarius, che testimoniava che i migranti a bordo erano felici, ben vestiti, e che passavano il tempo giocando a videogames e d'azzardo alla roulette. Si trattava di un falso, e Saolini afferma di aver diffuso il video con l'obiettivo di "mettere in luce il razzismo degli italiani". Tuttavia, quel video ha avuto, a quarantotto ore dalla pubblicazione, 4 milioni di visualizzazioni e 120.000 condivisioni. Quanti di questi utenti sono rimasti convinti che fosse vero?

Riferimenti bibliografici

Anolli, L., Ciceri, R., Infantino, M.G. (2002). From "blame by praise" to "praise by blame": Analysis of vocal patterns in ironic communication. *International Journal of Psychology*, *37*, 5, pp. 266-276.

Attardo, S. (2000). Irony markers and functions: Towards a goal-oriented theory of irony and its processing. *Rask*, *12*, 1, pp. 3-20.

Attardo, S., Eisterhold, J., Hay, J., Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor*, *16*, 2, pp. 243-260.

Barbieri, F., Saggion, H. (2014). Modelling irony in twitter. In *Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 56-64.

Bharti, S.K., Vachha, B., Pradhan, R.K., Babu, K.S., Jena, S.K. (2016). Sarcastic sentiment detection in tweets streamed in real time: a big data approach. *Digital Communications and Networks*, *2*, 3, pp. 108-121.

Bryant, G.A., Fox Tree, J.E. (2005). Is there an ironic tone of voice? *Language and speech*, *48*, 3, pp. 257-277.

[5] Si veda l'articolo di Chivers su *The Telegraph*, Internet rules and laws: the top 10, from Godwin to Poe, del 23 ottobre 2009, https://www.telegraph.co.uk/technology/news/6408927/Internet-rules-and-laws-the-top-10-from-Godwin-to-Poe.html.

Burgers, C.,van Mulken, M. (2017). Humor markers, in S. Attardo (a cura di), *The Routledge handbook of language and humor*. London-New York: Routledge, pp. 385-399.

Carvalho, P., Sarmento, L., Silva, M.J., De Oliveira, E. (2009). Clues for detecting irony in user-generated contents: oh...!! it's so easy ;-. In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*. Association for Computational Linguistics, pp. 53-56.

Chivers, T. (2009). Internet rules and laws: the top 10, from Godwin to Poe. *The Telegraph*, 23/10/2009, https://www.telegraph.co.uk/technology/news/6408927/Internet-rules-and-laws-the-top-10-from-Godwin-to-Poe.html (ultima consultazione 18/7/2019).

Clark, H.H., Gerrig, R.J. (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General*, *113*, 1, pp. 121-126.

Currie, G. (2006). Why irony is pretence. In S. Nichols (a cura di), *The architecture of the imagination*, Oxford: Oxford University Press, pp. 111-133.

Davidov, D., Tsur, O., Rappoport, A. (2010). Semi-supervised recognition of sarcastic sentences in twitter and amazon. In *Proceedings of the fourteenth conference on computational natural language learning*. Association for Computational Linguistics, pp. 107-116.

Ervas, F. (2011). Perché l'ironia riguarda il pensiero. *Esercizi Filosofici*, *6*, 2011, pp. 64-75.

Fahlman, S.E. (2002). Smiley Lore :-), http://www.cs.cmu.edu/~sef/sefSmiley.htm (ultima consultazione 18/7/2019).

Filippova, E., Astington, J.W. (2010). Children's Understanding of Social-Cognitive and Social-Communicative Aspects of Discourse Irony. *Child Development*, *81*, 3, pp. 913-928.

Fowler H.W. (1926). *A Dictionary of Modern English Usage*, Oxford: Oxford University Press.

Grice, H.P. (1975). Logic and conversation. In P. Cole (a cura di), *Syntax and Semantics Vol. 3: Speech Acts*. New York: Academic Press, pp. 41-58.

Grice, H.P. (1978). Further notes on logic and conversation. In P. Cole (a cura di), *Syntax and Semantics Vol. 9: Pragmatics*. New York: Academic Press, pp. 113-127.

Hootsuite & We Are Social (2019). Digital 2019: Italy, https://datareportal.com/reports/digital-2019-italy?rq=italy (ultima consultazione 18/7/2019).

Houston, K. (2013). *Shady characters: The secret life of punctuation, symbols, and other typographical marks*. New York: WW Norton & Company.

Kreuz, R.J. (1996). The use of verbal irony: Cues and constraints. In J.S. Mio e A.N. Katz (a cura di), *Metaphor: Implications and applications*. Mahwah, NJ: Lawrence Erlbaum, pp. 23-38.

Kreuz, R.J., Roberts, R.M. (1995). Two cues for verbal irony: Hyperbole and the ironic tone of voice. *Metaphor and symbol*, *10*, 1, pp. 21-31.

Kumon-Nakamura, S., Glucksberg, S., Brown, M. (1995). How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General*, *124*, 1, pp. 3-21.

Mantovan, L., Giustolisi, B., Panzeri, F. (2019). Signing something while meaning its opposite: The expression of irony in Italian Sign Language (LIS). *Journal of Pragmatics*, *142*, pp. 47-61.

Recanati F. (2005) Indexicality, context, and pretense: a speech-act theoretic

account. In N. Burton Roberts (a cura di), *Advances in Pragmatics*. London: Palgrave-Macmillan.
Reyes, A., Rosso, P., Veale, T. (2013). A multidimensional approach for detecting irony in twitter. *Language resources and evaluation*, *47*, 1, pp. 239-268.
Saban-Bezalel, R., Dolfin, D., Laor, N., Mashal, N. (2019). Irony comprehension and mentalizing ability in children with and without Autism Spectrum Disorder. *Research in Autism Spectrum Disorders*, *58*, pp. 30-38.
Seto, K.I. (1998). On non-echoic irony, in R. Carston e S. Uchida (a cura di), *Relevance theory: Applications and implications (Vol. 37)*. Amsterdam: John Benjamins Publishing, pp. 239-256.
Sperber, D. (1984). Verbal irony: Pretense or echoic mention? *Journal of Experimental Psychology: General*, *113*, 1, pp. 130-136.
Sperber, D., Wilson, D. (1981). Irony and the use-mention distinction. In P. Cole (a cura di), *Radical pragmatics*. New York: Academic Press, pp. 295-318.
Stalnaker, R. (2002). Common ground. *Linguistics and philosophy*, *25*, 5, pp. 701-721.
Sullivan, K., Winner, E., Hopfield, N. (1995). How children tell a lie from a joke: The role of second-order mental state attributions. *British Journal of Developmental Psychology*, *13*, 2, pp. 191-204.
Weinsberg, U., Adamic, L., Develin, M. (2015). The not-so-universal language of laughter, https://research.facebook.com/blog/the-not-so-universal-language-of-laughter (ultima consultazione 18/7/2019).
Wilson, D. (2006). The pragmatics of verbal irony: Echo or pretence? *Lingua*, *116*, 10, pp. 1722-1743.
Wilson, D., Sperber, D. (1992). On verbal irony. *Lingua*, *87*, 1, pp. 53-76.
Wilson, D., Sperber, D. (2012). Explaining irony, in *Meaning and relevance*. Cambridge (UK): Cambridge University Press, pp. 123-145.
Winner, E. (1997). *The point of words: Children's understanding of metaphor and irony*. Cambridge (MA): Harvard University Press.
Winner, E., Brownell, H., Happé, F., Blum, A., Pincus, D. (1998). Distinguishing lies from jokes: Theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and language*, *62*, 1, pp. 89-106.

Are you kidding me? (Not) detecting irony in social networks

Irony is a form of figurative language in which what is communicated does not correspond to the remark's literal meaning: the ironist intends to manifest her attitude of disapproval towards that message, from whose truth she, therefore, dissociates herself. Unlike other forms of non-literal language, such as metaphors or hyperboles, the ironic remark can have a plausible literal interpretation, and it can thus be misunderstood. To avoid communication failures, the ironist can therefore use particular meta-communicative clues, the so-called irony markers, to facilitate the identification of the intended meaning. We analyze the expression and the recognition of irony in social networks, aiming at identifying the irony markers in written texts, and at evaluating whether the risk of misunderstanding is higher than in face-to-face interactions. After reviewing the main theories on irony and irony markers, we discuss two cases in which the ironic intent of the posts has been misinterpreted. Our goal is to assess whether the

misunderstanding is due to the use of the medium, or whether it is intentional, as suggested by Clark and Gerrig's Pretense theory.

Keywords: irony, irony markers, social networks, communication failure.

Francesca Panzeri, Università degli Studi di Milano - Bicocca, Dipartimento di Psicologia, Piazza dell'Ateneo Nuovo 1, 20126 Milano, francesca.panzeri@unimib.it

BIANCA CEPOLLARO PAOLO LABINAZ

IDENTITÀ E LINGUAGGIO DISCRIMINATORIO NEI SOCIAL NETWORK

1. Introduzione

Con l'avvento dei servizi di social network, le possibilità di interazione tra individui in rete si sono allargate a dismisura. A partire da un profilo personale, che può essere pubblico, semi pubblico o anche privato, ciascun utente può condividere con singoli utenti, con la sua rete di amicizie o pubblicamente contenuti di vario formato e natura[1], siano essi testi, immagini, video o una combinazione di essi[2]. Ogni servizio di social network consente inoltre ai propri utenti di prendere posizione rispetto ai contenuti pubblicati da altri utenti o da pagine pubbliche; Facebook è particolarmente ricco in questo senso, dal momento che è possibile commentare un contenuto (e replicare direttamente ai commenti degli altri utenti con la funzione 'risposta'), esprimere (genericamente) il proprio apprezzamento grazie alla funzione 'Mi Piace' (Like) o emozioni di diverso tipo con le icone affiancate a essa (Love, Haha, Wow, Sigh,

Gli autori hanno collaborato nella ideazione e stesura dell'articolo. Tuttavia, i paragrafi 3 e 4 sono stati scritti da Bianca Cepollaro, mentre i paragrafi 1 e 2 da Paolo Labinaz.

[1] Un utente può condividere contenuti propri oppure rimandare (con un collegamento ipertestuale) a post di pagine o profili del sito di social network a cui è iscritto o di altri siti di social network (come ad esempio tweet o immagini postate su Instagram), o anche a contenuti multimediali presenti su social media, quali Youtube, Vimeo o Spotify, o più semplicemente alle pagine di qualunque altro sito presente in rete.

[2] Facebook è certamente il servizio di social network che offre le maggiori possibilità di interazioni nella condivisione di contenuti e per le discussioni in merito a essi sia sulle bacheche dei profili dei propri utenti sia sulle tante pagine e gruppi dedicati a tematiche di vario tipo. Altri siti di social network, come ad esempio Instagram e Twitter, sono più limitati negli spazi di condivisione e discussione, in quanto gli utenti hanno a diposizione solo la bacheca del proprio profilo personale e quelle degli utenti con cui possono interagire. Ciò dipende da fatto che questi servizi di social network sono stati progettati con scopi ben specifici: nel caso di Instagram, ad esempio, è la condivisione di fotografie personali, sebbene, nel tempo, a fronte della concorrenza agguerrita, si sia allargato anche ad altre funzioni.

Grrr)[3], o ancora condividerlo con l'aggiunta o meno di un commento proprio. Le frasi con cui si condivide un contenuto spesso chiariscono l'atteggiamento che chi posta esprime nei confronti di quanto condiviso: si può infatti condividere qualcosa in segno di supporto, ma anche di dissociazione, in modo derisorio, e così via. D'altra parte, ogniqualvolta si pubblica o condivide un qualche contenuto o si commentano contenuti postati da altri vi è sempre il rischio che il significato complessivo di quanto comunicato non sia compreso appieno, soprattutto quando ciò avviene, come nel caso di Facebook, su una pagina o un gruppo pubblico[4]. In questi contesti comunicativi virtuali, la platea a cui lo scrivente si rivolge, al di là di chi sia l'effettivo destinatario o gruppo di destinatari, è potenzialmente illimitata poiché quanto pubblicato è visibile a qualsiasi utente che abbia accesso a quella pagina o gruppo (almeno fino a quando la pagina o il gruppo non vengano eliminati, oppure gli amministratori di quella pagina o gruppo, o lo scrivente stesso, non decidano di cancellare il suo post o commento). È chiaro che i possibili fraintendimenti o incomprensioni circa quanto chi scrive ha inteso comunicare con il proprio post[5] o commento non dipendono dal numero potenzialmente illimitato di lettori e lettrici ma dall'alta probabilità che tra questi molti non possiedano informazioni condivise con chi scrive. Qui ci occuperemo di un caso particolare di questo problema generale: che riguarda le possibili conseguenze del mancato possesso di informazioni relative all'identità di chi scrive. Come vedremo, l'identità di chi pubblica un post o ne commenta uno è una fonte potenzialmente ricca di informazioni che possono contribuire a risolvere dubbi interpretativi che riguardano quanto comunicato, soprattutto quando si tratta di usi non letterali del linguaggio. Sotto il cappello di 'identità' possiamo fare rientrare una moltitudine di fattori che riguardano un individuo, dal suo aspetto, al suo genere, alla provenienza geografica, all'età, alla sua storia personale. Alcune di queste informazioni possono essere desunte dal nome, cognome o nickname con cui chi scrive è iscritto al social network. Un qualche tipo di nome infatti è sempre visibile agli altri utenti: il nome di battesimo può segnalare il genere di chi scrive, cognomi tipici di certe zone potrebbero indicare la sua provenienza geografica, un nickname

[3] L'icona 'Mi piace' e le icone affiancate a essa che esprimono diversi tipi di emozioni possono essere utilizzate per comunicare le proprie reazioni non solo a post e commenti, ma anche ai messaggi dei propri amici sul sistema di messaggistica istantanea di Facebook (noto come Facebook Messenger).

[4] Mentre nei gruppi solo gli utenti iscritti possono postare contenuti e fare commenti, ed è a discrezione degli amministratori di ciascun gruppo il fatto che gli utenti non iscritti possano leggere quanto viene postato sulla bacheca del loro gruppo, le pagine pubbliche sono visibili a tutti gli utenti di Facebook, sebbene in molto casi sia necessario mettere un 'Mi piace' alla pagina per poter commentare sulla bacheca.

[5] Il termine 'post', come utilizzato in questo lavoro, si riferisce (in senso lato) a tutte quelle funzioni dei servizi di sociali network che permettono di pubblicare contenuti, come ad esempio i tweet di Twitter.

può incorporare il suo nome oppure anche l'anno di nascita, pensiamo ad esempio a un nickname come 'Stefano_1975'. Di aiuto può essere anche l'immagine del profilo (se presente) o le informazioni relative a chi scrive presenti sul suo profilo, sebbene molto spesso tali informazioni siano visibili solo in maniera limitata per chi non fa parte della sua rete di amicizie. È possibile infine farsi qualche idea sull'identità di chi scrive anche osservando, tra le altre cose, il suo stile di scrittura, l'attenzione per il rispetto delle regole di netiquette[6] e l'uso delle emoticon. Insomma, stabilire l'identità di chi posta o commenta contenuti è un'attività per lo più indiziaria che non assicura l'univocità interpretativa degli indizi raccolti, né consente di verificare la loro veridicità.

Scopo di questo lavoro è esaminare quale sia il ruolo giocato dall'identità (o presunta tale) di chi scrive nel guidare i potenziali lettori e lettrici verso una corretta interpretazione di alcuni usi del linguaggio apparentemente discriminatori. Come vedremo, in questi casi l'identità di chi scrive può diventare un elemento cruciale nell'interpretazione di quanto scritto poiché perderla di vista può modificare radicalmente il significato complessivamente inteso e in certi casi persino mettere in discussione la legittimità del post pubblicato.

Presenteremo in primo luogo il quadro teorico entro cui ci muoveremo, evidenziando i vantaggi che esso offre nel rendere conto delle aspettative che guidano i processi interpretativi di certi usi del linguaggio non solo nella comunicazione faccia a faccia, ma anche in quella che avviene nelle interazioni 'anonime' tipiche dei social network. Esamineremo quindi alcuni casi di usi non denigratori degli slurs e di ironia e umorismo apparentemente discriminatori tratti da profili, pagine e gruppi di social network, quali Facebook e Twitter. Ci interrogheremo infine su come i social network complichino il ruolo giocato dall'identità di chi parla nell'interpretazione di questi usi del linguaggio e quale impatto ciò abbia sui sistemi di controllo e censura di post e commenti con contenuti apparentemente discriminatori.

2. Aspettative di pertinenza e usi del linguaggio nei social network

L'analisi che intendiamo condurre si muoverà all'interno del quadro teorico offerto dalla Teoria della Pertinenza (Sperber e Wilson, 1986/1995). Come messo in evidenza da Francisco Yus (2011), tale quadro teorico è

[6] Le regole di netiquette, come ad esempio quella secondo cui si deve evitare di scrivere parole o frasi con lettere maiuscole (in quanto questo tipo di carattere corrisponde al parlare ad alta voce, o peggio all'urlare), sono norme informali che disciplinano il buon comportamento degli utenti di *chat*, *mailing list* e altri spazi virtuali di comunicazione come i gruppi di Facebook. La loro violazione reiterata da parte di un utente può portare (a giudizio insindacabile degli amministratori) alla sua sospensione o cancellazione dalla *chat*, *mailing list* o gruppo dove tali violazioni sono avvenute.

particolarmente adatto allo studio della comunicazione mediata dalla rete. La Teoria della Pertinenza, nata dalla collaborazione tra l'antropologo e scienziato cognitivo Dan Sperber e la linguista Deirdre Wilson, propone un modello ostensivo-inferenziale della comunicazione, nel quale la comprensione linguistica è concepita come un un'attività inferenziale guidata dalla presunzione di pertinenza ottimale che accompagna ogni enunciato per il fatto stesso di essere stato proferito (Sperber e Wilson, 1986/1995, 50-54). Anche i processi interpretativi di post e commenti, siano essi composti da testi, immagini, video o una combinazione di essi, sembrano seguire un simile percorso. Quando un utente pubblica un post o ne commenta uno, di norma infatti intende sollecitare l'attenzione dei potenziali lettori, suggerendo che quanto comunicato può essere considerato di una certa pertinenza per loro. Di fatto, post e commenti vengono pubblicati proprio per essere fruibili, a seconda delle situazioni, da parte dei propri amici, degli utenti dei gruppi a cui si è iscritti o (solo nel caso dei commenti) di chi segue una certa pagina pubblica. La pubblicazione di un post o di un commento può essere equiparata quindi alla produzione di uno 'stimolo ostensivo': chi scrive non intende soltanto rendere manifesto un certo contenuto ai suoi potenziali lettori (intenzione informativa), ma anche rendere manifesto loro che ha intenzione di fare ciò (intenzione comunicativa). Così come gli enunciati proferiti in una conversazione faccia a faccia, anche post e commenti generano certe aspettative nei potenziali destinatari, portandoli a ritenere che i contenuti di tali post e commenti siano abbastanza pertinenti da valere la pena di essere elaborati. E, per i teorici della pertinenza, uno stimolo ostensivo vale la pena di essere elaborato poiché consente di ottenere un certo numero di effetti cognitivi positivi, che si concretizzano in cambiamenti che riguardano la nostra rappresentazione del mondo (Sperber e Wilson, 1986/1995, 118-122, 265-266).

A differenza tuttavia di quello che avviene nelle interazioni comunicative faccia a faccia, gli utenti dei social network si confrontano con un flusso continuo di stimoli ostensivi, come ben rappresentato graficamente dalla sezione News Feed presente sulla pagina home di tutti gli account di Facebook (ma sezioni simili si trovano anche su Twitter e Instagram). In questa sezione, gli utenti vengono continuamente aggiornati, sulla base dell'elaborazione di un algoritmo, circa i post pubblicati dagli utenti appartenenti alla loro rete di amicizie, quelli in cui sono taggati loro stessi o i loro amici, così come i post pubblicati nelle pagine che seguono e nei gruppi a cui sono iscritti. Si tratta chiaramente della sezione più 'vissuta' di Facebook, quella che ogni utente controlla più volte al giorno e da cui si fa almeno in parte guidare nel porre l'attenzione su certi post e commenti piuttosto che altri[7]. L'algoritmo del News Feed si sostituisce

[7] La selezione dei post da parte dell'algoritmo del News Feed avviene in base a criteri quali il numero di interazioni che l'utente ha con l'autore di quel post o com-

in un certo senso quindi agli utenti nel filtrare le informazioni in entrata, suggerendo che cosa può valere la pena di essere elaborato e che cosa no. È chiaro che si tratta di un'operazione di filtraggio preliminare: di tutte le notizie che scorrono nella sezione News Feed solo una piccola parte diviene effettivamente oggetto dell'attenzione di un utente e quindi elaborata. Questo dipende, tra le altre cose, dalle risorse cognitive richieste per l'elaborazione di quanto contenuto nel post o commento: uno stimolo è pertinente quanto più produce effetti cognitivi positivi con il minor sforzo di elaborazione (Sperber e Wilson, 1986/1995, 123-132). Come sostenuto dai teorici della pertinenza, le aspettative di pertinenza generate da uno stimolo ostensivo devono essere infatti sufficientemente precise per guidare il ricevente verso il significato complessivo inteso da chi l'ha prodotto. Questo significa che chi produce tale stimolo deve farlo in modo tale che sia abbastanza prevedibile da parte del ricevente quanto da lei o lui inteso. Le dinamiche appena descritte vengono amplificate nei contesti 'anonimi' tipici dei social network, quali bacheche di pagine e gruppi pubblici: in mancanza di un insieme di informazioni condivise, chi scrive un post o un commento deve formularlo in modo tale che i suoi potenziali lettori interpretino quanto comunicato con il minor sforzo possibile, in particolare senza che ciò comporti assunti contestuali difficilmente ricavabili da quanto scritto in quella data situazione. A partire dal modello teorico della procedura di comprensione basata sulla pertinenza (Sperber e Wilson, 2002), possiamo ipotizzare quindi che per interpretare il significato complessivamente inteso da chi scrive un post o commento chi legge deve formulare delle ipotesi a partire dagli indizi forniti dai contenuti disponibili, siano essi testi, immagini, video o una combinazione di essi, attraverso dei processi inferenziali che coinvolgono (per quanto possibile) anche assunti contestuali. In questa elaborazione, l'utente seguirà un percorso di minimo sforzo e si fermerà una volta raggiunta l'ipotesi interpretativa che soddisfa in maniera sufficiente la sua attesa di pertinenza ottimale.

Così come nelle interazioni comunicative faccia a faccia, anche in quelle 'virtuali' il compito interpretativo degli utenti è solitamente reso complesso dal fatto che ciò che chi scrive intende rendere manifesto può essere il proprio pensiero che un certo stato di cose (reale o concepibile) sia vero, oppure può richiamarsi al pensiero di qualcun altro, o ancora può esprimere il proprio atteggiamento nei confronti di un certo pensiero. Si tratta di usi del linguaggio che chiaramente differiscono nell'attesa di pertinenza che li accompagna e che chi legge deve riconoscere per poter arrivare all'interpretazione corretta di quanto comunicato. I teorici della

mento, l'interesse dimostrato precedentemente per un contenuto piuttosto che un altro (ad esempio, fotografie piuttosto che video), il numero di interazioni avute dal post in questione, la data di pubblicazione, sebbene post e commenti vecchi possano essere inseriti nella sezione News Feed quando ricevono nuove interazioni (si tratta del cosiddetto 'story bump') e così via.

pertinenza, in base alle differenze tra aspettative, distinguono in primo luogo tra usi descrittivi e attribuitivi (o, più in generale, interpretativi) (Sperber e Wilson, 1986/1995, 224-231; 2012, 128-129). Se per gli usi descrittivi l'attesa è che chi proferisce un certo enunciato intenda rappresentare un certo stato di cose del mondo, negli usi attributivi deve essere riconosciuto che chi parla intende evocare un certo contenuto per attribuirlo a qualcun altro, che può essere dissociato da sé (sia esso un singolo individuo o un gruppo di persone) o dal suo io attuale. Per esempio, quando Paolo dice qualcosa come (1), descrive uno stato di cose; quando dice qualcosa come (2), invece, attribuisce a chi lo ha proferito un pensiero o, in questo caso, un proferimento:

1. Piove a Londra.
2. Ho parlato con Bianca. Piove a Londra.

Mentre (1) è un uso descrittivo del linguaggio, (2) conta come uso attributivo: il contenuto 'Piove a Londra' è ascritto a Bianca. Oltre ad attribuire un contenuto a qualcun altro, si può, nell'atto stesso di attribuirlo, esprimere il proprio atteggiamento nei confronti del contenuto evocato. Siamo ancora di fronte a un uso attributivo del linguaggio, ma questa volta di tipo ecoico: si tratta di un uso attributivo del linguaggio in cui l'intenzione informativa di chi parla non è quella di informare sul contenuto evocato, ma di comunicare il proprio atteggiamento o reazione a esso. Ci si può richiamare a una frase, a un pensiero o a un'opinione. E lo si può fare per esprimere approvazione, perplessità, derisione, stupore e così via. Un caso molto studiato di uso ecoico del linguaggio, e di cui ci occuperemo nelle sezioni successive, è l'ironia (Sperber e Wilson, 1986/1995, 237-243; 2012, 123-145). Secondo i teorici della pertinenza, chi fa ironia non fa altro che dissociarsi o prendere le distanze dal contenuto evocato[8]. Per esempio, se Bianca proferisce (4), ma poi si presenta in ritardo a un appuntamento, il proferimento di Paolo in (5) è da considerarsi ironico: Paolo evoca il contenuto proferito da Bianca ed esprime al contempo il proprio atteggiamento di scherno.

3. Arriverò puntuale, come sempre.
4. Meno male che arrivi sempre puntuale.

Per interpretare il significato complessivamente inteso da Paolo chi ascolta deve formulare un'ipotesi circa quanto da lui comunicato di questo tipo:

5. Paolo crede che sia ridicolo pensare, come Bianca fa di sé, che ella sia sempre puntuale.

[8] Gli atteggiamenti a carattere dissociativo caratteristici dell'ironia coprono un ampio spettro che va dallo scorno fino al disprezzo.

Affinché chi ascolta arrivi a un'interpretazione di questo tipo, deve riconoscere l'uso attributivo di (4) da parte di Paolo e contemporaneamente l'atteggiamento di scherno che egli esprime verso l'enunciato precedentemente proferito da Bianca.

Gli usi ironici del linguaggio sono particolarmente interessanti perché per riconoscere l'ipotesi interpretativa che soddisfa la presunzione di pertinenza ottimale che li accompagna è necessario che l'uditorio possieda specifiche informazioni (nell'esempio sopra che Bianca abbia precedentemente affermato (4)) o, come diremo a breve, che siano presenti certi marcatori di ironia. In assenza di queste informazioni, l'ascoltatore potrebbe ritenere l'enunciato non abbastanza pertinente per essere preso in considerazione oppure potrebbe essere interpretato in senso letterale, dando luogo a incomprensioni o fraintendimenti. Se ciò può capitare in situazioni interattive faccia a faccia (e a ognuno di noi sarà capitato di fraintendere enunciati ironici o vedere fraintesi i propri), maggiore è la possibilità che tali incomprensioni o fraintendimenti avvengano nel caso delle interazioni 'anonime' tipiche dei social network, dove chi scrive ha un limitato spazio di manovra per segnalare l'uso non letterale (nel nostro caso ironico) che sta facendo di un certo enunciato.

3. Usi non letterali del linguaggio e social network

Un esempio paradigmatico di uso non letterale del linguaggio è l'ironia verbale, che può essere impiegata per molti scopi, tra cui quello umoristico. Umorismo e ironia non coincidono (vi sono casi di umorismo non ironico e di ironia non umoristica); tuttavia, qui non ci dilungheremo su come distinguere le due nozioni, ma ci interessiamo agli usi umoristici dell'ironia. Il fenomeno dell'ironia è estremamente complesso e in questo numero speciale Panzeri (2019) offre una ricca disamina degli aspetti che nella letteratura sono considerati marcatori di ironia; fra i tratti che guidano chi ascolta nel riconoscere un certo proferimento come ironico, vi sono la prosodia, le espressioni facciali (la cosiddetta 'blank face'), l'utilizzo di espressioni iperboliche o estreme, e così via (si vedano, tra gli altri, Attardo, Einsterhold, Hay e Poggi, 2003). L'interesse di Panzeri è individuare quali marcatori di ironia possano essere impiegati nel riconoscimento dell'ironia online, compito interpretativo difficile per cui si rende necessario integrare contemporaneamente molti fattori. In questo contributo, invece, il nostro interesse si concentra su un particolare tipo di informazione che guida l'interpretazione dell'ironia, cioè l'identità di chi parla; il caso specifico su cui ci soffermeremo è come l'identità di chi parla influenza il modo in cui leggiamo alcune forme di ironia e umorismo all'apparenza discriminatorie. Prima di considerare il caso particolare degli usi discriminatori del linguaggio, due parole sul ruolo dell'identità del parlante nell'interpretazione di enunciati ironici

in generale. Prendiamo un celebre aneddoto su Sigmund Freud riportato dal figlio in Freud (1957, 217): poco dopo l'Anschluss – l'annessione dell'Austria alla Germania nazista (1938) – Sigmund Freud è costretto a lasciare il Paese per trovare rifugio in Inghilterra; prima di partire, la Gestapo lo obbliga a sottoscrivere un testo in cui dichiara di essere stato trattato col massimo rispetto dalle autorità tedesche; pare che Freud, dopo aver firmato, abbia voluto aggiungere un ulteriore commento "Ich kann die Gestapo jedermann auf das beste empfehlen", cioè "Posso consigliare vivamente la Gestapo a chiunque". Sapere chi sia Freud – conoscere le sue origini ebraiche, ma anche la forte ostilità del Nazismo alle sue tesi e ai suoi studi – permette di interpretare con facilità l'enunciato come ironico: costretto a firmare la dichiarazione per poter lasciare l'Austria e mettersi in salvo, Freud segnala il proprio dissenso e la propria prese di distanze inserendo nel testo che sottoscrive una frase di apprezzamento della Gestapo palesemente ironica. In questo caso, l'identità del parlante è un forte indizio per riconoscere il carattere beffardo di questo uso linguistico (oltre naturalmente ad alcune conoscenze di sfondo, tra cui la condotta e la pericolosità della Gestapo). È facile immaginare che la stessa frase possa essere stata pronunciata del tutto seriamente da un/a simpatizzante della Gestapo ed è proprio di questa prospettiva che Freud si fa beffa. L'identità di chi parla è una fonte potenzialmente molto ricca di informazioni che contribuiscono a cogliere l'eventuale ironia degli usi linguistici.

Oltre alla questione del *riconoscimento* dell'ironia – inteso proprio come la capacità di cogliere che un dato enunciato è usato ironicamente – vi è poi quella della *legittimità*, specie nei casi di linguaggio discriminatorio. Che non tutti possano fare ironia su qualunque argomento è quasi proverbiale. Nell'ambito della comicità è noto come spesso battute che colpiscono temi socialmente delicati quali razzismo, omofobia e sessismo siano più facilmente consentite a chi appartiene al gruppo colpito. La cornice comica chiarirebbe già di per sé che si è in un ambito di usi non seri del linguaggio, ma pur *riconoscendo* che chi parla *vorrebbe* essere ironico o umoristico, spesso questo non basta ad attribuire la *legittimità* della battuta. Esiste infatti una dimensione morale rispetto a cui certi usi non seri del linguaggio come ironia e umorismo possono essere riconosciuti come moralmente accettabili[9] o meno; l'identità di chi parla è uno dei fattori che giocano un ruolo in questa complessa negoziazione. Prendiamo *Baby Cobra*, celebre show Netflix della comica Ali Wong, statunitense di origine asiatiche, i cui spezzoni vengono spesso condivisi e commentati sui social[10]. Lo spettacolo è un'aspra critica alla

[9] Nel dibattito sull'umorismo, si discute se l'immoralità di una battuta interferisca o meno con la sua comicità e se lo fa, in che direzione (aumentando o diminuendo l'effetto comico). Si veda in proposito la disamina offerta in Anderson (2015).

[10] All'11/09/19, troviamo spezzoni del suo spettacolo condivisi e sottotitolati in molte lingue; solo per fare un esempio, la versione sottotitolata in inglese (www.face-

misoginia di certi ambienti lavorativi, incluso quello dei commedianti: la comica, incinta di sette mesi e mezzo al momento della performance, apre lo show con un pezzo sul fatto che non si vedono molto spesso delle comiche incinte; anzi, non si vedono proprio. E questo perché mentre l'arrivo di un figlio non danneggia ma anzi incoraggia la carriera di un comico, segna troppo spesso la fine di quella di una comica. D'altro canto, *Baby Cobra* include anche battute all'apparenza sfacciatamente antifemministe, come: "I think feminism is the worst thing that ever happened to women" "Our job used to be no job" "We had it so good!... And then all these women had to show off and say, 'We could do it; we could do anything!... They ruined it for us!" (trad. it. nostra: "Penso che il femminismo sia la cosa peggiore che sia mai capitata alle donne. Il nostro lavoro era non avere un lavoro. Eravamo messe bene! E poi queste donne hanno dovuto mettersi in mostra e dire 'Possiamo farlo; possiamo fare qualunque cosa!'"). L'umorismo di Wong è caratterizzato (anche) da battute apparentemente antifemministe di questo genere, ma la sua storia di comica in un mondo lavorativo prevalentemente maschile permette di leggere questo pezzo comico come ironico e di interpretarlo come, di fatto, *non* misogino: *Baby Cobra* evoca un pensiero radicato in una certa parte della società per esprimere al contempo un'istanza critica. Analogamente, la performance di Wong include battute razziste, spesso contro gli asiatici, come: "But I think that for marriage it can be nice to be with someone of your own race. The advantage is that you can go home and be racist together. You get to say whatever you like" (tr. it. nostra: "Ma penso che per quanto riguarda il matrimonio, può essere carino stare con qualcuno della tua stessa razza. Il vantaggio è che puoi tornare a casa ed essere razzisti insieme. Puoi dire tutto quello che vuoi"). Ancora una volta, il fatto che Ali Wong sia un'americana asiatica e abbia vissuto sulla propria pelle cosa vuol dire fare parte di una minoranza etnica rende chiaro che lo guardo benevolo nei confronti del razzismo è una presa in giro. Le battute apparentemente razziste di Ali Wong non hanno lo stesso status morale di una battuta razzista qualunque, anzi. L'identità di questa artista – il suo genere, il suo mestiere, le sue origini etniche – chiarisce che la prospettiva discriminatoria delle sue battute non è sul serio abbracciata, ma è solo evocata per diventare il vero oggetto dello scherno e permette di attribuire uno status morale diverso da quello di una battuta genuinamente razzista o misogina.

Una possibile interpretazione di questo fenomeno è sostenere che nel caso di usi non letterali del linguaggio come ironia e umorismo, esiste

book.com/netflixcanada/videos/1253678437976553/) ha accumulato 95000 reazioni, 42000 commenti, 175000; la versione sottotitolata in portoghese (www.facebook.com/watch/?v=1167096710013787) ha accumulato circa 37000 reazioni e 16000 commenti, con 3300000 visualizzazioni; la versione sottotitolata in francese (www.facebook.com/netflixfrance/videos/1036548976427351/) ha ricevuto circa 6500 reazioni, 714 commenti, 7100 condivisioni e 389000 visualizzazioni e così via.

un certo margine di fraintendimento circa cosa intenda chi parla. Quando si evocano, scherzando, prospettive discriminatorie, lo si può fare con un grado di dissociazione dalle stesse molto variabile. L'identità di chi parla può in questi frangenti fungere da *garanzia* che l'atteggiamento razzista/misogino/omofobo evocato non sia neppure in minima parte appoggiato, ma sia anzi l'oggetto dello scherno. Ciò non vuol dire che chi appartiene alla categoria derisa abbia un lasciapassare che permette di fare impunemente dell'umorismo discriminatorio, né che questo genere di umorismo sia consentito *solo* ai membri del gruppo in questione. Piuttosto, diremo che alcuni aspetti dell'identità dei parlanti, come l'appartenenza alla categoria oggetto dell'ironia o della derisione, costituiscono un forte *indizio* che chi parla non abbraccia davvero la prospettiva discriminatoria evocata. In altre parole, è certamente possibile che individui che *non* fan parte del gruppo colpito da una battuta si dissocino ironicamente dall'atteggiamento discriminatorio che essa evoca, ma l'appartenenza alla categoria in questione è un indizio particolarmente efficace. L'identità di chi parla in questi casi dissipa possibili dubbi sulla sua buona fede e contribuisce a determinare non solo l'interpretazione dell'enunciato ma anche il suo status morale.

Riconoscere l'ironia e l'umorismo online presenta due ordini di complessità aggiuntivi. In primo luogo, come illustrato molto bene da Panzeri (2019), l'espressione scritta non consente di cogliere gli indizi che caratterizzano l'ironia in modo più prototipico (espressioni facciali e prosodia); in secondo luogo – e questo è l'aspetto su cui qui ci vorremmo soffermare – i social network possono rendere meno chiara l'identificazione di chi parla, o, in questo caso, di chi ironizza. Ci sono almeno due sensi in cui questo avviene. Da un lato, post, tweet o condivisioni sono ascrivibili a un *account* e non propriamente a un individuo: un profilo può facilmente camuffare nome, volto, e storia della persona che effettivamente fa uso dell'account; d'altro canto, specie nel caso delle pagine, un account può appartenere a un collettivo di persone o a un personaggio finzionale.

Dall'altro lato, quando un contenuto viene pubblicato, i meccanismi di condivisione possono far perdere di vista chi in origine aveva usato gli enunciati in questione ironicamente; in molti casi, questo non importa e non snatura il contenuto; in altri, l'identità di chi parla è un elemento cruciale nell'interpretazione del contributo e perderla di vista modifica radicalmente il senso di ciò che è comunicato. Una delle critiche sollevate, ad esempio, allo spettacolo di Ali Wong – i cui spezzoni, come si diceva, sono stati molto condivisi e commentati sui social – è proprio che presti il fianco all'antifemminismo. Per chi conosce il lavoro di Wong, la critica è non solo infondata ma anzi mal riposta. Tuttavia, la questione è proprio che per interpretare correttamente il pezzo comico è necessario avere un insieme molto ricco di informazioni su chi parla. Inoltre, spesso sui social vengono diffusi solo dei frammenti di un pezzo

comico e questo mette ulteriormente a rischio l'impresa di una corretta interpretazione. Si pensi ad esempio al fatto che su Twitter è stato diffuso uno stralcio di Baby Cobra al grido di 'Feminism is bad'[11]: difficile capire se ironicamente o no. In alcuni commenti su Twitter all'annuncio di un ritorno di Wong con un nuovo spettacolo (Wong appariva nel video nuovamente incinta), un utente ha lasciato un commento ironico, sentendosi in dovere di chiarire l'intento ironico con una postilla: "Surely she must be carrying the spawn of Satan Himself, for who else will fornicate with her. She is the one who will finally birth the Antichrist & bring about the Apocalypse! The End is Upon Us. Repent Ye None Believers. Yes, this was meant to be satirical." (trad. it: "Porterà di certo in grembo la prole di Satana in persona, perché chi altro mai potrebbe fornicare con lei. È lei che darà infine la luce all'anticristo e causerà l'Apocalisse. La fine è alle porte. Pentitevi non credenti. Sì, questo voleva essere satirico").

Per concludere, nonostante in alcuni casi interpretare correttamente l'ironia sui social sia semplice, spesso l'identità di chi parla – intesa in senso ampio – è una chiave di lettura imprescindibile. Nel prossimo paragrafo discuteremo del perché interpretare correttamente certi usi apparentemente discriminatori del linguaggio sia così importante per i social network.

3.1. *Usi riappropriativi degli slurs: dimmi chi sei e ti dirò cosa puoi dire*

La diffusione dei social network, con la possibilità di una comunicazione a diffusione 'virale', ha reso particolarmente cogente la necessità di identificare il cosiddetto discorso d'odio (o hate speech). Per 'hate speech' si intende l'insieme delle pratiche linguistiche che colpiscono gli individui sulla base di caratteristiche quali la nazionalità, l'origine geografica o etnica, l'identità di genere, l'orientamento sessuale, la fede religiosa e così via. Il discorso d'odio può prendere forme piuttosto variegate e di non sempre banale individuazione. L'esempio più prototipico consiste nei cosiddetti slurs o epiteti denigratori: termini quali 'negro', 'frocio', 'terrone', e così via, che colpiscono persone e gruppi sulla base della loro appartenenza a una certa categoria sociale. Se gli slurs costituiscano una classe linguistica vera e propria è una questione dibattuta (si vedano Bolinger, 2017; Nunberg, 2018; Stanley, 2015; Pullum, 2018; ma anche le evidenze sperimentali presentate in Cepollaro, Sulpizio e Bianchi, 2019). Filosofi del linguaggio e linguisti hanno proposto vari approcci[12] per spiegare come questo genere di

[11] twitter.com/LynseyAmanda83/status/1101426212617052161. Ultima consultazione: 29/07/19.

[12] Per un approccio vero-condizionale, si veda: Hom, 2008; Hom e May, 2013; 2018; Diaz-Leon, 2019; per un'analisi in termini di implicature convenzionali: Potts, 2005; 2007;

espressioni veicoli contenuti dispregiativi. In questa indagine, hanno concentrato la propria attenzione sugli usi più tipici di questi termini, ovvero quelli denigratori. Tuttavia, un'osservazione a più ampio raggio rivela che queste espressioni – considerate l'esempio emblematico di hate speech – sono anche usate in modo solidale e positivo. Si tratta del fenomeno della riappropriazione, per cui alcuni soggetti (tipicamente, ma non necessariamente, membri del gruppo target) si impossessano di una locuzione usata per umiliare e svilire una categoria di individui e l'utilizzano in un modo nuovo e non convenzionale, privandola della sua accezione negativa. In ambito anglofono, i casi più noti di riappropriazione sono quelli dello slur 'nigger' (specie nella sua variante 'nigga') e di 'queer'. Gli usi riappropriativi interessano inizialmente una nicchia ristretta di persone (solitamente, membri del gruppo target), ma si possono gradualmente diffondere. Uno dei possibili effetti della riappropriazione è indebolire la forza offensiva degli epiteti, fino anche a farla scomparire, come è accaduto per 'gay' in inglese (Brontsema, 2004; Bianchi, 2014): l'uso positivo ha poco a poco soppiantato quello negativo, fino a far diventare 'gay' il modo standard in inglese (e non solo) per riferirsi alle persone omosessuali.

Bianchi (2014) si rifà alla teoria ecoica dell'ironia per analizzare gli usi riappropriativi degli epiteti denigratori: chi usa uno slur in modo riappropriativo evoca la prospettiva discriminatoria di chi impiega tipicamente quei termini e ne prende le distanze esprimendo il proprio atteggiamento dissociativo. I social network possono oggi ricoprire un ruolo importante nella diffusione rapida e capillare di nuovi usi non convenzionali del linguaggio, ma amplificano anche alcune difficoltà. Osservando una piattaforma social come Facebook, si notano molti usi riappropriativi di slurs. Limitandosi solo ai nomi, troviamo gruppi come 'Terroni al Nord'[13], 'Terroni a Verona', 'Terroni a Padova' e così via: si adotta il termine usato per offendere e lo si utilizza con orgoglio, sbeffeggiando la discriminazione. Nella descrizione del gruppo 'Terroni al Nord', ad esempio, si legge "Questo gruppo è stato creato per i meridionali che sono emigrati al nord Italia, per scambiare opinioni... darci dei consigli su dove trovare ristoranti dove si mangia bene oppure dove si mangiano cose tipiche del sud che a noi fanno gola. [...]". Altri gruppi hanno invece una natura più fortemente politica: nella descrizione del gruppo 'Terroni'[14], legato al giornalista e scrittore Pino Aprile, leggiamo "A cosa serve questa pagina? A incontrarci e concentrarci. [...] Cercando di riunirli in

McCready, 2010; per un approccio presupposizionale, si veda: Macià, 2002; Schlenker, 2007; Cepollaro, 2015; Cepollaro e Stojanovic, 2016; per degli approcci pragmatici, secondo cui il contenuto denigratorio degli epiteti non è parte del significato lessicale, si vedano: Anderson e Lepore, 2013; Bolinger, 2017; Nunberg, 2018; Rappaport, 2019.

[13] www.facebook.com/groups/260040218130060/. Ultima consultazione: 29/07/19.

[14] www.facebook.com/terronidipinoaprile/. Ultima consultazione: 29/07/19.

una sola pagina e riempendo questa pagina di contenuti che alle ragioni del nostro incontro si rifanno (la denuncia delle discriminazioni a danno del Sud, ma non solo; la ricerca delle vie per eliminarle; l'informazione su chi fa cosa, a questo fine)". In entrambi i casi, il termine 'terrone' viene rivendicato con orgoglio; nella prospettiva di Bianchi (2014), questa riappropriazione passa dall'evocazione della prospettiva razzista, accompagnata dall'espressione della propria distanza. Si noti per altro che l'insulto antimeridionalista non è affatto innocuo oggi. Molti casi di cronaca riportano multe salatissime per chi ha usato questo epiteto per denigrare qualcuno: uno dei più recenti risale all'aprile 2019, quando un autista di un bus urbano trevigiano ha insultato due turiste meridionali chiamandole 'terrone', ottenendo una sospensione e decurtamento dello stipendio (l'autista si era per altro difeso dicendo di aver voluto solo scherzare, ma questa giustificazione non è stata accettata dai suoi datori di lavoro)[15]. Possiamo quindi scartare l'ipotesi per cui l'uso positivo di 'terrone' sui social rifletta il fatto che il termine antimeridionalista non sia ormai più percepito come offensivo.

Ciò che distingue gli usi riappropriativi degli slurs sui social da quelli nelle interazioni faccia a faccia è la difficoltà di stabilire l'identità di chi parla, che pure gioca un ruolo tanto centrale nella possibilità di usare un termine denigratorio in modo non offensivo. Prendiamo ad esempio il post[16] dell'1 maggio 2017 di Carlo Giuseppe Gabardini, l'attore, comico e scrittore noto per aver interpretato Olmo in *Camera Caffè* (di cui era anche autore). Gabardini racconta su Facebook un episodio comico che ruota attorno all'acquisto di prodotti cosmetici assai costosi nel Marais, a Parigi, orchestrato da un abile commesso. Parte del racconto è un dialogo virgolettato tra lo scrittore e il suo compagno: "Ah dici che ha capito che siamo ricchioni?" "Siamo al Marais, non credo sia una novità esotica, non qui". L'episodio prosegue, ma quel che ci interessa è che quando Gabardini usa sui social questo tipo di linguaggio, omofobo per eccellenza, è lampante che si tratta di un uso non convenzionale, che sovverte le aspettative. Gli elementi che consentono di interpretare l'uso di 'ricchione' come positivo sono molteplici. Il fattore principale è la consapevolezza che l'autore del post è molto attivo nella difesa e promozione dei diritti delle persone LGBTQ+. Un dettaglio aggiuntivo che può cogliere chi conosce il lavoro dell'artista è che nel momento in cui pubblica il post, l'immagine del profilo ritrae l'attore mentre mangia una fetta di pane e Nutella, un chiaro riferimento al video "La marmellata e la Nutella (ci si innamora di chi ci s'innamora)"[17], un monologo di

[15] Per la notizia si veda: www.trevisotoday.it/attualita/multa-autista-mom-terrone-treviso-12-aprile-2019.html. Ultima consultazione: 29/07/19.

[16] www.facebook.com/carlogabardini/photos/a.479651715467727/1038627826236777/?type=3&theater. Ultima consultazione: 29/07/19.

[17] www.youtube.com/watch?v=axdRn-ai758. Ultima consultazione: 29/07/19.

Gabardini che racconta l'omosessualità in una società eternonormativa attraverso la metafora dei gusti culinari. Per chi invece non conoscesse approfonditamente il personaggio e i suoi lavori, basta il contesto conversazionale in cui l'uso dell'epiteto si inserisce: nel post si fa esplicito riferimento al fatto che l'autore abbia un compagno e questo spesso spinge a interpretare l'uso di un epiteto omofobo come non offensivo.

La difficoltà nell'interpretare come riappropriativi questi usi del linguaggio è resa ancora più acuta quando questi non appaiono sulla pagina di un personaggio (come nel caso di Carlo Giuseppe Gabardini), ma di un gruppo, la cui identità è molto più sfuggente. Prendiamo il caso dei gruppi Facebook: non esiste la certezza di chi siano le persone che gestiscono concretamente la pagina scrivendo i post. Nel caso del gruppo 'Il Terrone Fuori Sede', per esempio, non si ha la certezza che chi scrive i post e compone i cosiddetti meme sia davvero una persona meridionale; è probabile, ma non è detto, e non è tanto la provenienza geografica effettiva di chi scrive quanto piuttosto l'appartenenza a un gruppo che, come si legge nella descrizione, "porta ovunque con ironia la propria meridionalità" a guidare l'interpretazione riappropriativa dell'uso di epiteti come appunto 'terrone'.

3.2. *Usi prospettici degli slurs*

Anche sui social network si possono poi trovare degli usi non denigratori degli epiteti che non si caratterizzano, come i casi considerati fin qui, per l'*orgoglio* con cui un termine viene usato; piuttosto, si tratta di usi 'prospettici', in cui chi parla prende il punto di vista di un razzista per parlarne o prenderne le distanze (cfr. Jeshion, in corso di stampa). Questi usi vanno interpretati come attributivi, poiché i contenuti discriminatori sono ascritti a qualcuno di diverso da chi sta effettivamente proferendo gli enunciati. Sembra anzi che chi parla prenda le distanze dalla prospettiva discriminatoria evocata; se così fosse, questi conterebbero a tutti gli effetti come usi ironici degli epiteti.

Consideriamo due casi che ci paiono interessanti: il primo è un tweet[18] dell'11 gennaio 2019 di un giovane scrittore calabrese che commenta così una notizia – un ospedale di Crotone che ospita alcune mamme curde con bambini e neonati ha dovuto chiedere alla cittadinanza di bloccare le donazioni di abiti e simili da cui sono stati sommersi –: "Il tipico #buonismo dei #terroni! <3". Due giorni dopo, il gruppo 'I Sentinelli di Milano', pubblica lo screenshot[19] del tweet col commento

[18] twitter.com/Gaepanz/status/1083675488608899072. Ultima consultazione: 29/07/19.

[19] www.facebook.com/isentinellidimilano/photos/p.1113070115542074/1113070115542074/?type=1. Ultima consultazione: 29/07/19.

'Ma tu guarda sti terroni'. I Sentinelli sono un movimento nato proprio su Facebook in opposizione alle 'Sentinelle in piedi'; questi ultimi nascono sulla scorta di un'iniziativa francese del 2013 che protesta contro la legge del 'mariage pour tous'[20]; al contrario, i Sentinelli di Milano vengono fondati da una coppia omosessuale meneghina con l'iniziale scopo di difendere e promuovere i diritti delle persone LGBTQ+, poi impegnata su più fronti contro molti generi di discriminazione. Il post dei Sentinelli loda l'iniziativa solidale dei cittadini di Crotone, ma lo fa evocando la prospettiva antimeridionalista tramite l'adozione del suo stesso lessico. È importante qua notare che chi parla non è un cosiddetto in-group (cioè un meridionale). Mentre l'autore del tweet postato dai Sentinelli dichiara nel profilo Twitter di essere calabrese, la situazione dei Sentinelli è molto diversa. Per quanto non sia immediato risalire con precisione a una persona specifica, la pagina si caratterizza geograficamente come nordica: I Sentinelli *di Milano*. Nella letteratura sugli usi non offensivi degli epiteti vi è una forte insistenza sul fatto che appartenere al gruppo target sia fondamentale per poter impiegare uno slur in modo non denigratorio; qui vediamo come questo sia lungi dall'essere una regola inviolabile: proprio una pagina che si caratterizza come nordica fin dal nome si permette di pubblicare un post in cui occorre 'terrone'. Ciò può succedere perché l'identità di chi parla non si riduce all'appartenenza al gruppo target, ma include moltissime dimensioni, ben più complesse. Nel caso dei Sentinelli di Milano, l'impegno nella lotta alla discriminazione – attraverso la promozione di laicità, antirazzismo e antifascismo – sembra permettere a chi gestisce questa pagina di poter usare in modo non convenzionale e positivo quello dovrebbe essere per eccellenza un esempio di linguaggio d'odio. Si noti tuttavia che non manca chi non coglie il modo in cui l'epiteto è usato; ecco uno scambio tra gli utenti che hanno commentato nel gennaio 2019 (abbiamo sostituito i loro nomi con '*A*' e '*B*'): *A*: Ma quelli che hanno messo la faccia che ride, che problemi hanno?; *B*: Hanno probabilmente capito la battuta sul buonismo dei terroni e sorriso di conseguenza. *A*: Intanto quelli che tu chiami terroni, sono persone che hanno salvato 51 vite. Mi pare che ci sia poco da ridere; *B*: 1) lo so; 2) sono terrone anch'io; 3) ne io ne l'autore del tweet abbiamo mai avuto l'intenzione di offendere nessuno; 4) non sono io ad averli chiamato terroni ma l'autore del tweet; 5) ERA UNA BATTUTA: se non l'hai capita te la si può spiegare facilmente, basta chiedere. ☺ ; *A*: probabilmente non ho capito che era una battuta.

Il secondo caso è un esempio di uso finzionale di uno slur in cui lo scopo con cui gli epiteti denigratori sono usati non è solo di mettere in scena il discorso di un razzista, ma quello di criticarne l'ideologia. Si tratta di un video intitolato "Quando i neri erano i meridionali: ovvero,

[20] www.gouvernement.fr/action/le-mariage-pour-tous. Ultima consultazione: 29/07/19.

l'ultimo è 'il più terrone' di tutti"[21], pubblicato sulla Pagina Facebook 'This is Racism'[22] il 19 ottobre 2018. Nel video, un attore impersona un antimeridionalista che ricorda il disgusto provato per i meridionali; ma dopo essersi scagliato contro quelli che identifica come 'terroni', passa a parlare dell'arrivo di migranti neri in Italia, che avrebbe compiuto quel miracolo che neppure a Cavour era riuscito: creare gli Italiani, in opposizioni agli stranieri ("dopo 300 anni ci siam scoperti tutti fratelli nel dare addosso al negro", tr. nostra dal dialetto veneto). Oltre a usare un linguaggio fortemente discriminatorio ('terroni' e 'negri' abbondano), chi parla umilia i meridionali storpiandone le cadenze e i dialetti, e dileggiandone la condotta ("Delinquente e codardo, non neanche stai a casa tua a combattere la mafia, vieni qua a non fare un cazzo"). Il personaggio prosegue avvertendo i meridionali che se avessero davvero compreso il disprezzo che lui stesso e la sua fazione politica (la Lega) provavano per loro, non si sarebbero sognati di votare la Lega Nord alle elezioni ("Terroni, ma che cazzo di problema avete? Dovreste vergognarvi" tr. nostra dal dialetto veneto). La presenza di un attore dovrebbe rendere chiara la cornice finzionale del monologo in cui chi parla recita la parte del razzista adottandone lessico e luoghi comuni, così come dovrebbe essere abbastanza trasparente che, al di là della durezza dei modi, il messaggio del pezzo è che affidarsi a fazioni politiche che seminano odio e intolleranza è poco lungimirante, poiché mentre nel tempo il target del disprezzo e della discriminazione può cambiare, restano invariate le modalità dell'esclusione e della frattura sociale. Uno sguardo ai commenti sulla pagina Facebook, tuttavia, mostra come tutti questi livelli di senso non siano affatto trasparenti quando un contenuto viene diffuso sui social e raggiunge un numero impressionante di utenti (68000 reazioni circa, 9400 commenti circa e 101853 condivisioni al 29/07/2019), con potenzialmente sempre meno informazioni sul contesto in cui è stato prodotto il video. Abbiamo ristretto la nostra osservazione a un sottoinsieme estremamente limitato di commenti (quelli lasciati nella seconda metà del luglio 2019). Si osserva che vari utenti rimangono feriti dal linguaggio discriminatorio e dai contenuti offensivi e sembrano non cogliere o non considerare l'elemento fortemente polemico del monologo ("Buffone razzista schifoso"; "Ma nn ti fai schifo da solo per quello che dici...!!!!!! Sai quanti terroni vorrebbero stare al loro paese... a respirare l'aria buona invece che respirare smog per un misero stipendio di €1200 coglionazzo"; ma anche "Ma vai a fanculo uomo di merda polentone e imbecille"); altri cercano di spiegare il senso

[21] www.facebook.com/thisisrazzismo/videos/551229588670224/. Ultima consultazione: 29/07/19.

[22] www.facebook.com/thisisrazzismo/?eid=ARAzYbhgBKJygJGqCpZFMHZyMs0d4vcV8qDXUZkXONNdrnEFhyeRnmDODLEKo_FaxsAyDiMcw-D-0r-. Ultima consultazione: 29/07/19.

del monologo ("Si è sempre il sud di qualcun altro"; "Guardatelo tutto fino alla fine. È chiaro che è un attore. Da terrone la trovo una disamina stupenda. Meditate amici terroni che avete votato Salvini, meditate"); altri ancora se la prendono con chi, insultando l'attore che impersona il razzista, non sembra cogliere il senso del monologo ("una delle cose più intelligenti che abbia visto in 10 anni di facebook (chi non capisce che non è un monologo contro i meridionali ma contro l'intolleranza e la metodica politica e sociale del capro espiatorio dovrebbe valutare l'idea di non andare più a votare)", ma anche "Ma possibile che siate così ritardati da non capire il messaggio che vuole mandare? Ma lo avete visto tutto il video? Avete sentito quello che ha detto? O vi siete fermati al "terroni di merda"? Italiani medi con zero cervello..."); altri utenti esplicitano la presenza di due livelli di senso ("Esistono due tipi di persone quelli che hanno capito il video e quelli a cui è rivolto"), altri ancora riconoscono la bontà del messaggio, ma restano scioccati dalla durezza dei modi ("Sei stato giusto però davvero un cesso, non so se essere offeso o meno"); altri si chiedono perché il contenuto non sia stato censurato ("Perché non bloccano questa 'persona'" – qui sono interessanti le virgolette: potrebbero essere un modo per offendere chi parla insinuando che non sia nemmeno meritevole di essere chiamato persona o piuttosto di sottolineare come sia difficile individuare una persona in un monologo recitato da un attore e postato su una pagina Facebook). La difficoltà nell'interpretare l'uso degli epiteti denigratori come standard (e dunque denigratorio) oppure come prospettico (cioè come se a parlare fosse un razzista) emerge anche dai botta e risposta dei commenti (abbiamo sostituito i nomi degli utenti con 'A' e 'B'): A: "Tranquilli che quando Salvini avrà finito con i negri si occuperà dei terroni"; B: "A: MANGIA POLENTA DEL CAZZO"; A: "Quando c'è l'intuito per capire il sarcasmo c'è tutto. ciao furrrrreettto".

Nei due casi di usi prospettici del linguaggio, chi parla prende il punto di vista di un antimeridionalista e ne adotta il lessico discriminatorio. Nel primo caso, lo scopo ultimo del parlante è lodare la solidarietà dimostrata dalla città di Crotone fingendo di insultarli adottando il lessico antimeridionalista; nel secondo caso, è sostenere l'assurdità di aderire a un partito xenofobo che ha rivolto a chi emigra (dal Sud Italia o dal mondo intero) lo stesso odioso trattamento.

Nei casi presi in esame, l'uso del termine 'terrone', e in generale del discorso apparentemente discriminatorio, può essere analizzato non solo come un uso prospettico del linguaggio, ma anche come ironico e/o comico. Tuttavia, come si è detto a proposito dell'ironia e dell'umorismo, non esiste solo il problema del riconoscere un proferimento come qualcosa che vorrebbe essere ironico, ma anche la questione della legittimità dell'ironia. A questo proposito, risalendo ai commenti lasciati nei primi giorni successivi alla pubblicazione del video di 'This is racism' (ottobre

2018), troviamo uno scambio di commenti particolarmente interessante. Un utente (che chiamiamo A) cerca di illustrare il senso del video a chi lo interpreta come antimeridionalista; un altro utente (B) risponde a questa spiegazione sostenendo che il problema non è tanto *capire* che il messaggio veicolato non è antimeridionalista, quanto il fatto che resta offensivo usare il linguaggio d'odio (pur per finta) contro una categoria di persone che sono *già* oggetto di discriminazione.

A: "Per chi non l'avesse capito: la persona presente nel video è un ATTORE che sta recitando un COPIONE. Non sta offendendo i meridionali o il sud. L'intento del monologo è far comprendere che c'è sempre bisogno di un capro espiatorio, prima erano i cosiddetti "terroni", poi, una volta che la loro presenza al nord ha iniziato ad essere considerata normale, si è passati ai neri, e "polentoni" e "terroni" si sono finalmente trovati d'accordo su qualcosa, ovvero dare addosso al "nemico comune". Questo per ricordare che siamo sempre i "terroni" di qualcuno, e che gli episodi di razzismo che si stanno verificando in Italia contro i neri potrebbero accadere anche a noi, ad esempio, in un paese dell'Europa del nord. L'amara ironia sta nel fatto che per 300 anni italiani del nord e del sud non sono mai riusciti a formare un popolo unito se non quando hanno trovato una categoria considerata più debole da maltrattare [...]"

B: "(...) All'inizio della lettura mi sono indignata, non per il concetto, ma per le parole usate, per me che sono una terrona, anzi no una marina, visto quanto mare abbiamo! Tu non fai molto testo sei una polentona, per noi é diverso visto che siamo continuamente oggetto dell'altrui retorica! Basta vedere i cori ignoranti e cattivi che fanno negli stadi! C'era proprio bisogno, per esprimere un concetto tanto importante, usare un testo così offensivo!

Questo scambio esemplifica bene la distinzione che tracciamo tra riconoscere e legittimare gli usi ironici del linguaggio. Sempre tra i commenti dell'ottobre 2018, troviamo poi questo, che ci conduce al tema del prossimo paragrafo: "Video scandaloso, andrebbe segnalato per incitazione all'odio [...]".

4. La censura

Uno dei motivi che spingono gli amministratori dei social network, e più in generale dei social media, a implementare sistemi che distinguano usi letterali e non letterali del linguaggio è la possibilità di gestire la censura dei contenuti. I contenuti pubblicati sui social sono sottoposti a una serie di requisiti; questo richiede di affrontare questioni teoriche e pratiche insieme. Pensando ai fenomeni che abbiamo considerato in questo articolo – riappropriazione e usi ironici del linguaggio d'odio –, per arrivare a formulare delle policies, bisogna rispondere a domande

teoriche molto complesse, quali ad esempio cosa conti come hate speech[23], cosa sia l'ironia, l'umorismo, se e quando l'umorismo razzista/sessista/omofobo/antisemita ecc. possa essere legittimo e così via. D'altro canto, per mettere in pratica le policies che si sono formulate, è necessario affrontare problemi pratici, per esempio come concretamente riconoscere un enunciato ironico da uno serio, una battuta di spirito da un'affermazione sincera, un caso di linguaggio d'odio da un uso riappropriativo; avere una ricetta del genere permetterebbe di facilitare il lavoro degli esseri umani e idealmente di insegnare a delle macchine a effettuare la stessa operazione su larga scala e in tempi più brevi.

Il fatto che interventi che sembrano adoperano linguaggio d'odio come quelli dei Sentinelli o di Carlo Gabardini o come il monologo pubblicato da This is Racism non siano oggetto di censura rivela che i criteri di un social network come Facebook devono includere una qualche distinzione tra usi riappropriativi e usi denigratori, così come una distinzione tra usi seri e usi ironici o comici dell'hate speech. Il caso del monologo "Quando i neri erano i meridionali: ovvero, l'ultimo è 'il più terrone' di tutti" è ancora più interessante in questo senso perché tra i commenti alcuni utenti, come si accennava sopra, hanno evocato l'idea che il contenuto andasse censurato ("Video scandaloso, andrebbe segnalato per incitazione all'odio [...]").

A testimonianza del fatto che la questione circa quale uso ironico o comico degli epiteti sia concesso resta aperta, vi sono interessanti casi di censura che riguardano i comici; si prenda per esempio lo Sgargabonzi, uno scrittore comico italiano, il cui lavoro si esprime anche tramite gli status di Facebook. La pagina dello Sgargabonzi è stata chiusa varie volte, arrecando un danno importante per un comico che lavora soprattutto tramite Facebook. Alcune delle segnalazioni anonime che hanno portato alla chiusura temporanea della pagina avevano proprio a che fare con l'uso di termini come 'frocio', che doveva però essere attribuito a un personaggio omofobo e cripto-omosessuale[24]. Lo stesso comico ha espresso, col registro che gli compete, il proprio dissenso nei confronti delle politiche adottate dal social network da cui era stato bannato: "Nel frattempo, impaurito come una mammola, ho tolto ogni parola a rischio dai commenti: frocio, gay ma per sicurezza pure parrucchiere. E ci sarà chi leggendo i futuri status politicamente correttissimi (per forza!) dirà che Lo Sgargabonzi s'è rammollito. Se poi vai a leggere il codice di comportamento su Facebook ti metti le mani nei capelli per

[23] Per una trattazione filosofica delle leggi sul discorso d'odio, si veda Brown (2015).

[24] La vicenda è ricostruita dal giornalista Claudio Giunta per Internazionale: www.internazionale.it/opinione/claudio-giunta/2017/03/15/sgargabonzi-facebook-pagina-bannata?fbclid=IwAR3xCpta3YsXiQlOdy-xksmhl2S4ZC6_JNfpU5Uw8uKtMeJhc_Xpr56sRtQ. Ultima consultazione: 29/07/19.

come tutto è vago, vaporizzato e opinabile"[25]. Senza entrare nel merito su chi avesse ragione – se cioè l'umorismo dello Sgargobonzi violasse o meno le politiche di Facebook – è utile ribadire ancora una volta la necessità di distinguere il problema del riconoscimento dell'ironia (non difficile quando si tratta di una pagina comica), dalla questione della legittimità dell'ironia e della comicità rispetto a temi quali la discriminazione razziale, di genere, ecc. L'identità di chi parla non gioca un ruolo centrale solo rispetto alla prima questione, ma anche rispetto alla seconda. Ammesso che l'algoritmo di un social network possa essere addestrato a *riconoscere* l'ironia (nel senso di distinguere enunciati seri da enunciati ironici), difficilmente potrà stabilire se si tratti di ironia da legittimare o da condannare.

La questione della censura sui social network è molto complessa; si tenga per esempio conto del fatto che il nostro sguardo si è concentrato sulla dimensione verbale, che non è che un aspetto della comunicazione. Uscendo dall'ambito linguistico, un caso interessante è la censura delle immagini di nudo. Gli algoritmi che individuano i nudi – atti a individuare oscenità ed eventualmente pornografia – faticano a distinguere gli usi artistici (concessi) dagli usi non artistici. Come sanno bene gli iscritti a pagine di storia dell'arte come Le Connoisseur[26], Facebook domanda ai singoli utenti se un certo quadro contenente un nudo sia da ritenersi osceno o meno. In questo modo, integra l'apporto di un essere umano al lavoro di riconoscimento dell'algoritmo, assumendo che le opinioni raccolte tramite gli utenti siano una buona guida per stabilire cosa sia da censurare e cosa no. Una procedura del genere potrebbe essere adottata nei casi linguistici discussi fin qui, ma ancora una volta l'identità di chi valuta è un fattore fondamentale e problematico da tenere in considerazione: per esempio, si può sostenere che la sensibilità dei gruppi colpiti da un certo tipo di umorismo sia da tenere più in considerazione di quella di chi non è toccato direttamente (uno dei commenti citati contro il video di *This is racism* sembrava presupporre considerazioni di questo tipo).

Quel che è certo è che il funzionamento di algoritmi simili, atti a censurare i contenuti vietati sul social in questione, porta a galla gli aspetti contraddittori e problematici che vanno al di là di cosa sia consentito dire sui social network. In altre parole, la questione di cosa sia ammesso sui social attira l'attenzione sugli usi e costumi fuori-da-internet, cioè della società che usa i social. La rapidità con cui i contenuti possono viaggiare e diffondersi online richiede di sviluppare degli algoritmi che soddisfino il più precisamente possibile i desiderata delle politiche di censura, ma questo richiederebbe di avere chiaro quali tali desiderata siano.

[25] La vicenda è ricostruita dal giornalista Claudio Giunta per Internazionale: www.internazionale.it/opinione/claudio-giunta/2017/03/15/sgargabonzi-facebook-pagina-bannata?fbclid=IwAR3xCpta3YsXiQlOdy-xksmhl2S4ZC6_JNfpU5Uw8uKtMeJhc_Xpr56sRtQ. Ultima consultazione: 29/07/19.

[26] www.facebook.com/groups/leconnoisseur/. Ultima consultazione: 29/07/19.

5. Considerazioni conclusive

I social network possono sollevare problemi inediti nello studio della comunicazione, ma più spesso amplificano fenomeni già esistenti. In questo lavoro, ci siamo occupati del ruolo giocato dall'identità (o presunta tale) di chi scrive nel guidare chi legge verso una corretta interpretazione di alcuni usi del linguaggio apparentemente discriminatori (in particolare, usi non denigratori degli slurs da un lato e ironia e umorismo a prima vista discriminatori, dall'altro). Come abbiamo visto, in questi casi l'identità di chi scrive è un elemento cruciale nell'interpretazione di un certo contenuto: perderla di vista può modificare radicalmente il significato complessivamente inteso e in certi casi persino mettere in discussione la legittimità dell'ironia e della comicità rispetto a temi quali la discriminazione razziale, di genere, ecc.

Più in generale, l'impossibilità di stabilire l'identità di chi scrive pone seri problemi per quanto riguarda il controllo e la censura di post e commenti con contenuti come quelli da noi esaminati: anche se si riuscisse ad addestrare un algoritmo a distinguere tra usi seri e ironici del linguaggio, rimane una questione aperta come sia possibile stabilire, qualora si tratti di un uso ironico riferito a contenuti apparentemente discriminatori, se tale uso sia da legittimare o da censurare. Ci sono infatti molte variabili da tenere in conto e che coinvolgono sia aspetti teorici che pratici.

Speriamo che la discussione offerta in questo articolo possa aiutare a mettere a fuoco i problemi sollevati dall'impiego sui social di questi particolari usi non letterali del linguaggio, così come i desiderata per le policies che ne regolano l'utilizzo.

Riferimenti bibliografici

Anderson, L. (2015). Racist humor. *Philosophy Compass*, *10*, pp. 501-509.

Anderson, L., Lepore, E. (2013). Slurring words. *Noûs*, *47*, pp. 25-48.

Attardo, S., Eisterhold, J., Hay, J., Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor*, *16*, pp. 243-260.

Bianchi, C. (2014). Slurs and appropriation: An echoic account. *Journal of Pragmatics*, *66*, pp. 35-44.

Bolinger, R.J. (2017). The pragmatics of slurs. *Noûs*, *51*, pp. 439-462.

Brontsema, R. (2004). A queer revolution: Reconceptualizing the debate over linguistic reclamation. *Colorado Research in Linguistics*, *17*, pp. 1-17.

Brown, A. (2015). *Hate Speech Law*. London: Routledge.

Cepollaro, B. (2015). In defense of a presuppositional account of slurs. *Language Sciences*, *52*, pp. 36-45.

Cepollaro, B., Stojanovic, I. (2016). Hybrid evaluatives. *Grazer Philosophische Studien*, *93*, pp. 458-488.

Cepollaro, B., Sulpizio, S., e Bianchi, C. (2019). How bad is it to report a slur? An empirical investigation. *Journal of Pragmatics*, *146*, pp. 32-42.

Diaz-Leon, E. (2019). Pejorative terms and the semantic strategy. *Acta Analytica*, pp. 1-12, https://doi.org/10.1007/s12136-019-00392-2 (ultima consultazione 15/09/2019).
Freud, M. (1957). *Glory reflected: Sigmund Freud, man and father*. London: Angus & Robertson.
Hom, C. (2008). The semantics of racial epithets. *Journal of Philosophy*, *105*, pp. 416-440.
Hom, C. (2010). Pejoratives. *Philosophy Compass*, *5*, pp. 164-685.
Hom, C., May, R. (2013). Moral and semantic innocence. *Analytic Philosophy*, *54*, pp. 293-313.
Hom, C., May, R. (2018). Pejoratives as fiction. In D. Sosa (a cura di), *Bad words*. Oxford: Oxford University Press, pp. 108-131.
Jeshion, R. (in pubblicazione). Pride and prejudiced: On the appropriation of slurs. *Grazer Philosophische Studien*.
Macià, J. (2002). Presuposición y significado expressivo. *Theoria: Revista de Teoria, Historia y Fundamentos de la Ciencia*, *3*, pp. 499-513.
McCready, E. (2010). Varieties of conventional implicature. *Semantics and Pragmatics*, *3*, pp. 1-57.
Nunberg, G. (2018). The social life of slurs. In D. Fogal, D. Harris e M. Moss (a cura di), *New Work on Speech Act*. Oxford: Oxford University Press, pp. 237-295.
Panzeri, F. (2019). Stai scherzando? (Non) riconoscere l'ironia nei social network. *Sistemi Intelligenti*, *31*, 3, pp. 489-506**.**
Potts, C. (2005). *The logic of conventional implicatures*. Oxford: Oxford University Press.
Potts, C. (2007). The centrality of expressive indexes. Reply to commentaries. *Theoretical Linguistics*, *33*, pp. 255-268.
Pullum, G.K. (2018). Slurs and obscenities: Lexicography, semantics, and philosophy. In D. Sosa (a cura di), *Bad words*. Oxford: Oxford University Press, pp. 168-192.
Rappaport, J. (2019). Communicating with slurs. *The Philosophical Quarterly*, https://doi.org/10.1093/pq/pqz022 (ultima consultazione 15/09/2019).
Schlenker, P. (2007). Expressive presuppositions. *Theoretical Linguistics*, *33*, pp. 237-245.
Sperber, D., Wilson, D. (1986/1995). *Relevance. Communication and Cognition*. 2ª ed. Oxford: Blackwell (trad. it. della 1ª ed. *La pertinenza*. Milano: Anabasi, 1993).
Sperber, D., Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind and Language*, *17*, pp. 3-23.
Sperber, D., Wilson, D. (2012). *Meaning and Relevance*. Cambridge (MA): Cambridge University Press.
Stanley, J. (2015). *How propaganda works*. Princeton: Princeton University Press.
Yus, F. (2011). *Cyberpragmatics. Internet-Mediated Communication in Context*. Amsterdam: John Benjamins.

Identity and discriminatory language in social networks

Social networks pose new problems in the study of communication, while also amplifying old ones. In this paper we assess the role that the identity of users plays in guiding their readers towards the correct interpretation of their apparently discriminatory uses of language on social network such as Facebook and Twitter: in particular, we focus on non-derogatory uses of slurs on the one hand, and apparently discriminatory irony and humor on the other hand. In these cases, the user's identity plays a crucial role in the interpretation of the content conveyed by her utterance: to lose track of it may dramatically change the intended meaning, as well as, in some cases, question the legitimacy of her ironic and humoristic uses of language in relation to issues such as racial and gender discrimination. Moreover, the fact that the identity of the writer is hard to pin down poses serious challenges to the issue of censorship: to develop specific policies – which is of the highest importance in managing social networks – requires addressing many theoretical as well as practical issues that we tried to illustrate in this paper.

Keywords: social networks, slurs, hate speech, irony, racist humor, sexist humor.

Bianca Cepollaro, Università Vita-Salute San Raffaele, Milano, Facoltà di Filosofia, via Olgettina 58, 20132, Milano, bianca.cepollaro@gmail.com

Paolo Labinaz, Università di Trieste, Dipartimento di Studi Umanistici, Androna Campo Marzio 10, plabinaz@units.it

GIOVANNI TUZET

MOLESTE ANALOGIE? SOCIAL NETWORK E NORME PENALI

1. Diritto e social network

Internet e i social network pongono ai giuristi una serie di questioni nuove, fra cui la protezione dei dati elettronici, la tutela della privacy, la protezione da forme di sorveglianza e influenza occulta da parte di chi può seguire le nostre navigazioni digitali, la difficoltà di mantenere il legittimo segreto di date informazioni stante la facilissima diffusione, il difficile equilibrio fra tutela della libertà d'espressione e contenimento della violenza verbale. Il doppio volto di Internet, fra libertà e controllo, chiede cose diverse e spesso ambiguamente[1].

L'attività normativa del diritto si svolge attraverso la qualificazione deontica di determinate condotte (permesse, vietate, imposte) e la determinazione di sanzioni in capo a chi ne violi i precetti. In relazione al nostro tema, si tratta della qualificazione deontica di condotte proprie della società dell'informazione e della comunicazione digitale. Talvolta è una qualificazione controversa. Per esempio, viola il diritto penale chi partecipi a un social network con una falsa identità, o ingannando gli altri utenti circa la propria professione o altra caratteristica rilevante? Oppure, chi ha la responsabilità giuridica di quanto accade sulle piattaforme digitali? O ancora, chi è responsabile dei danni provocati a terzi nello svolgimento di un'attività lavorativa coordinata da una piattaforma digitale per trasportare persone o consegnare cibo?

La comunicazione attraverso social network presenta alcune caratteristiche comuni alla comunicazione tradizionale e altre caratteristiche che la rendono diversa e peculiare. Per le prime sono applicabili e sufficienti le norme che regolano condotte di tipo tradizionale; per le seconde caratteristiche occorrono strumenti nuovi, o almeno l'estensione di strumenti tradizionali.

Il presente contributo si occupa dell'estensione (analogica?) di norme giuridiche di natura penale, pensate per la comunicazione tradi-

[1] V. fra gli altri Pollicino, Bertolini e Lubello (2013); Ziccardi (2015; 2016); Pitruzzella, Pollicino e Quintarelli (2017).

zionale, alla comunicazione attraverso social network. In particolare, si può estendere alla comunicazione via social network una norma sulle molestie recate "in luogo pubblico o aperto al pubblico"? Il problema giuridico è che nel nostro ordinamento – come nella maggioranza di quelli contemporanei – non è consentita l'estensione analogica delle norme incriminatrici, mentre non è proibita la loro estensione "interpretativa". Presenterò un argomento in favore di questo secondo tipo di estensione (§ 5). Prima di ciò farò alcuni cenni al ruolo conoscitivo dell'analogia e al suo ruolo normativo (§§ 2-3), richiamando in particolare la discussione su determinate norme relative a veicoli; quindi mi concentrerò sul problema della distinzione fra ragionamento analogico e interpretazione estensiva, indicando una maniera di tracciarne il confine e sostenendo al contempo che l'onere argomentativo è in capo a chi voglia tracciarlo (§ 4). Chiedo al lettore la pazienza di seguire questi passaggi per arrivare con cognizione di causa alla questione delle molestie via social network.

2. L'analogia come strumento conoscitivo

In un testo pubblicato originariamente nel 1898, il sociologo Émile Durkheim sostiene che la società deve essere intesa come intendiamo la mente e intelligenza umana; ossia, così come la mente non è riducibile alle cellule cerebrali, la società non è riducibile agli individui che la compongono[2]. In termini contemporanei si direbbe che le proprietà mentali non sono riducibili a quelle neuronali e che l'ontologia sociale non è riducibile a quella degli individui. La mente umana presenta un'elevata complessità derivante dall'interazione fra le sue componenti, il corpo e l'ambiente. Allo stesso modo, la società si istituisce e trasforma con dinamiche complesse di cui non può rendere conto uno stretto individualismo metodologico e ontologico.

È chiaro che l'argomento di Durkheim ha una struttura analogica: la società sta agli individui come la mente sta alle cellule; pertanto, se non vale il riduzionismo neurologico rispetto alla mente non vale neppure quello individualista rispetto alla società. L'argomento ha la pretesa di trarre una conclusione su un dominio più controverso a partire da un dominio meno controverso con cui il primo ha una relazione analogica. Agli occhi di Durkheim è meno controverso il rapporto fra mente e cellule del cervello, poiché nessuno può sostenere ragionevolmente che la prima sia riducibile alle seconde; mentre è più controverso il rapporto fra società e individui, dato che diversi sociologi e filosofi autorevoli hanno difeso un riduzionismo individualista. Durkheim intende mostrare che,

[2] *Représentations individuelles et représentations collectives*, ora in Durkheim (2010, cap. 1). Cfr. Elster (1983, 35 ss.) per una critica all'uso di metodi sociologici in biologia e di paradigmi biologici nelle scienze sociali.

partendo dal dominio meno controverso, si può trarre una conclusione analogica su quello più controverso, allo scopo di migliorarne la comprensione o conoscenza, in virtù della somiglianza strutturale fra i due.

L'argomento di Durkheim è un esempio di uso conoscitivo dell'analogia. Per ripeterlo: al fine di avanzare nella comprensione o conoscenza di un dominio meno noto o più controverso, si muove da un dominio più noto o meno controverso con cui il primo ha una relazione analogica. Un altro esempio è il tentativo di utilizzare strumenti economici per approfondire la conoscenza e la spiegazione delle dinamiche giuridiche[3], assumendo che le decisioni assunte in ambito giuridico siano analoghe a quelle prese in ambito economico, che i comportamenti dei destinatari delle norme giuridiche siano analoghi a quelli di mercato, ecc.

Questo è un uso tradizionale dell'analogia. "Comparare ciò che è meno conosciuto con ciò che è maggiormente conosciuto assume da sempre il nome di *analogia*"[4]. Nella filosofia della scienza non sono mancati gli studi sul ruolo delle analogie, delle similitudini, delle metafore e dei modelli nell'impresa scientifica[5]. Discutere tutto ciò ci porterebbe però lontano dal tema del presente contributo.

Come strumento conoscitivo, in estrema sintesi, l'analogia ha lo scopo di integrare la conoscenza di un dominio *meno conosciuto* ricorrendo a un dominio *più conosciuto* con cui il primo ha una somiglianza rilevante. Si noti: non una somiglianza qualsiasi, ma una somiglianza *rilevante* sotto un profilo strutturale o causale. Data una certa struttura o un certo aspetto causale nel dominio più conosciuto si argomenta analogicamente sostenendo che la stessa struttura o lo stesso aspetto conta pure nel dominio meno conosciuto. C'è analogia se c'è somiglianza rilevante fra il dominio di partenza (quello più conosciuto) e quello d'arrivo (meno conosciuto).

3. L'analogia come strumento normativo

Nell'argomentazione giuridica l'analogia ha lo scopo di estendere, a una fattispecie *non espressamente regolata*, la disciplina *espressamente prevista* per almeno una fattispecie regolata con cui la prima ha in comune almeno una *proprietà rilevante* (in virtù di cui le fattispecie hanno una *somiglianza rilevante*)[6]. A differenza di quanto accade in ambito epistemico, l'analogia giuridica risponde a un problema normativo[7]: come trattare una fattispecie che non è stata espressamente regolata?

[3] Vedi ad es. Cooter, Mattei, Monateri, Pardolesi e Ulen (2006).
[4] Kaufmann (2001, 323).
[5] V. per tutti Hesse (1966). Cfr. Lakoff (1987).
[6] Ho sviluppato questa tesi in Tuzet (2010, cap. 5).
[7] Come segnala un anonimo revisore, la questione è però controversa: c'è chi ritie-

Un presupposto dell'argomento è che, tecnicamente, ci sia una *lacuna* nel diritto. Ossia una sorta di "buco", la mancata regolazione di una fattispecie che è significativa sotto il profilo giuridico. Intendiamoci bene su questo punto: non tutto ciò che passa sotto silenzio si configura giuridicamente come una lacuna. A quanto ne so nessun ordinamento giuridico disciplina l'ordine con cui allacciarsi le scarpe (se prima la destra, poi la sinistra, o viceversa) ma la cosa non preoccupa alcuno data la sua irrilevanza sotto il profilo giuridico. Quando invece a non essere regolata è una fattispecie giuridicamente significativa si ha una lacuna in senso tecnico.

È perciò corretto ma poco caratterizzante dire che l'analogia in ambito giuridico è un "ragionamento che lavora con le somiglianze e differenze"[8], o che ogni applicazione di diritto configura "un processo di tipo analogico"[9]. In senso ampio è analogia qualsiasi ragionamento che comporta una comparazione; in senso tecnico-giuridico è il ragionamento che presuppone una lacuna.

Facciamo un esempio: se un ordinamento consente il risarcimento per il furto subito dal cliente in una camera d'hotel e parallelamente *non* consente il risarcimento per il furto subito dal passeggero su un treno, cosa dire del furto subito nella cabina di un battello a vapore? Questo furto è risarcibile? La compagnia che presta il servizio è tenuta a compensare il cliente della perdita subita? L'esempio non è immaginario. L'ordinamento statunitense di fine Ottocento considerava risarcibile il furto in hotel ma non il furto in treno. Nel caso *Adams* (1896)[10] si pose il problema del furto in battello e tecnicamente i decisori si trovavano difronte a una lacuna: l'ordinamento non disciplinava espressamente tale fattispecie. È così che si apre la questione dell'analogia, se cioè la lacuna possa essere colmata in via analogica.

Con un altro esempio, se un ordinamento non stabilisce espressamente chi è proprietario di risorse quali il gas naturale, questo è da considerare simile a un frutto della terra o a un animale selvatico? Nel primo caso, la proprietà del gas è del titolare del terreno; nel secondo è di chi lo estrae, come se catturasse una lepre[11].

Con ciò arriviamo a un ulteriore punto che vorrei sottolineare: la somiglianza invocata per argomentare analogicamente deve essere rilevante. A differenza di quanto accade nelle analogie conoscitive in cui la rilevanza delle proprietà è data dal loro profilo causale (*P* è causa di *Q*),

ne che il giudice debba "scoprire" come il legislatore *tratterebbe* (controfattualmente) la fattispecie che non ha regolato. O ancora, per qualcuno, il giudice deve "scoprire" come *si debba trattare* la fattispecie; in questo senso il problema normativo avrebbe una dimensione epistemica.

[8] Trujillo (2014, 8).

[9] Zaccaria (1990, 175), che cita una tesi di A. Kaufmann, ora in Kaufmann (2003).

[10] Adams v. New Jersey Steamboat Co., su cui cfr. Posner (2006).

[11] Vedi Posner (2006, 765-766).

nelle analogie giuridiche la rilevanza è stabilita dalle ragioni normative della disciplina (*P* è una ragione per *Q*).

Da una parte, se sappiamo che un'entità meno nota ha in comune con una più nota una proprietà *P* causalmente rilevante per la proprietà *Q*, possiamo inferirne per analogia che anche l'entità meno nota ha la proprietà *Q*[12]. Dall'altra, se una fattispecie non regolata ha in comune con una fattispecie regolata una proprietà *P* che è una ragione normativa della disciplina *Q*, possiamo inferirne per analogia che la disciplina *Q* va estesa alla fattispecie non regolata.

Posto questo, come trattare il caso del furto in battello? La lacuna poteva essere colmata con un'analogia rispetto al furto in treno, o doveva essere trattata come il furto in hotel? Con la prima soluzione il furto non era risarcibile, mentre lo era con la seconda. L'esempio mostra come non sia facile indicare la somiglianza rilevante che giustifica un'estensione analogica del diritto, specie quando sono in gioco più somiglianze. Il battello è simile al treno in quanto mezzo di trasporto, ma è simile all'hotel in quanto è possibile ottenervi uno spazio riservato (camera o cabina) dove riposare e lasciare i propri beni per un certo tempo.

I giudici di *Adams* ritennero che la disciplina sul furto in hotel proteggesse l'affidamento del cliente, cioè il suo confidare nel prestatore del servizio: il cliente che paga per una cabina sul battello riceve per contratto uno spazio chiuso e riservato in cui può lasciare i propri beni, confidando nelle garanzie dategli dal prestatore del servizio, così come in hotel e a differenza di quanto accade in treno[13]. Pertanto fu deciso che la somiglianza rilevante era quella con l'hotel e che la compagnia prestatrice del servizio in battello era tenuta a risarcire il furto analogamente a un albergatore e diversamente da una compagnia ferroviaria.

La struttura dell'argomento era essenzialmente questa: dato che si vuole proteggere l'affidamento del cliente, la disciplina *Q* (risarcimento) va estesa al furto in battello poiché questo condivide con il furto in hotel la proprietà rilevante *P* (furto in uno spazio riservato prestato da chi fornisce il servizio).

Adams offre uno dei numerosi esempi in cui sono in gioco analogie a proposito di veicoli o mezzi di trasporto[14]. La casistica è molto nutrita, alimentata dalla vaghezza dell'iperonimo "veicoli" e dal moltiplicarsi delle sue specie. Se in *Adams* si discuteva di treni e battelli, in *McBoyle*

[12] Ulteriore questione, colta da un altro revisore e che non posso approfondire qui, è se *P* sia condizione sufficiente di *Q*, o sua condizione necessaria, o si limiti ad aumentarne la probabilità. La formulazione usata nel testo è volutamente generica.

[13] Ma si può discutere dei vagoni con spazi riservati in cui chiudersi, dormire, ecc.

[14] Un'altra vicenda interessante è quella degli oggetti "inerentemente pericolosi", su cui v. Levi (1948). Per una lettura pragmatista della questione, cfr. Haack (2018, 1065-1066), secondo cui, come le lingue e le tecnologie, i sistemi giuridici mutano nel tempo, si adattano a nuovi problemi, ecc.

(un altro caso americano, del 1931)[15] si discuteva di veicoli e aeroplani, poiché un testo normativo sanzionava il furto di "veicoli a motore" e l'imputato era riuscito a sottrarre un aeroplano; in *Corkery* (un caso inglese del 1951)[16] si discuteva di veicoli e biciclette, poiché ai sensi di una norma del 1872 era sanzionata la guida di "carrozze" in stato di ebbrezza e l'accusato andava ubriaco in bicicletta, mentre in un famoso esempio teorico[17] si vieta ai "veicoli" l'ingresso nel parco e si discute se ammettere biciclette, veicoli giocattolo, skateboard, pattini a rotelle, ecc. Per non parlare del caso tedesco in cui, stante una legge penale del XIX secolo che sanzionava come furto forestale aggravato quello operato con "bestie da soma, barche e carri da traino", alla metà del XX secolo la Corte suprema federale si trovò a decidere se fosse condannabile per furto forestale aggravato chi rubava legname in un bosco trasportando con un'automobile la refurtiva[18]. In tutti questi casi ci si chiede come sia corretto argomentare e con quali ragioni.

Nella letteratura sul tema c'è chi ha tentato di ricondurre l'analogia al ragionamento deduttivo. Norberto Bobbio, in particolare, ha sostenuto[19] che c'è analogia quando *P* è *ragione sufficiente* di *Q*. La struttura dell'argomento sarebbe questa: se la fattispecie 1 è *Q* in quanto *P*, se la fattispecie 2 è simile alla 1 in quanto *P*, e se *P* è condizione sufficiente di *Q*, allora la fattispecie 2 è necessariamente *Q*.

Se è così, è chiaro che l'analogia si trasforma in una deduzione in quanto *P* è condizione sufficiente di *Q*. Una volta risaliti alla comune proprietà *P* e riconosciuta in questa una condizione sufficiente di *Q*, non v'è dubbio sulla conclusione dell'argomento. Invece la configurazione dell'argomento presentata qui sopra è incentrata sulla rilevanza della somiglianza (nelle sue declinazioni strutturali, causali, normative) ed è più debole poiché non equipara la rilevanza a una ragione sufficiente. Ritengo che questa configurazione sia preferibile in quanto rende conto di una più vasta e articolata serie di argomenti rispetto a quelli cui si riduce l'analogia nel senso di Bobbio. Ma non è il punto che voglio discutere qui.

Il problema che voglio trattare ora è che molti ordinamenti giuridici contemporanei, fra cui il nostro, *vietano* l'analogia in diritto penale. Per la precisione, vietano l'applicazione analogica delle norme incriminatrici, cioè delle norme che qualificano certe condotte come crimini e stabiliscono

[15] McBoyle v. United States, su cui si può vedere Canale e Tuzet (2016).

[16] Corkery v. Carpenter, su cui v. Velluzzi (2008, 505-506).

[17] In Hart (1958, 607-608) e Hart (1961, cap. 7); a riguardo v. Schauer (2008).

[18] Si veda Hassemer (1997, 191). Cfr. Carlizzi (2012, 152 ss.).

[19] Bobbio (1938, 124 ss.). Di Lucia (2003, 114) ha commentato che due sono le tesi principali in tale lavoro di Bobbio: 1) la logicità della relazione analogica tra norme e 2) (come corollario della prima) l'impossibilità del divieto di analogia (su questo diremo poi). Cfr. Gianformaggio (1987, 322), secondo cui la riduzione dell'analogia a un nucleo deduttivo ne smarrisce le caratteristiche.

delle sanzioni a riguardo[20]. Perché questo divieto? La risposta classica sta nel *favor libertatis*, cioè nella protezione della libertà individuale e nella tutela da interventi punitivi dello Stato che travalichino il dettato normativo. Dove il legislatore penale non si è espresso o si è espresso in modo dubbio, la questione va risolta in favore della libertà. Il che può essere riassunto nel cd. principio generale esclusivo, valido per il diritto penale: *tutto ciò che non è espressamente vietato è permesso*. Se dunque vi sono lacune nel diritto penale, queste vanno colmate con il principio generale esclusivo e non per analogia con altre fattispecie regolate da norme incriminatrici[21]. Si noti, a tal proposito, che *Adams* era un caso civile, in quanto l'autore del furto non era stato identificato e la vittima chiedeva un risarcimento in sede civile, dove l'analogia è pacificamente ammessa.

Se però, da un lato, nel nostro ordinamento penale è vietata l'analogia, dall'altro non è vietata la cd. *interpretazione estensiva*. Questa consisterebbe in una lettura dei testi normativi che, anziché rilevare la lacunosità della disciplina, vi include la fattispecie in esame interpretando i testi in modo ampio.

4. Un problema ricorrente: analogia o interpretazione estensiva?

Nello stesso lavoro menzionato in precedenza Bobbio sostiene che "per la sua logica irresistibilità" l'analogia "incalza il giurista", il quale "non può non riconoscerne l'esigenza" e non potendo ammetterla stante il divieto legislativo è "costretto ad ammetterla" entro certi limiti o con certi espedienti come la distinzione fra analogia e interpretazione estensiva[22]; insomma l'analogia, come intesa da Bobbio, si impone al

[20] Nel gergo giuridico, sono penalmente vietate le analogie *in malam partem* mentre non lo sono quelle *in bonam partem* (cioè quelle che estendono norme di favore come le "scriminanti"). Vedi fra gli altri Ferrajoli (1989, 378-381, 732).

[21] Ma più radicalmente si potrebbe dire che, stante tale principio, non c'è nessuna lacuna nel diritto penale (appunto perché quanto non è espressamente vietato è permesso). In questo senso il principio generale esclusivo fa parte del sistema e funge da norma di chiusura dello stesso, per cui non v'è lacuna. Come nota un anonimo revisore, così il problema diventa quello della distinzione tra (i) interpretazione non estensiva cui non può essere combinata la tecnica analogica e (ii) interpretazione estensiva della fattispecie con correlativa estensione della sanzione.

[22] Bobbio (1938, 230). La tesi mi pare scorretta se fa coincidere necessità psicologica e incontinenza decisionale, se cioè consiste nel dire che il giurista non può frenare il proprio uso decisionale dell'analogia una volta ragionato in termini analogici. Come è falso che non si possa tacere quando si ha qualcosa da dire, è falso che in sede decisionale non ci si possa astenere da un argomento pur impellente. Sui controversi aspetti politici di tale tesi di Bobbio, v. invece Guastini (1976, 587-591), secondo cui agli intenti liberali di Bobbio – consistenti nel voler vincolare analogia e interpretazione estensiva a un'espansione "logica" del diritto – corrisposero gli esiti illiberali di una moltiplicazione giudiziaria delle restrizioni alla libertà personale.

ragionamento. Il lavoro in questione è del 1938 e al regime dell'epoca non dispiaceva l'idea di poter estendere per analogia le norme incriminatrici. Nel 1942 entrano comunque in vigore le cd. Preleggi al codice civile, le quali, riprendendo le Preleggi al codice del 1865 (artt. 3-4), ammettono l'analogia in sede civile (art. 12 c. II) ma la vietano espressamente in materia penale (art. 14). Nel 1950 Bobbio torna sulla questione e in un famoso articolo, che segna la nascita ufficiale della filosofia analitica del diritto, dice fra le altre cose che l'interpretazione estensiva è un'opera di "completamento del sistema"[23]. Così egli ribadisce che, nonostante l'iterato divieto nelle Preleggi, non c'è sostanziale differenza fra estensione analogica e interpretazione estensiva:

> l'estensione analogica è un modo della interpretazione estensiva, cioè è un mezzo, se vogliamo usare un'altra formula, con cui si opera l'interpretazione estensiva. Così concepita l'analogia, cioè come operazione controllata da una regola del discorso giuridico, non è, come pur da taluni giuristi si sostiene, un atto creativo, ma una operazione in stretto senso logica che non esorbita dalla considerazione della scienza giuridica come analisi del linguaggio; è quindi una delle operazioni con cui si effettua quell'analisi linguistica del diritto in cui consiste la giurisprudenza (Bobbio, 1950, 362).

Altri hanno sostenuto che non si può tracciare una distinzione "qualitativa" fra analogia e interpretazione estensiva: fra le due c'è un passaggio graduale e continuo, come momenti "di un processo fondamentalmente unitario"[24]. In maniera più ponderata, Bobbio ha sostenuto in opere successive che la distinzione fra estensione analogica e interpretativa può tuttavia essere tracciata in linea teorica, poiché "l'effetto della prima è la creazione di una nuova norma giuridica; l'effetto della seconda è l'estensione di una norma a casi non previsti da questa"[25]. Ma anche di recente non sono mancate le voci scettiche, fra cui l'opinione che "non esistendo formule che consentano di definire, a priori ed in astratto, i confini semantici delle parole e quindi delle proposizioni, non ha senso distinguere tra interpretazione *restrittiva* ed interpretazione *estensiva* del testo"[26].

Un aspetto interessante del problema è che lo scetticismo sulla distinzione fra analogia e interpretazione estensiva viene usato per portare acqua a due mulini diversi: quello di chi, essendo bandita l'analogia, vorrebbe bandire anche l'interpretazione estensiva[27] e quello di chi, essendo impossibile la loro distinzione, è incline a ridimensionare la portata

[23] Bobbio (1950, 361-362).
[24] Caiani (1958, 354).
[25] Bobbio (1993, 269) (da lezioni del 1959-60).
[26] Di Giovine (2010, 358).
[27] Io stesso sono stato incline a questa idea in Tuzet (2011). Mentre l'argomento che svilupperò di seguito è stato prospettato in Canale e Tuzet (2014).

del divieto. Gli uni si appellano al *favor libertatis* e ai principi della legalità e tassatività penale, gli altri osservano la necessità di adeguare il diritto alle mutevoli circostanze della vita sociale, tenendo specialmente conto delle innovazioni tecnologiche come i social network. Per i primi è un onere del legislatore quello di intervenire per aggiornare i testi normativi; per i secondi, dati i tempi e le difficoltà della legislazione, sarebbe opportuno lasciare ai giudici tale adeguazione del diritto in via interpretativa, consentendo senza troppi patemi d'animo l'estensione di norme il cui tenore scritto sconta i limiti del tempo e delle circostanze genetiche.

Quello che intendo sostenere è che la distinzione fra analogia e interpretazione estensiva è possibile e chiara in linea teorica e che però, stante la difficoltà di operarla in pratica, l'onere argomentativo è in capo a chi vuole invocare un'interpretazione estensiva, con la conseguenza che il mancato soddisfacimento dell'onere, o anche solo un motivato dubbio a riguardo, giustifica la conclusione che l'estensione invocata non è interpretativa bensì analogica e pertanto vietata.

In linea teorica l'argomento prende il via dalla distinzione fra testo e interpretazione del testo. I testi normativi (o disposizioni normative) sono suscettibili di essere interpretati secondo una serie di canoni che il diritto stesso pone (come all'art. 12 delle Preleggi) o che la giurisprudenza elabora. Alcuni di questi canoni sono l'interpretazione secondo il significato letterale del testo, l'intenzione del legislatore, o lo scopo della disciplina. L'utilizzo corretto dei canoni riconosciuti in un ordinamento configura la "cornice" delle interpretazioni ammissibili, le quali costituiscono le norme vere e proprie; in questo senso, le norme non sono il punto di partenza ma il risultato dell'attività interpretativa[28]. Ora, se con nessuna delle interpretazioni ammissibili si può sostenere che la fattispecie in esame è regolata dal diritto, allora se ne deve concludere che il sistema presenta una lacuna. A tal punto, dove l'analogia è permessa, si vedrà se la lacuna può essere colmata analogicamente. In questo senso l'analogia non è un argomento interpretativo, bensì *integrativo* del diritto esistente. Ad esempio è analogica la decisione in *Adams*, dove si applica al furto in battello la disciplina risarcitoria per il furto in hotel.

Quanto all'interpretazione estensiva, *in senso lato* la si può intendere come un'interpretazione più estesa di un'altra fra quelle ammissibili; *in senso stretto* come un'interpretazione più estesa di quella secondo il significato letterale (standard) del testo. Ne parlo in questi due sensi per segnalare che il più delle volte ciò che si cerca di estendere è il significato letterale (standard). Naturalmente quest'ultima nozione andrebbe discussa in dettaglio ma non è la sede per farlo. Si assuma semplicemente che almeno alcuni testi normativi hanno un significato letterale (standard). Allora, quando c'è un significato letterale (standard) e ci sono ragioni

[28] Vedi per tutti Guastini (2011) e Velluzzi (2013).

per estenderlo senza arrivare al punto di un'integrazione analogica, c'è interpretazione estensiva. Il fenomeno della *vaghezza* viene in aiuto a questo proposito: se un testo normativo è vago sotto un qualche profilo e la fattispecie in esame ricade nella zona umbratile del significato testuale, allora ci possono essere ragioni per estendere la disciplina dal nucleo alla penombra[29]. Ad esempio si può considerare estensiva l'interpretazione di "veicolo a motore" che vi include gli aeroplani, come viene deciso in *McBoyle* dalla Corte d'appello, se è vero che un aeroplano non è nel nucleo semantico di "veicolo a motore" ma nella sua penombra[30]; o si può considerare estensiva l'interpretazione di "veicoli" che vi include le biciclette. Questa è la linea argomentativa che proporrò circa le molestie tramite social network.

In un quadro dialettico vorrei aggiungere ancora che c'è un *onere argomentativo* in capo a chi solleva una pretesa di questo genere. Chi in una controversia giuridica di carattere penale vuole ricorrere a un'interpretazione estensiva dei testi ha l'onere di mostrare che non si tratta di ragionamento analogico (vietato). Se l'onere non è soddisfatto, l'argomento si considera un tentativo di estensione analogica e dunque un argomento illegittimo.

5. Molestie e social network

Una prestazione sessuale per via telematica è più simile alla prostituzione da strada o a un film a luci rosse?[31] E se è più simile alla prima si tratta di un'analogia o di un'interpretazione estensiva di "prostituzione"? Questo è un esempio di condotta nuova, resa possibile da innovazioni tecnologiche e nuove forme di comunicazione, rispetto a cui si pone il problema dell'applicazione di norme penali pensate per circostanze tradizionali.

Un altro esempio sono le molestie tramite social network. Il nostro codice penale (art. 660) sanziona le molestie recate in un "luogo pubblico o aperto al pubblico". Ora, se il molestatore agisce tramite social network si può dire che la sua condotta avvenga in un "luogo pubblico o aperto al pubblico"? Chi risponde positivamente opera un'analogia fra luoghi fisici e contesti digitali o piuttosto interpreta estensivamente tale espressione?

[29] Sulla "penombra" dei significati v. il classico Hart (1961, cap. 7).

[30] Però il legislatore americano aveva definito "veicoli a motore" indicando "un'automobile, un camion, un motociclo o qualsiasi altro veicolo semovente non progettato per andare su rotaie" e questa specificazione lasciava intendere un'applicazione a soli veicoli terrestri. Infatti la Corte suprema rovesciò la decisione d'appello considerando l'estensione come un'indebita analogia.

[31] Cfr. Vogliotti (2011, 132-133), dove si propende per la seconda opinione, con considerazioni relative allo scopo della legge contro lo sfruttamento della prostituzione.

Si ricordi che devono esservi buone ragioni per ritenere che non si tratti di un'analogia nascosta, cioè di un'operazione analogica mascherata da interpretazione estensiva. Tali ragioni devono consistere in ciò che giustifica un'interpretazione più estesa di quella letterale (standard) o un'interpretazione più estesa di un'altra comunque ammissibile.

Un caso in cui si è presentata la questione è una vicenda di molestie (soprattutto a sfondo sessuale) recate dal caporedattore di un quotidiano a una redattrice dello stesso. Parte dei fatti era stata commessa presso gli uffici del giornale e parte tramite un account fittizio sulla pagina Facebook della redattrice, dove l'imputato – sotto pseudonimo – aveva fatto apparire diversi messaggi di tale tenore.

Come hanno scritto Paolo Labinaz e Marina Sbisà, un motivo di interesse per i social network "è dato dal fatto che essi si presentano come spazi virtuali (sufficientemente aperti) in cui è possibile non solo creare e condividere contenuti, ma anche discutere e prendere posizione su di essi"[32]. Nel caso in esame il contenuto era molesto e la difesa dell'imputato lo ammetteva. La questione giuridica era se le molestie fossero state recate "in un luogo pubblico o aperto al pubblico". In primo grado l'imputato fu assolto e in secondo condannato, poiché la Corte d'appello di Firenze ritenne (diversamente dal Tribunale di Livorno) che la redazione di un giornale e la pagina Facebook della vittima costituissero dei luoghi siffatti.

La Cassazione (sez. I, sentenza n. 37596/2014) ha confermato l'indirizzo della Corte d'appello. Ciò appare meno controverso per quanto riguarda la redazione di un giornale se questa ha almeno alcuni ambienti aperti al pubblico; più controverso per una pagina elettronica o un social network. Seguiamo allora l'argomentazione della Corte.

Al fine di intendere l'espressione "luogo pubblico o aperto al pubblico" la Cassazione ne nota innanzitutto la dimensione sistemica (l'espressione ricorre negli artt. 266, 352, 404, 405, 660, 663, 688, 689, 690, 718, 720, 725, 726 del codice penale) e ne richiama l'interpretazione di dottrina e giurisprudenza consolidate: "per luogo *pubblico*" si deve intendere "quello di diritto o di fatto continuativamente libero a tutti, o a un numero indeterminato di persone; per luogo *aperto al pubblico*, quello, anche privato, ma al quale un numero indeterminato, ovvero un'intera categoria, di persone, può accedere, senza limite o nei limiti della capienza, ma solo in certi momenti o alle condizioni poste da chi esercita un diritto sul luogo"[33]. Su questa base, con quali argomenti si può sostenere che la pagina personale di un social network come Facebook sia un luogo pubblico o aperto al pubblico ai sensi del diritto penale italiano?

[32] Labinaz e Sbisà (2017, 64). Si noti l'inciso "sufficientemente aperti", che richiama uno dei punti discussi qui.

[33] Punto 3.1 della motivazione in diritto. I corsivi nel testo sono miei.

Chiediamoci se abbia più eco un evento che si verifica in un determinato luogo (pur frequentato) senza che nessuno ne registri una traccia elettronica, o piuttosto con qualcuno che lo registra e diffonde in rete. La risposta è ovvia. Il numero di persone che potrà vedere la scena nel secondo caso è incommensurabilmente superiore a quello che vi assiste dal vivo. Mi hanno di recente raccontato, ad esempio, di un matrimonio religioso in cui l'organista (amico dello sposo) ha inserito una canzonetta nella musica di cerimonia; nessuno se ne è accorto al momento, ma quando la registrazione è stata diffusa tramite social network la cosa ha assunto proporzioni e conseguenze ben diverse.

Meno diretto è il confronto fra un evento fisico diffuso in rete e un evento che accade esclusivamente in rete. Se infatti nella prima fattispecie c'è *ab origine* un luogo fisico i cui accadimenti vengono amplificati dalla rete, nella seconda non è chiaro se ci sia un "luogo" e di che tipo sia.

Per risolvere il problema dobbiamo partire esattamente da questo: se un social network elettronico sia configurabile come "luogo" e per di più sia qualificabile come "pubblico o aperto al pubblico".

Se è vero quanto detto sopra dobbiamo partire dal significato letterale delle espressioni in gioco. Mi sembra di poter dire che esse rimandino a spazi fisici e che dunque il loro significato letterale (standard) escluda i contesti digitali. Vediamo cosa ne dicono alcuni dizionari online (che in quanto tali dovrebbero essere più sensibili ai mutamenti della lingua): il vocabolario Treccani indica come primo significato di "luogo", in senso ampio, "una parte dello spazio, idealmente o materialmente circoscritta"[34]; il dizionario Garzanti una "porzione determinata dello spazio, considerata in senso generale o in astratto"[35]; il dizionario Hoepli una "parte di spazio, idealmente o materialmente determinata, che un corpo può occupare"[36]. La materialità è fuori discussione, mentre è opinabile se l'idealità così richiamata riguardi anche gli spazi digitali come un social network.

Ma allo stesso tempo mi pare asseribile che nella nostra lingua è crescente l'uso di tali e simili espressioni in riferimento a contesti digitali. Indirizzi elettronici, siti, piattaforme, bacheche... ci sono molti termini e locuzioni che dall'uso fisico si sono estesi all'uso digitale. Forse per il termine "luogo" non è altrettanto evidente che per "indirizzo" o per altri, ma proprio gli utilizzatori di certi dispositivi hanno familiarità con tale accezione di "luogo". Se non lo intendono così l'anziano e l'eremita che non conosce tali dispositivi, discorso diverso va fatto per coloro che li utilizzano e ne fanno anzi uno strumento principe della propria socia-

[34] http://www.treccani.it/vocabolario/luogo/ (ultimo accesso 4 giugno 2019).

[35] https://www.garzantilinguistica.it/ricerca/?q=luogo (ultimo accesso 4 giugno 2019).

[36] http://www.grandidizionari.it/Dizionario_Italiano/parola/L/luogo.aspx?query=luogo (ultimo accesso 4 giugno 2019).

lità. Per queste persone non è urlando in una piazza[37] che si fa sapere qualcosa al mondo, ma attraverso un social o una bacheca elettronica.

Insomma, mi sembra si possa dire che i luoghi digitali appartengono quantomeno alla penombra del significato di "luogo" nella lingua corrente. È forse un'accezione di "luogo" che inizialmente si profila come metaforica, ma nell'uso acquista una relativa diffusione e stabilità in grado di giustificare a mio avviso la sua collocazione nella penombra semantica del termine. Se è così, i luoghi digitali si collocano oltre il significato letterale (standard) di "luogo", ma non tanto lontano da esso da richiedere la creazione di nuovo diritto per via analogica. È configurabile infatti un'interpretazione estensiva del termine, che copra più fattispecie di quella standard.

Con quali ragioni? Almeno due: 1) il già detto uso crescente di questo e simili termini in chiave elettronica o digitale; 2) lo scopo della disciplina. In questa linea mi pare di poter leggere le motivazioni della sentenza di Cassazione[38]. Della prima ragione abbiamo rapidamente detto; aggiungiamo alcune considerazioni sulla seconda. La protezione della tranquillità delle persone richiede la sanzione di molestie e forme di disturbo recate in luoghi pubblici o aperti al pubblico. Se tale è lo scopo, a maggior ragione andranno sanzionate quelle condotte che hanno un'offensività superiore date le caratteristiche dello spazio elettronico e di un social network. Si noti che con ciò non sto offrendo degli argomenti teleologici o consequenzialisti per modificare il diritto esistente; offro degli argomenti teleologici che supportano un'interpretazione estensiva delle disposizioni vigenti.

Una molestia tramite social network raggiunge un più ampio numero di persone rispetto a quelle che fisicamente il molestatore potrebbe coinvolgere; ha una più facile replicabilità e ha un minor costo per il molestatore, il quale può diffondere e iterare le molestie quando voglia e dovunque si trovi purché dotato di un dispositivo funzionante. Per altro verso la non immediatezza fisica della comunicazione molesta la rende meno viva e plausibilmente meno forte sotto l'aspetto emotivo; se però lo scopo della disciplina è proteggere la tranquillità delle persone e non esporle a pubblico disturbo, andranno disincentivate le molestie tramite social network. Infatti, che incentivi darebbe il *non* considerare questi spazi come luoghi penalmente rilevanti? Premesso che la forza di condizionamento del diritto penale nelle dinamiche psicologiche degli agenti non deve essere sopravvalutata[39] – specie se è la forza delle

[37] A meno che non sia l'*agorà* virtuale di cui parla la sentenza in esame (punto 4.1 della motivazione in diritto).

[38] Si veda in particolare il punto 4.1 in diritto, pur se la Cassazione non parla espressamente di "interpretazione estensiva". Per un approfondimento sui profili della decisione giudiziale e i rapporti fra l'analogia e altri argomenti impiegati nella prassi giuridica, cfr. Canale e Tuzet (2019).

[39] Vedi Giunta (2000, 271).

sentenze penali e non della legislazione – credo si possa comunque prevedere che le molestie si moltiplicherebbero se fossero permesse o tollerate, dati da un lato i costi minimi per i molestatori (il solo tempo di postare le molestie creando magari un account fittizio) e dall'altro gli elevati benefici in termini di diffusione (date le caratteristiche del mezzo). Questo non significa, si badi, ricorrere allo scopo della disciplina per mostrare la somiglianza rilevante fra la molestia fisica e quella elettronica. Se si trattasse di questo si sarebbe sulla strada di un'analogia vietata dal diritto penale. La strada argomentativa è piuttosto quella di un'estensione interpretativa della disciplina testuale data la vaghezza di "luogo pubblico o aperto al pubblico" e date le considerazioni sullo scopo della disciplina.

Ciò per quanto riguarda il termine "luogo". Per quanto concerne l'espressione "pubblico o aperto al pubblico" credo si possa semplicemente aggiungere una considerazione di fatto: se le molestie vengono recate su un profilo "aperto" dotato di una bacheca siffatta (il cui contenuto è visibile a qualsiasi utente), non vi sono margini per dubitare di tale qualifica. Ammesso che sia un luogo, si deve ammettere che si tratta quantomeno di un luogo aperto al pubblico[40].

In conclusione, le molestie digitali e tramite social network sono plausibilmente più moleste di quelle fisiche. E sono recate in un luogo "pubblico o aperto al pubblico" se, data la vaghezza di questa espressione che ricorre nei testi normativi, è possibile interpretarla in modo estensivo tenendo conto dello sviluppo della lingua e delle nuove forme di comunicazione. Se però così non fosse, e fosse più corretta la definizione di "luogo" che ne conserva il significato tradizionale, a essere molesta sarebbe non solo la molestia in sé ma anche l'analogia a carico degli imputati, in quanto violazione del divieto di applicazione analogica delle norme incriminatrici.

Riferimenti bibliografici

Bobbio, N. (1938). *L'analogia nella logica del diritto*. Ed. 2006 a cura di P. Di Lucia, Torino: Giappichelli.

Bobbio, N. (1950). Scienza del diritto e analisi del linguaggio. *Rivista trimestrale di diritto e procedura civile*, *4*, pp. 342-367.

Bobbio, N. (1993). *Teoria generale del diritto*. Torino: Giappichelli.

Caiani, L. (1958). *Analogia*. In *Enciclopedia del diritto*, vol. 2. Milano: Giuffrè, pp. 348-378.

[40] Infatti la decisione in esame osserva che la sentenza di secondo grado "fa difetto nel dare conto della base fattuale" di tale qualifica (punto 4.2 in diritto), in quanto la Corte d'appello assume e non verifica che le espressioni moleste siano state inserite sulla pagina pubblica della vittima. Uno degli argomenti difensivi era appunto l'invio dei messaggi molesti tramite chat privata e non su pagina pubblica.

Canale, D., Tuzet, G. (2014). Sulla distinzione tra analogia e interpretazione estensiva nel ragionamento giuridico. *Materiali per una storia della cultura giuridica*, *44*, pp. 149-173.
Canale, D., Tuzet, G. (2016). What the legislature did not say. Legislative intentions and counterfactuals in legal argumentation. *Journal of Argumentation in Context*, *5*, pp. 249-270.
Canale, D., Tuzet, G. (2019). *La giustificazione della decisione giudiziale*. Torino: Giappichelli.
Carlizzi, G. (2012). *Contributi alla storia dell'Ermeneutica Giuridica Contemporanea*. Napoli: La Scuola di Pitagora.
Cooter, R., Mattei, U., Monateri, P.G., Pardolesi, R., Ulen, T. (2006). *Il mercato delle regole. Analisi economica del diritto civile*. Nuova ed. in 2 voll., Bologna: Il Mulino.
Di Giovine, O. (2010). Tra analogia e interpretazione estensiva. *Criminalia*, *2010*, pp. 355-366.
Di Lucia, P. (2003). *Normatività. Diritto linguaggio azione*. Torino: Giappichelli.
Durkheim, É. (2010). *Sociologie et philosophie*. Paris: Puf.
Elster, J. (1983). *Ulisse e le sirene. Indagini sulla razionalità e l'irrazionalità*. Trad. di P. Garbolino, Bologna: Il Mulino.
Ferrajoli, L. (1989). *Diritto e ragione. Teoria del garantismo penale*. Roma-Bari: Laterza.
Gianformaggio, L. (1987). Analogia. In *Digesto delle discipline privatistiche*, quarta ed., vol. 1. Torino: Utet, pp. 320-329.
Giunta, F. (2000). Quale giustificazione per la pena? Le moderne istanze della politica criminale tra crisi dei paradigmi preventivi e incanti scientistici. *Politica del diritto*, *31*, pp. 265-282.
Guastini, R. (1976). Completezza e analogia. Studi sulla teoria generale del diritto italiana del primo Novecento. *Materiali per una storia della cultura giuridica*, *6*, pp. 511-591.
Guastini, R. (2011). *Interpretare e argomentare*. Milano: Giuffrè.
Haack, S. (2018). The pragmatist tradition: Lessons for legal theorists. *Washington University Law Review*, *95*, pp. 1049-1082.
Hart, H.L.A. (1958). Positivism and the separation of law and morals. *Harvard Law Review*, *71*, pp. 593-629.
Hart, H.L.A. (1961). *The concept of law*. Terza ed. 2012 a cura di L. Green, Oxford: Oxford University Press.
Hassemer, W. (1997). Diritto giusto attraverso un linguaggio corretto? Sul divieto di analogia nel diritto penale. *Ars Interpretandi*, *2*, pp. 171-195.
Hesse, M.B. (1966). *Models and analogies in science*. Notre Dame: University of Notre Dame.
Kaufmann, A. (2001). Il ruolo dell'abduzione nel procedimento di individuazione del diritto. *Ars Interpretandi*, *6*, pp. 319-332.
Kaufmann, A. (2003). *Analogia e "natura della cosa". Un contributo alla dottrina del tipo*. Ed. a cura di G. Carlizzi, Napoli: Vivarium.
Labinaz, P., Sbisà, M. (2017). Credibilità e disseminazione di conoscenze nei social network, *Iride*, *30*, pp. 63-86.
Lakoff, G. (1987). *Women, fire and dangerous things. What categories reveal about the mind*. Chicago: Chicago University Press.

Levi, E.H. (1948). An introduction to legal reasoning. *The University of Chicago Law Review*, *15*, pp. 501-574.
Pitruzzella, G., Pollicino, O., Quintarelli, S. (2017). *Parole e potere. Libertà d'espressione, hate speech e fake news*. Milano: Egea.
Pollicino, O., Bertolini, E., Lubello, V. (a cura di) (2013). *Internet: regole e tutela dei diritti fondamentali*. Roma: Aracne.
Posner, R. (2006). Reasoning by analogy. *Cornell Law Review*, *91*, pp. 761-774.
Schauer, F. (2008). A critical guide to vehicles in the park. *New York University Law Review*, *83*, pp. 1109-1134.
Trujillo, I. (2014). *Sulla specificità del ragionamento giuridico. L'analogia nel* Civil *e nel* Common Law. Napoli: Editoriale Scientifica.
Tuzet, G. (2010). *Dover decidere. Diritto, incertezza e ragionamento*. Roma: Carocci.
Tuzet, G. (2011). La storia infinita. Ancora su analogia e interpretazione estensiva. *Criminalia*, *2011*, pp. 507-519.
Velluzzi, V. (2008). Interpretazione degli enunciati normativi, linguaggio giuridico, certezza nel diritto. *Criminalia*, pp. 493-507.
Velluzzi, V. (2013). *Le Preleggi e l'interpretazione*. Pisa: Ets.
Vogliotti, M. (2011). *Dove passa in confine? Sul divieto di analogia nel diritto penale*. Torino: Giappichelli.
Zaccaria, G. (1990). *L'arte dell'interpretazione. Saggi sull'ermeneutica giuridica contemporanea*. Padova: Cedam.
Ziccardi, G. (2015). *Internet, controllo e libertà. Trasparenza, sorveglianza e segreto nell'era tecnologica*. Milano: Raffaello Cortina.
Ziccardi, G. (2016). *L'odio online. Violenza verbale e ossessioni in rete*. Milano: Raffaello Cortina.

Harassing analogies? Social networks and criminal law

Starting from some remarks on the legal aspects of communication via social networks, the article deals with the (analogical?) extension of certain legal norms, designed for traditional communication, to communication through social networks. In particular, can a norm on harassment "in a public place or open to the public" be extended to social media communication? The legal problem with this is that our legal system forbids the analogical extension of incriminating norms, while their "interpretive" extension is not prohibited. The article presents an argument in favor of this second type of extension.

Keywords: analogy, communication, criminal law, extensive interpretation, social networks.

Giovanni Tuzet, Università Bocconi, Dipartimento di Studi Giuridici, via Roentgen 1, 20136 Milano, giovanni.tuzet@unibocconi.it

FILIPPO DOMANESCHI LUCA DE VITA SIMONA DI PAOLA

LA PRAGMATICA ☺ E ☹

Quando e quanto usiamo le emoticon su WhatsApp

1. Introduzione

Con *Computer-mediated communication* (CMC) si fa riferimento a tutte quelle forme di comunicazione in cui due o più interlocutori interagiscono tra loro tramite il supporto di un computer o di altri strumenti informatici (Herring, 1996). Una questione centrale nell'ambito della CMC riguarda il problema di stabilire in che misura gli strumenti informatici che mediano le interazioni verbali modificano le pratiche comunicative dei parlanti, dal punto di vista linguistico, psicologico e sociologico (Rice e Love, 1987). In una prospettiva linguistica, ad esempio, svariate ricerche hanno messo in evidenza che nelle interazioni verbali CMC, rispetto al linguaggio parlato faccia-a-faccia, è possibile osservare un ampio repertorio di variazioni nella produzione e comprensione linguistica sul piano sintattico, semantico e pragmatico (Herring, 1996). Questo articolo si inserisce nell'ambito della pragmatica della CMC, in quanto prende in esame in che misura il tipo di interazione verbale influisce sull'uso delle emoticon nelle conversazioni WhatsApp. In particolar modo, scopo del presente contributo è valutare se il tipo di interazione verbale e la natura conflittuale o conciliatoria di una conversazione WhatsApp influenzino la frequenza d'uso di emoticon con diverso valore semantico. In § 2, illustreremo brevemente le principali teorie sulla funzione pragmatica delle emoticon nella CMC. In § 3 presenteremo un lavoro sperimentale di Derks, Bos e Von Grumbkow (2007) e solleveremo alcune considerazioni critiche in merito. In § 4 presenteremo i dati di un lavoro sperimentale di replica e critica a Derks *et al*. (2007). In § 5 concluderemo sostenendo che nelle conversazioni di natura non conflittuale e negli scambi verbali di natura socio-emotiva, in cui, in altre parole, i partecipanti non sono chiamati a cooperare al fine del raggiungimento di uno scopo pratico, vi è una maggiore frequenza d'uso di emoticon facciali.

2. Le emoticon nella CMC

Le emoticon sono una serie di segni grafici utilizzati per arricchire e supportare le interazioni verbali nella *Computer-mediated communication* (Dresner e Herring, 2010). Possono configurarsi come segni iconici che rappresentano oggetti o simboli di varia natura. Nel vasto repertorio di emoticon realizzabili con i simboli di una tastiera, la tipologia più nota è costituita dalle icone di espressioni facciali[1]: ☺, ☹, 😐, ecc., che occorrono con maggiore frequenza in contesti CMC a modalità sincronica come la messaggistica istantanea e nei messaggi di posta elettronica (Baron, 2004; Huffaker e Calvert, 2005; Markman e Oshima, 2007; Provine, Spencer e Mandell, 2007).

2.1. *La funzione delle emoticon nella CMC*

La visione classica nell'ambito della CMC rende conto delle emoticon come di indicatori paralinguistici di stati affettivi ed emotivi. Rezabek e Cochenour (1998, 201), ad esempio, definiscono le emoticon come segnali visivi formati da simboli tipografici ordinari che rappresentano sentimenti o emozioni; Crystal (2001, 36), le caratterizza come combinazioni di caratteri della tastiera progettati per mostrare espressioni facciali emotive mentre Baron (2000, 42) le considera veri e propri marcatori emotivi alla stregua delle interiezioni.

Se, da un lato, una consistente letteratura attribuisce alle emoticon la funzione di veicolo di emozioni e stati d'animo dell'emittente di un messaggio, dall'altro, diversi autori, in tempi più recenti, hanno provato a rendere conto della più articolata funzione comunicativa di tali segni iconici. Ad esempio, Dresner e Herring (2010) trattano le emoticon come strumenti ausiliari alla comunicazione verbale mediata dal computer, volti a fare le veci di un complesso repertorio di comportamenti non verbali e di espressioni facciali che integrano a livello multimodale la comunicazione verbale faccia-a-faccia. Wang, Zhao Qiu e Zhu (2014) rendono conto del fatto che l'uso delle emoticon può influenzare la disponibilità di un destinatario ad accettare commenti e osservazioni negative o conflittuali. Thompsen e Foulger (1996), invece, sostengono che, sebbene l'uso delle emoticon sia generalmente inteso come funzionale alla attenuazione della percezione di ostilità di un messaggio connotato negativamente,

[1] In questa sede, scegliamo deliberatamente di non trattare la distinzione tra emoticon ed emoji. La differenza tra le due categorie di simboli, tuttavia, andrebbe considerata con maggiore attenzione in eventuali futuri studi sperimentali. In estrema sintesi: mentre con emoticon si fa riferimento a un repertorio di realizzazioni grafiche iconiche prodotte utilizzando i soli elementi di testo della tastiera, le emoji sono vere e proprie immagini che possono essere visualizzate sol tramite il supporto di un apposito software.

l'ostilità percepita in un messaggio può anche aumentare laddove le emoticon vengono utilizzate in associazione ad atti linguistici ostili. Non in ultimo, l'uso delle emoticon può influenzare altresì il contenuto di un messaggio, contribuendo a disambiguare diverse interpretazioni concorrenti e ugualmente plausibili di un medesimo enunciato (Walther e D'Addario, 2001). In altre parole, le emoticon costituiscono un segno grafico che, nella comunicazione mediata dal computer, adempie a un ampio repertorio di funzioni comunicative di natura non solo emotiva ma anche sociale.

Un'idea condivisa nella letteratura su emoticon e CMC è che, nella comunicazione mediata dal computer, le emoticon si facciano carico di quelle funzioni tipicamente svolte dalla comunicazione non verbale nelle interazioni faccia-a-faccia (Fisher, 2011). Questo vale in particolare per le icone facciali, le quali sembrano svolgere tre funzioni chiave già attribuite alle espressioni facciali da Ekman e Friesen (1969) nel contesto del Facial Action Coding System (Derks *et al*., 2007): primo, le espressioni facciali veicolano informazioni sullo stato emotivo di colui o colei che le produce; secondo, regolano l'interazione tra gli interlocutori (e.g., il turn-taking); terzo, rafforzano il legame di intimità tra gli interlocutori o, al contrario, alimentano il conflitto e il distacco (cfr. Harrison, 1973). Tali funzioni comunicative sono prevalentemente focalizzate sullo scambio di informazioni di natura sociale e affettiva, volte a determinare e regolare la relazione sociale ed emotiva degli interlocutori. Analogamente alle espressioni facciali nel linguaggio parlato, le emoticon nella CMC svolgono tali funzioni socio-emotive, fornendo indizi di natura visiva che integrano il testo di un messaggio scritto e contribuiscono a costruire e regolare la relazione tra gli interlocutori ai due capi dello strumento informatico (Thompson e Foulger, 1996; Rezabek e Cochenour, 1998).

L'esigenza di utilizzare emoticon in contesti di CMC sembra essere legata al fatto che le interazioni verbali mediate dal computer risultano più povere di segnali sociali (Fisher, 2011; Kiesler, Siegel e McGuire, 1984). Le interazioni verbali CMC che avvengono *in absentia*, vale a dire, senza la presenza fisica degli interlocutori, generano infatti un ambiente socialmente povero che può favorire una distanza psicologica tra gli interlocutori (Rutter, 1987). La carenza di segnali sociali e la conseguente maggiore distanza sociale tra gli interlocutori può altresì favorire un linguaggio più disinibito e informale, una maggiore tendenza a intraprendere scelte più affrettate e rischiose di quanto avverrebbe in interazioni faccia-a-faccia (Kiesler *et al*., 1984; Siegel, Dubrovsky, Kiesler e McGuire, 1986; Sproull e Kiesler, 1986), nonché una maggiore propensione per scambi comunicativi strumentali, ovvero, conversazioni orientate alla risoluzione di specifici compiti.

In definitiva, la CMC predispone un ambiente comunicativo carente di segnali che consentono di veicolare contenuti di natura sociale ed emotiva. In questo contesto, le emoticon sopperiscono a tale mancanza

facendo le veci delle espressioni facciali o di altri segnali non verbali di natura sociale ed emotiva nelle conversazioni faccia-a-faccia. Da un lato, dunque, la minore presenza sociale e la maggiore distanza psicologica tra gli interlocutori nella CMC producono contesti di interazione verbale meno familiari e amichevoli e più orientati a compiti specifici (Rice e Love, 1987). Dall'altro, l'uso delle emoticon può favorire la trasmissione di contenuti sociali e rendere le interazioni CMC meno strumentali e più orientate alla dimensione socio-emotiva. Tuttavia, l'esatta relazione che intercorre tra l'uso delle emoticon e il contesto sociale in una conversazione CMC rimane un tema ancora quasi del tutto inesplorato. In particolare, due domande risultano centrali in questo ambito di ricerca: primo, in quali tipi di contesti di interazione sociale CMC prevale l'uso delle emoticon? Secondo, quali tipi di emoticon vengono utilizzate nei diversi contesti interazionali CMC?

3. Emoticon e interazioni su Internet: Derks, Bos e Grumbkow (2007)

In un articolo del 2007 dal titolo "Emoticons and social interaction on the Internet: the importance of social context", Derks, Bos e Von Grumbkow presentano uno studio sperimentale volto a stabilire in che misura l'uso di emoticon e, in particolare, di icone emotive come le emoticon facciali, dipenda dal tipo di interazione avente luogo su Internet.

Secondo gli autori, quattro differenti ipotesi possono essere formulate rispetto alla frequenza d'uso delle emoticon nelle interazioni verbali su chat online. La prima ipotesi (Hyp 1) è che, poiché nelle interazioni faccia-a-faccia la frequenza di segnali emotivi non verbali aumenta nelle relazioni familiari e amichevoli (Wagner e Lee, 1999), è plausibile aspettarsi che i parlanti utilizzino più emoticon nelle conversazioni CMC in contesti socio-emotivi rispetto a contesti orientati alla risoluzione di compiti specifici.

Per la seconda ipotesi, Derks e colleghi partono da Lee e Wagner (2002), i quali hanno mostrato che le persone tendono maggiormente ad esprimere contenuti emotivi in contesti sociali positivi e collaborativi rispetto a contesti negativi di natura conflittuale. Alla luce di tale considerazione, Derks e colleghi ipotizzano dunque un effetto della valenza del contesto (Hyp 2): è più probabile osservare una maggiore frequenza d'uso di emoticon nei contesti positivi rispetto a quelli negativi. Sulla base delle due ipotesi principali di cui sopra, gli autori predicono inoltre due possibili interazioni tra i tipi di contesti e la loro valenza. Innanzitutto, terza ipotesi (Hyp 3), un utilizzo meno frequente di emoticon nei contesti negativi orientati a uno scopo, vale a dire in quei contesti in cui due interlocutori coinvolti nella risoluzione di un compito pratico si trovano in una circostanza conflittuale. Infine, quarta ipotesi (Hyp 4), poiché le icone facciali possono avere una valenza sia positiva (e.g., ☺)

che negativa (e.g., ☹), gli autori si aspettano di osservare una maggiore frequenza d'uso delle emoticon a valenza positiva nei contesti positivi e un uso più ricorrente di emoticon a valenza negativa nei contesti negativi, ove si genera conflitto o disaccordo.

Per verificare le ipotesi di cui sopra, Derks *et al*. (2007) somministrano a 158 studenti di scuola secondaria un questionario cartaceo individuale della durata di circa 15 minuti. In tale questionario, ai partecipanti vengono presentate brevi conversazioni estratte da alcune chat scritte al computer. Gli autori producono un disegno sperimentale 2x2, in cui vengono manipolati due aspetti delle conversazioni oggetto del questionario: la tipologia (i.e., conversazioni orientate a un compito vs. conversazioni socio-emotive) e la valenza delle interazioni (i.e., contesti positivi vs. negativi). Ad esempio, ai partecipanti venivano presentati esempi di conversazioni orientate al compito o socio-emotive. Nel primo caso, un gruppo di studenti discuteva al fine di suddividersi le mansioni per lo svolgimento di un compito scolastico; nel secondo caso, un gruppo di studenti si scambiava reciprocamente idee su un possibile regalo da acquistare per il compleanno di un amico. In entrambe le condizioni sperimentali, la conversazione poteva presentarsi come positiva, ovvero cooperativa, o negativa, ovvero conflittuale con opinioni divergenti. Il compito richiesto ai partecipanti consisteva nel rispondere ad ogni chat in un'apposita sezione. I partecipanti potevano rispondere in tre modi: (i) producendo un enunciato in reazione allo scambio conversazionale in oggetto; o (ii) selezionando un'emoticon tra una lista di 6; o (iii) optando per una combinazione di emoticon e testo verbale. Le sei emoticon a disposizione erano: la faccia sorridente, la faccia con il sorriso esteso, la faccia triste, la faccia con gli occhiali da sole, la faccia che strizza l'occhiolino e la faccia con le corna da diavolo. Derks e colleghi prendono in considerazione due variabili dipendenti: la frequenza d'uso delle emoticon e la valenza delle emoticon (positiva o negativa) utilizzate nelle quattro condizioni sperimentali – ovvero, contesti orientati al compito vs. socio-emotivi e contesti positivi vs. negativi.

Sulla base dei dati raccolti, gli autori mostrano che il tipo di contesto interazionale in cui ha luogo una conversazione CMC influenza significativamente l'uso di emoticon facciali da parte dei parlanti. In primo luogo, conformemente alla prima ipotesi di ricerca, nei contesti di natura socio-emotiva, sia positivi sia negativi, si osserva una maggiore frequenza d'uso di emoticon (sia positive sia negative) rispetto ai contesti orientati a un compito. Poiché il confronto tra la percentuale di emoticon utilizzate in contesti a valenza positiva e in quelli a valenza negativa non rivela alcuna differenza significativa, la seconda ipotesi sperimentale non trova conferma. Viceversa, l'interazione fra la tipologia e la valenza del contesto emerge come significativa: nei contesti negativi orientati a un compito i partecipanti utilizzano meno emoticon rispetto alle altre condizioni sperimentali; al contrario, nei contesti

negativi di tipo socio-emotivo si osserva una percentuale più alta di emoticon selezionate. L'ipotesi Hyp3, dunque, pare confermata. Infine, da un'analisi condotta sulla frequenza di emoticon a valenza positiva e a valenza negativa all'interno dei contesti positivi e negativi emerge che, conformemente all'Hyp4, nei contesti a valenza positiva vi è una maggiore frequenza d'uso di emoticon a valenza positiva mentre nei contesti negativi occorrono più frequentemente emoticon a valenza negativa.

Il lavoro di Derks e colleghi offre un primo approccio pioneristico allo studio della funzione comunicativa delle emoticon nelle conversazioni CMC. In particolare, lo studio si configura come un progetto di grande interesse per la pragmatica della CMC, dal momento che l'analisi della frequenza d'uso delle emoticon si concentra sulla funzione giocata dal tipo di scopo conversazionale e dalla natura cooperativa o conflittuale dello scambio verbale in atto. Nondimeno, il lavoro presenta alcuni punti deboli che rendono l'interpretazione dei risultati solo parzialmente soddisfacente.

In primo luogo, il metodo sperimentale adottato dagli autori risulta non pienamente ecologico. Ai partecipanti viene infatti chiesto di reagire a uno scambio verbale avente luogo su internet producendo un enunciato. Sia la lettura dello scambio chat, sia la produzione dell'enunciato da parte dei partecipanti, tuttavia, vengono prodotti e registrati su un questionario in forma cartacea. Tale soluzione riduce drasticamente la spontaneità e la naturalezza delle risposte dei partecipanti, prima di tutto, a causa della scarsa abitudine e disposizione a utilizzare emoticon in testi scritti su supporto cartaceo. In secondo luogo, la costruzione degli stimoli sperimentali risente di una variabile confondente a nostro avviso fatale dovuta a un'interferenza fra la tipologia del contesto sociale (socio-emotivo vs. orientato al compito) ed il ruolo del partner con cui avviene la comunicazione: nei contesti socio-emotivi il partner di interazione era rappresentato da un amico o un'amica, mentre nei contesti orientati al compito il partner era un compagno o una compagna di classe. Le due figure in questione generano una disparità di interpretazione della situazione comunicativa che può aver influenzato l'utilizzo delle emoticon in modo più sostanziale di quanto non abbia fatto la manipolazione sperimentale primaria, ovvero, la tipologia del contesto in sé.

La letteratura psicologica e linguistica su emoticons ed emoji nel corso degli ultimi anni ha riscosso un grande interesse. Numerosi sono i lavori che hanno indagato il ruolo di tali componenti segniche nella comunicazione mediata dal computer. Il presente lavoro ha uno scopo deliberatamente modesto: offrire un tentativo di replica e modifica del lavoro di Derks e colleghi. Proprio per questo, conformemente alla scelta degli autori di riferimento, in questo lavoro non verrà trattata, ad esempio, la distinzione tra emoticon ed emoji.

4. L'uso delle emoticon su WhatsApp: uno studio sperimentale

Il presente studio si propone come uno sviluppo del lavoro di Derks *et al*. (2007). In particolar modo, l'obiettivo di questo lavoro sperimentale è provare a comprendere se le conversazioni orientate allo scopo vs. socio-emotive con valenza positiva vs. negativa influenzino la frequenza d'uso di emoticon con diversa valenza (i.e., positiva e negativa) in contesti di interazione verbale più ecologici. Il presente studio si differenzia da Derks *et al*. (2007) in tre aspetti cruciali. In primo luogo, a differenza di Derks *et al*. (2007), viene qui preso in esame l'uso delle emoticon in conversazioni diffuse tramite l'applicazione di messaggistica istantanea WhatsApp. Le conversazioni oggetto di analisi non vengono dunque somministrate da uno sperimentatore tramite un questionario cartaceo, ma vengono diffuse on-line tramite un modulo fruibile direttamente dal telefono cellulare dei partecipanti. In secondo luogo, nel presente esperimento non è stato coinvolto un campione ristretto di studenti della scuola secondaria ma è stato preso in considerazione un più ampio campione di partecipanti volontari selezionati sul web e con diversa caratterizzazione individuale in termini di sesso, età, livello d'istruzione e confidenza con gli strumenti informatici. In terzo luogo, l'utilizzo di un'applicazione on-line ha consentito di mettere a disposizione dei partecipanti tutte le emoticon previste nella tastiera del loro smartphone, evitando di limitare le possibilità espressive alle sole sei icone facciali pre-selezionate da Derks e colleghi.

4.1. *Metodo e Design*

4.1.1. *Partecipanti*

137 partecipanti volontari hanno preso parte all'esperimento, diffuso on-line tramite Google Moduli, utilizzando il proprio Smartphone personale. 27 soggetti sono stati esclusi dal campione per aver completato l'esperimento solo in forma parziale. Complessivamente, 113 partecipanti [MA = 26.84; SD = 9.37; 86 F; 27 M] hanno portato a termine l'intero questionario. I partecipanti erano tutti madrelingua italiani. Il consenso informato è stato ottenuto tramite un modulo online da tutti i partecipanti.

4.1.2. *Stimoli e Procedura*

Sono state create 24 conversazioni fittizie (16 target, 8 filler) su temi ordinari di diversa natura. Ogni conversazione era costituita da quattro scambi di battute alternati tra due interlocutori. La presenza/assenza di emoticon positive/negative all'interno delle conversazioni è stata

randomizzata. Il genere degli interlocutori è stato bilanciato nei diversi item sperimentali (6m/m, 6f/f, 6m/f, 6f/m). Le conversazioni sono state presentate ai partecipanti sotto forma di screenshot di conversazioni WhatsApp.

Nella costruzione degli stimoli sperimentali sono state manipolate due *variabili indipendenti*, seguendo il design sperimentale 2x2 adottato da Derks *et al*. (2007): la tipologia del contesto conversazionale (orientato al compito vs. socio-emotivo) e la valenza (positiva vs. negativa) – esempio in Fig. 1. Le 16 conversazioni target, dunque, sono state distribuite rispettivamente nelle quattro condizioni sperimentali: 4 conversazioni orientate al compito (i.e., task-oriented, TO) con valenza positiva (TO+), 4 conversazioni orientate al compito con valenza negativa (TO–), 4 conversazioni socio-emotive (i.e., SE) con valenza positiva (SE+) e 4 socio-emotive con valenza negativa (SE–). Tale manipolazione sperimentale è stata operata con l'obiettivo di esaminare due *variabili dipendenti*: la frequenza d'uso generale delle emoticon e la frequenza d'uso di emoticon con valenza positiva e negativa al variare delle quattro condizioni sperimentali.

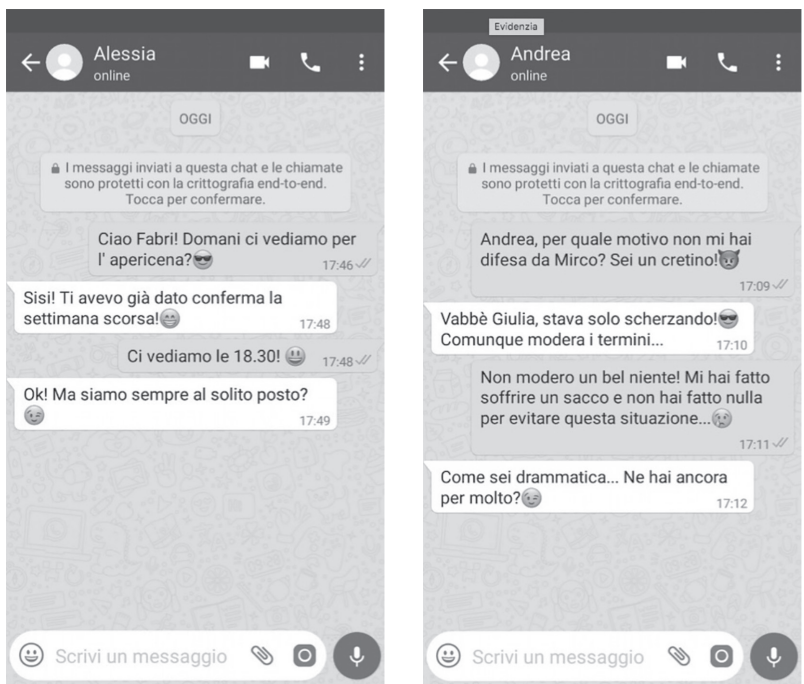

Fig. 1. Esempi di screenshot di due conversazioni presentate nel modulo sperimentale on-line. Le conversazioni sono caratterizzate da un'interazione fra soggetti di genere diverso. Lo screenshot a sinistra presenta una conversazione in condizione TO+, quello a destra un'interazione in condizione SE–.

Le 24 conversazioni sono state presentate ai partecipanti con ordine randomizzato. Ai partecipanti veniva richiesto di leggere attentamente ogni conversazione e di produrre una replica spontanea ad ognuna di esse, immedesimandosi nel ruolo di interlocutore/trice. Al termine di ogni conversazione è stata predisposta una barra di risposta dove i partecipanti potevano inserire una frase di replica costituita da (i) solo testo verbale, (ii) solo emoticon, o (iii) sia testo che emoticon. Ai partecipanti era consentito di utilizzare tutte le emoticon a disposizione nella tastiera del loro Smartphone. Non sono state fornite specifiche indicazioni sulle modalità di risposta per evitare un possibile effetto di interferenza sull'utilizzo o meno delle varie emoticon. L'utilizzo di icone emoticon in fase di risposta, dunque, è avvenuto in modo del tutto spontaneo. Le emoticon prodotte dai partecipanti sono state codificate tenendo conto della frequenza d'uso generale e della loro valenza (positiva o negativa) utilizzando il sistema di codifica Unicode. In particolare, è stato adottato un criterio normativo di categorizzazione delle emoticon: tutte le emoticon dalla U+1F600 alla U+1F60F sono state valutate come positive. Invece, tutte le emoticon dalla U+1F612 alla U+1F616 e dalla U+1F61E alla U+1F633 sono state considerate di valore negativo.

4.1.3. *Analisi statistiche*

I dati raccolti sono stati trattati in due macro-analisi statistiche volte ad esaminare quanto segue:

(i) La presenza di emoticon nelle risposte dei partecipanti in base alla condizione sperimentale;

(ii) La valenza positiva o negativa delle emoticon utilizzate dai partecipanti in base alla condizione sperimentale.

La prima analisi era finalizzata a verificare se il tipo di contesto conversazionale (i.e., condizione TO+, TO–, SE+, SE–) influenzasse il numero di emoticon utilizzate dai partecipanti nelle loro risposte (i.e., frequenza d'uso). A tal fine, due procedure statistiche sono state utilizzate: (i) un Modello Lineare Generalizzato (i.e., GLM, regressione classica) in cui la condizione è stata inserita come variabile indipendente a quattro livelli (TO+, TO–, SE+, SE–) e la frequenza d'uso delle emoticon è stata trattata come una variabile dipendente continua. Tale statistica è stata poi seguita da confronti multipli fra i livelli del fattore condizione con correzione di Bonferroni; (ii) un'analisi del chi-quadrato con la correzione di continuità di Yates, in cui la frequenza delle emoticon è stata trattata come una variabile categorica e codificata come 1 o 0 relativamente alla presenza o assenza di emoticon in base alla condizione sperimentale.

La seconda analisi era finalizzata a verificare variazioni di valenza delle emoticon (positiva o negativa) al variare delle condizioni sperimentali. Pertanto, due analisi del chi-quadrato con la correzione

di continuità di Yates sono state condotte. Nella prima, si misurava la frequenza d'uso della valenza delle emoticon fra condizioni. Nella seconda, invece, l'analisi del chi-quadrato è stata condotta all'interno di ciascuna condizione per analizzare la frequenza della valenza (positiva o negativa) delle emoticon in ciascun contesto conversazionale (i.e., TO+, TO–, SE+, SE–).

4.2. *Risultati*

(i) Frequenza d'uso delle emoticon

La frequenza d'uso di emoticon è pari in media a 0.51(0.44) in condizione SE+, 0.35(0.76) in condizione TO+, 0.25(0.60) in SE– e 0.20(0.78) in TO–, Fig. 2.

La statistica GLM ha rivelato che tali differenze sono significative, come suggerito dall'effetto significativo della condizione (c^2 (3) = 58.2; p < .001). Il numero di emoticon utilizzato dai partecipanti, dunque, varia significativamente in base alla condizione sperimentale. In particolare, i confronti multipli post-hoc hanno rivelato che la frequenza d'uso di emoticon varia significativamente in condizione SE– vs. SE+ (t = –5.87; p <.001), SE+ vs. TO– (t = 7.07; p < .001), SE+ vs. TO+ (t = 3.51; p < .005) e TO– vs. TO+ (t = –3.56; p < .005). Viceversa, la frequenza d'uso delle emoticon in condizione SE– non è risultata variare significativamente rispetto alle condizioni TO– (t = 1.20; p = n.s.) e TO+ (t = –2.36; p = n.s.).

Tale pattern di risultati è stato ulteriormente confermato anche nel caso in cui la presenza di emoticon è stata trattata come una variabile categorica (1 = presenza; 0 = assenza). In questo caso, infatti, i partecipanti hanno utilizzato emoticon il 39.15% delle volte in condizione SE+, il 31.19% delle volte in TO+ e il 18.14% e 17.47% delle volte rispettivamente in condizione TO– e SE–. L'analisi del chi-quadrato ha confermato che la frequenza d'uso differisce significativamente in tutti i confronti fra condizione, eccetto TO– vs. SE– (TO+ vs. TO–: c^2 (1) = 20.02; p < .0001; TO+ vs. SE+: c^2 (1) = 5.94; p < .05; TO+ vs. SE–: c^2 (1) = 22.35; p < .0001; SE+ vs. SE–: c^2 (1) = 51.27; p < .0001; TO– vs. SE–: c^2 (1) = 0.03; p = n.s.).

(ii) Valenza delle emoticon

Le statistiche descrittive hanno rivelato una prevalenza di emoticon con valenza positiva in tutte le condizioni tranne in condizioni SE-, dove invece sembra prevalere la presenza di emoticon con valenza negativa – si veda fig. 3.

L'analisi del chi-quadrato fra condizioni ha confermato quanto emerso dai dati descrittivi, rivelando che le differenze di valenza delle emoticon fra le quattro condizioni sono significative (TO+ vs. TO–: c^2 (1) =

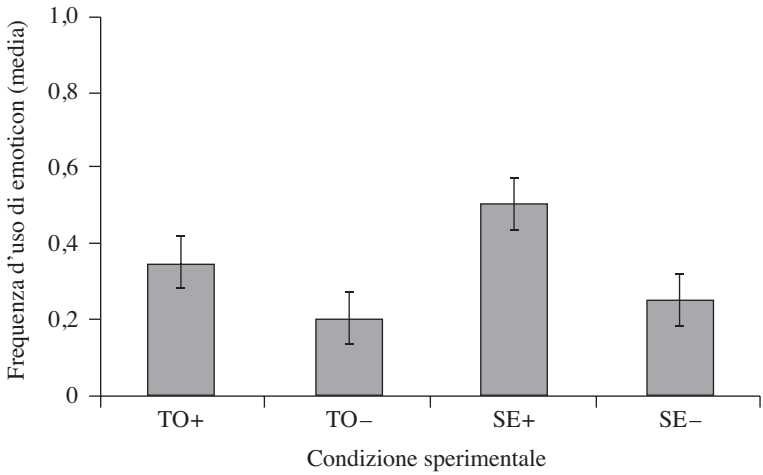

FIG. 2. Frequenza media dell'uso di Emoticon in ciascuna condizione sperimentale: TO+ = contesto Task-oriented positivo; TO– = contesto Task-oriented negativo; SE+ = contesto socio-emotivo positivo; SE– = contesto socio-emotivo negativo.

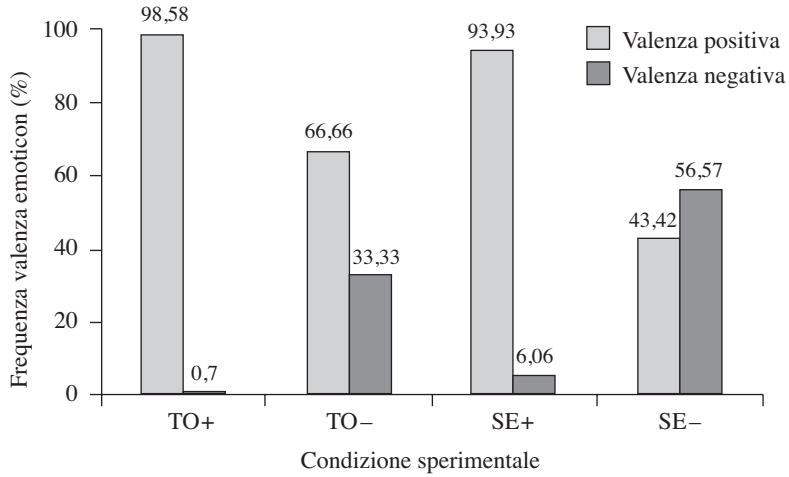

FIG. 3. Frequenza di emoticon in percentuale con valenza positiva e negativa in ciascuna condizione sperimentale: TO+ = contesto Task-oriented positivo; TO– = contesto Task-oriented negativo; SE+ = contesto socio-emotivo positivo; SE– = contesto socio-emotivo negativo.

46.43; $p < .0001$; TO+ vs. SE+: c^2 (1) = 4.78; $p < .05$; TO+ vs. SE–: c^2 (1) = 91.36; $p < .0001$; TO– vs. SE+: c^2 (1) = 29.52; $p < .0001$; TO– vs. SE–: c^2 (1) = 7.66; $p < .005$; SE+ vs. SE–: c^2 (1) = 74.49; $p < .0001$).

In linea con il pattern di risultati di cui sopra, l'analisi del chi-quadrato all'interno di ciascuna condizione ha rivelato che la frequenza di emoticon con valenza positiva differisce significativamente dalla frequenza

di emoticon con valenza negativa – con prevalenza significativa di emoticon con valenza positiva (Fig. 3) – all'interno di ciascuna condizione sperimentale eccetto SE– in cui la differenza fra valenza positiva e negativa non è significativa (valenza positiva vs. negativa in TO+: c^2 (1) = 268.12; $p < .0001$; TO–: c^2 (1) = 16.69; $p < .0001$; SE+: c^2 (1) = 251.34; $p < .0001$; SE–: c^2 (1) = 2.13; p = n.s.).

5. Discussione

Scopo di questo lavoro era stabilire se, in contesti verosimili ed ecologici di interazione verbale CMC, il tipo di interazione e la natura cooperativa o conflittuale dello scambio verbale influenzino (i) la frequenza d'uso di emoticon in generale e (ii) la frequenza d'uso di emoticon con valenza positiva o negativa in particolare. Tale studio si propone quindi come un lavoro di replica e sviluppo del progetto di ricerca proposto da Derks *et al*. (2007).

In riferimento alla prima ipotesi sperimentale, i dati raccolti in questo lavoro sembrano essere in linea con quelli di Derks e colleghi: nel nostro paradigma sperimentale le persone tendono a utilizzare le emoticon (indipendentemente dalla loro valenza) con maggiore frequenza in contesti conversazionali di natura socio-emotiva rispetto ai contesti di interazione orientati a un compito. Questo dato può essere spiegato facendo riferimento non solo all'idea che nelle relazioni di natura familiare e amichevole la frequenza di segnali emotivi non verbali aumenta (Wagner e Lee, 1999), ma anche alla maggiore tendenza dei parlanti a mostrare le proprie emozioni in presenza di persone con le quali si intrattengono relazioni amicali piuttosto che in presenza di colleghi o figure istituzionali (Fussell, 2002).

Per quanto riguarda la seconda ipotesi sperimentale di Derks e colleghi, invece, è interessante rilevare una discrepanza nei dati raccolti nel presente studio. Infatti, mentre Derks *et al*. non osservano alcuna differenza nella frequenza d'uso delle emoticon in contesti positivi versus quelli negativi, dall'analisi dei nostri dati è emersa una differenza significativa tra i due contesti, con una maggiore frequenza d'uso di emoticon nei contesti positivi cooperativi rispetto a quelli negativi conflittuali. Questo risultato è in linea con gli studi di Lee e Wagner (2002), i quali hanno evidenziano che nelle interazioni faccia-a-faccia è possibile osservare una maggiore tendenza da parte degli interlocutori ad esprimere più emozioni in contesti positivi rispetto a quelle circostanze conflittuali in cui i parlanti sono portati a non adottare un comportamento cooperativo. Una possibile spiegazione per la generale reticenza dimostrata dai partecipanti a manifestare emozioni nei contesti negativi può risiedere nel fatto che l'espressione di emozioni negative in tali circostanze comunicative è generalmente percepita come meno

appropriata dagli interlocutori, in quanto presuppone una stretta relazione di intimità tra i partecipanti allo scambio verbale (Chaikin e Derlega, 1974). Nel complesso, dunque, i dati raccolti suggeriscono che anche nelle conversazioni CMC si osserva un pattern comportamentale affine a quello delle interazioni faccia-a-faccia: i contesti positivi e cooperativi favoriscono l'espressione di emozioni e di comportamenti non verbali codificati nella forma di icone emoticon.

Questo pattern rappresenta un dato interessante e non del tutto prevedibile. Nelle interazioni CMC, infatti, complessivamente prevale una ridotta spontaneità nella produzione linguistica e un maggiore controllo da parte dei parlanti sull'espressione delle loro emozioni. Tale controllo può senz'altro essere legato allo sfasamento temporale che caratterizza la scrittura dei messaggi nella messaggistica istantanea, il quale mette a disposizione degli interlocutori una risorsa temporale in fase di produzione linguistica utile per poter elaborare, filtrare, sopprimerne o eventualmente codificare opportunamente le emozioni negative e di elaborare, con maggiore consapevolezza, messaggi più favorevoli ai fini della conversazione e più cooperativi (cfr. Walther, 2011). La maggiore frequenza d'uso di emoticon nei contesti positivi, rilevata nel nostro studio, può essere dovuta al diverso supporto di fruizione del materiale sperimentale utilizzato nei due studi: cartaceo in un caso, digitale nell'altro. Poiché il divertimento e la giocosità percepite durante una conversazione di messaggistica istantanea sono in grado di aumentare e rafforzare la complicità e la connessione tra due interlocutori legati da un rapporto di amicizia pre-esistente, è possibile che l'uso del supporto informatico abbia maggiormente favorito, rispetto a quello cartaceo, l'espressione di stati emotivi in forma di emoticon nelle interazioni positive e cooperative.

In riferimento alla terza ipotesi sperimentale, i risultati del nostro lavoro offrono una replica dei dati raccolti da Derks *et al*. (2007): nei contesti negativi orientati a uno scopo, si osserva una minore frequenza d'uso di emoticon rispetto alle altre condizioni sperimentali. Si tratta di un risultato che può essere spiegato chiamando in causa due processi che hanno luogo nei contesti di interazione verbale orientati all'adempimento di un compito in cui gli interlocutori si relazionano in modo conflittuale o non cooperativo: primo, all'interno di contesti operativi di questo tipo, l'interesse principale degli interlocutori è focalizzato sulla risoluzione di un problema e i loro sforzi di natura cognitiva e relazionale sono perlopiù finalizzati alla individuazione di una strategia di azione condivisa da tutti gli interlocutori. Questo può chiaramente aver favorito l'espressione di contenuti linguistici a discapito della manifestazione di segnali emotivi in forma iconica. Secondo, la natura conflittuale dell'interazione fa si che i parlanti privilegino una forma espressiva quanto più chiara possibile, favorevole e funzionale alla risoluzione di un conflitto, nel tentativo di ridurre possibili ambiguità o interpretazioni compromettenti. In tal

senso, le emoticon costituiscono un elemento segnico confondente che può sfavorire lo svolgimento di una conversazione su un piano razionale portandolo, invece, su un livello emotivo.

Infine, per quanto riguarda la quarta ipotesi sperimentale al vaglio della nostra analisi, i dati raccolti rivelano un'ulteriore discrepanza con i risultati ottenuti da Derks *et al*. (2007). Nello studio di Derks e colleghi si osserva infatti che la frequenza d'uso delle emoticon con valenza positiva (e.g., ☺) aumenta nei contesti con valenza positiva e che, viceversa, le emoticon con valenza negativa (e.g., ☹) vengono prodotte più frequentemente nei contesti a valenza negativa (a prescindere dalla loro tipologia). I dati raccolti nel nostro studio, invece, rivelano che le emoticon con valenza positiva vengono generalmente utilizzate con maggiore frequenza in tutti i contesti tranne in quelli di natura socio-emotiva con valenza negativa (i.e., SE–). In tali contesti prevale invece un uso di emoticon con valenza negativa rispetto a quelle con valore positivo. Pertanto, rispetto ai risultati di Derks *et al*. (2007), il presente studio sembra suggerire che le emoticon positive vengano prodotte con maggiore frequenza di quelle negative nei contesti negativi orientati a un compito (TO–). Si tratta di un risultato coerente con le conclusioni tratte in relazione alla terza ipotesi di lavoro: utilizzare emoticon negative in contesti non cooperativi orientati a uno scopo potrebbe favorire l'inquadramento dello scambio verbale in una cornice emotiva e conflittuale, piuttosto che razionale, dialettica e conciliatoria, favorendo il conflitto e compromettendo, in questo modo, il perseguimento dello scopo comune oggetto dello scambio verbale.

6. Conclusioni

Lo studio sperimentale presentato in questo lavoro ci consente di mettere in evidenza due risultati rilevanti rispetto ai dati preliminari raccolti da Derks *et al*. (2007). Primo, i soggetti tendono a produrre più emoticon in contesti di CMC positivi rispetto a quelli conflittuali e non cooperativi. Ciò correla con la tesi di Lee e Wagner (2002), secondo i quali la stessa dinamica ha luogo nelle interazioni verbali faccia-a-faccia. Secondo, nei contesti orientati a un compito in cui tra i parlanti si instaura una relazione negativa di natura conflittuale, i partecipanti alla conversazione tendono a produrre più emoticon con valenza positiva: poiché in un contesto conflittuale orientato allo svolgimento di un compito la priorità è quella di risolvere un problema e individuare una strategia di azione condivisa dagli interlocutori, l'uso delle emoticon con valenza negativa rischia di favorire l'insorgenza di un conflitto e di compromettere il raggiungimento dello scopo di comune interesse.

Il principale limite imputabile a questo lavoro sperimentale è riconoscibile nell'impossibilità di valutare il ruolo giocato dalle variabili

individuali dei soggetti. Un esempio su tutti: è ragionevole prevedere che il genere del partecipante possa costituire un predittore della frequenza di produzione di emoticon di diversa valenza al variare delle condizioni sperimentali. Un'indagine di questo tipo, tuttavia, è compromessa dalla numerosità esigua del nostro campione di riferimento e, ancor più, dalla assenza di una adeguata proporzione tra il numero di partecipanti uomini e donne. Futuri eventuali sviluppi del lavoro dovranno controllare più accuratamente tale variabile considerando altresì l'introduzione di scale di misura volte a stabilire se la tendenza alla produzione di un linguaggio emotivo tramite icone emoticon correli con variabili individuali di natura psico-sociale. La provenienza culturale dei partecipanti all'esperimento, infine, è un altro fattore cruciale al quale è ragionevole riconoscere un'influenza nella frequenza d'uso delle emoticon. Studi futuri dovranno considerare con maggiore attenzione, insieme ad altri fattori psico-sociali, come varia in una prospettiva cross-culturale la tendenza d'uso di emoticon ed emoji.

Il lavoro sperimentale condotto nel presente studio si configura come un tentativo di sviluppare una linea di ricerca nella prospettiva di una pragmatica della CMC. L'influenza delle emoticon sui processi pragmatici nell'ambito della comunicazione mediata dal computer è uno spazio di ricerca in larga parte ancora inesplorato che coinvolge una molteplicità di quesiti sul rapporto tra CMC e contesto d'uso. Per esempio, in che misura le emoticon contribuiscono a regolare i turni di avvicendamento di una conversazione CMC? Nella comprensione degli atti linguistici, le emoticon hanno un ruolo analogo a quello delle espressioni facciali nella comunicazione faccia-a-faccia? Soprattutto, l'uso di emoticon favorisce la velocità e la facilità di trasmissione delle intenzioni comunicative nella comunicazione mediata dal computer? Per il momento i dati raccolti suggeriscono che la disponibilità d'uso di segnali iconici quali le emoticon è mediata da due fattori che, all'interno di una teoria griceana della conversazione (Grice, 1975), si configurano come fattori chiave: il tipo di scopo conversazionale e la natura cooperativa o conflittuale di uno scambio verbale. Solo un'indagine sperimentale più approfondita potrà aiutare a caratterizzare il reale contributo pragmatico di questa forma di comunicazione non verbale ☺.

Riferimenti bibliografici

Baron, N.S. (2000). *Alphabet to email: how written English evolved and where it's heading*. London-New York: Routledge.

Baron, N.S. (2004). See you online: Gender issues in college student use of instant messaging. *Journal of language and social psychology*, *23*, 4, pp. 397-423.

Chaikin, A.L., Derlega, V.J. (1974). Variables affecting the appropriateness of self-disclosure. *Journal of Consulting and Clinical Psychology*, *42*, 4, p. 588.

Crystal, D. (2001). *Language and the Internet*. Cambridge, UK: Cambridge University Press.

Derks, D., Bos, A.E., Von Grumbkow, J. (2007). Emoticons and social interaction on the Internet: the importance of social context. *Computers in human behavior*, *23*, 1, pp. 842-849.

Dresner, E., Herring, S.C. (2010). Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication theory*, *20*, 3, pp. 249-268.

Ekman, P., Friesen, W.V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage and codings. *Semiotica*, *1*, p. 4997.

Fischer, A. (submitted) Gender and emotion in face-to-face and computer-mediated communication.

Fussell, S.R. (2002). The verbal communication of emotions: Introduction and Overview. In S.R. Fussell (a cura di), *The verbal communication of emotion: Interdisciplinary perspectives*. Mahwah, NJ: Lauwrence Erlbaum associates publisher, pp. 1-16.

Grice, H.P. (1975). Logic and conversation. In P. Cole e J. Morgan (a cura di), *Syntax and semantics*, Vol. 3. New York: Academic Press, pp. 41-58.

Harrison, R.P. (1973). Nonverbal communication. In I.S. Pool, W. Schramm, N. Maccoby, F. Fry, E. Parker e J.L. Fern (a cura di), *Handbook of communication*. Chicago: Rand McNally, pp. 93-115.

Herring, S.C. (a cura di) (1996). *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives*, Vol. 39. John Benjamins Publishing.

Huffaker, D.A., Calvert, S.L. (2005). Gender, identity, and language use in teenage blogs. *Journal of computer-mediated communication*, *10*, 2, JCMC10211.

Kiesler, S., Siegel, J., McGuire, T.W. (1984). Social psychological aspects of computer-mediated communication. *American psychologist*, *39*, 10, p. 1123.

Lee, V., Wagner, H. (2002). The effect of social presence on the facial and verbal expression of emotion and the interrelationships among emotion components. *Journal of Nonverbal Behavior*, 26, 1, pp. 3-25.

Markman, K.M., Oshima, S. (2007). Pragmatic play? Some possible functions of English emoticons and Japanese kaomoji in computer-mediated discourse. In *Association of Internet Researchers Annual Conference*, Vol. 8.

Provine, R.R., Spencer, R.J., Mandell, D.L. (2007). Emotional expression online: Emoticons punctuate website text messages. *Journal of Language and Social Psychology*, *26*, 3, pp. 299-307.

Rezabek, L., Cochenour, J. (1998). Visual cues in computer-mediated communication: Supplementing text with emoticons. *Journal of Visual Literacy*, *18*, 2, pp. 201-215.

Rice, R.E., Love, G. (1987). Electronic emotion: Socioemotional content in a computer-mediated communication network. *Communication research*, *14*, 1, pp. 85-108.

Rutter, D.R. (1987). *International Series in Experimental Social Psychology*, *Vol. 15. Communicating by Telephone*. Elmsford, NY: Pergamon Press.

Siegel, J., Dubrovsky, V., Kiesler, S., McGuire, T.W. (1986). Group processes in computer-mediated communication. *Organizational behavior and human decision processes*, *37*, 2, pp. 157-187.

Sproull, L., Kiesler, S. (1986). Reducing social context cues: Electronic mail in organizational communication. *Management science*, *32*, 11, pp. 1492-1512.

Thompsen, P.A., Foulger, D.A. (1996). Effects of pictographs and quoting on flaming in electronic mail. *Computers in Human Behavior*, *12*, 2, pp. 225-243.
Wagner, H., Lee, V. (1999). Facial Behavior Alone and in the Presence of Others. In P. Philippot, R.S. Feldman e E.J. Coats (a cura di), *Studies in emotion and social interaction. The social context of nonverbal behavior*. New York: Cambridge University Press; Paris, France: Editions de la Maison des Sciences de l'Homme, pp. 262-286.
Walther, J.B. (2011). Theories of computer-mediated communication and interpersonal relations. *The handbook of interpersonal communication*, *4*, pp. 443-479.
Walther, J.B., D'Addario, K. (2001). The impacts of emoticons on message interpretation in computer-mediated communication. *Social science computer review*, *19*, 3, pp. 324-347.
Wang, W., Zhao, Y., Qiu, L., Zhu, Y. (2014). Effects of emoticons on the acceptance of negative feedback in computer-mediated communication. *Journal of the Association for Information Systems*, *15*, 8, p. 454.

The pragmatics of :-) and :-(When and how much we use emoticons on WhatsApp

Emoticons are graphic signs used to enrich and support computer-mediated verbal interactions (CMC). The most famous type of emoticon is represented by the icons of facial expressions: :-), :-(, ;-), etc., that most frequently occur in instant messagging and e-mails. Falling within the scope of the pragmatics of CMC, this paper examines the extent to which the aim of a conversation and the conflicting or cooperative nature of a conversation influence the frequency of using emoticons (with either a positive or a negative value) on WhatsApp. The data collected from a behavioural experiment support some preliminary conclusions: WhatsApp users tend to use emoticons most frequently in positive and cooperative contexts as well as in socio-emotional verbal exchanges. Even though emoticons are used less frequently during conversations that are geared towards the accomplishment of a given goal, it is in such contexts that the emoticons with a positive value (e.g., :-)) occur most frequently.

Keywords: emoticons, computer mediated communication, pragmatics, contextual dependency,

Filippo Domaneschi, Università di Genova, Laboratory of Language and Cognition, Via Balbi 30, 16126 Genova, filippo.domaneschi@unige.it

Luca De Vita, Università di Genova, Studente, LucaDevita1996@hotmail.com

Simona Di Paola, Università di Genova, DISFOR, Corso Podestà 2, 16128 Genova, simona.dipaola@edu.unige.it

MARIA GRAZIA ROSSI JULIA MENICHETTI

SALUTE E PARTECIPAZIONE: FACEBOOK COME STRUMENTO EDUCATIVO PER IL COINVOLGIMENTO ATTIVO DELLA PERSONA CON DIABETE

1. Introduzione

L'utilizzo del web e dei social media come spazio in cui procacciare informazioni e supporto nell'ambito della salute è in continua crescita (Hamm *et al*., 2013). In modo parallelo, si allarga il dibattito attorno all'efficacia dei comportamenti di ricerca, condivisione e discussione di informazioni mediche online. Moorhead e colleghi (Moorhead *et al*., 2013) hanno fatto riferimento ai social media nei termini di una nuova dimensione dell'assistenza sanitaria, facendo il punto rispetto ai benefici e alle criticità che riguardano l'utilizzo di questa nuova frontiera della comunicazione. Accanto a questioni enormi relative ai problemi di privacy e riservatezza generati dalla condivisione e dall'immagazzinamento di dati sensibili, la principale criticità riguarda, come è ovvio, la questione della qualità delle informazioni ricercate in rete, il modo in cui queste informazioni vengono valutate e adattate alla propria condizione di malattia e, di conseguenza, il tipo di decisioni che fanno seguito a una tale consultazione. In relazione ai benefici, gli studi effettuati associano l'uso dei social media per la comunicazione sanitaria a un aumento delle interazioni con altri utenti e del supporto ricevuto da questi ultimi, come pure a un ampliamento della popolazione capace di usufruire di informazioni mediche complesse e, almeno in alcuni casi, personalizzate rispetto alla propria condizione di malattia (Markham, Gentile e Graham, 2017).

Smailhodzic, Hooijsma, Boonstra e Langley (2016) hanno condotto una rassegna sistematica della letteratura per fare il punto sulla ricerca empirica relativamente (1) al modo in cui i social media vengono usati dai pazienti e (2) all'effetto che l'informazione e le interazioni on-line producono all'interno della relazione terapeutica con i medici. Emergono dati interessanti: le informazioni ricercate e l'interazione on-line tra gli utenti vengono percepite dai pazienti stessi e dalle loro famiglie come complementari rispetto a quelle ricevute all'interno della consultazione medica; per esempio, i pazienti trovano benefici dalla condivisione di

informazioni che riguardano il vissuto di altri pazienti (e delle loro famiglie) che, trovandosi in situazioni analoghe di malattia, hanno già avuto modo di applicare (e adattare) le raccomandazioni tecniche e generiche ricevute dai medici nella vita concreta di ogni giorno (Colineau e Paris, 2010; Gómez-Zúñiga, Fernandez-Luque, Pousada, Hernández-Encuentra e Armayones, 2012; Kofinas *et al*., 2014). Soltanto due studi riportano un effetto negativo che spinge i pazienti a cambiare il proprio medico per via di discussioni on-line (Wicks *et al*., 2010), o in seguito alle reazioni negative espresse dal medico stesso rispetto all'opportunità della richiesta di un secondo parere suggerita on-line da altri pazienti (Rupert *et al*., 2014). Complessivamente, l'uso dei social media sembra avere effetti positivi all'interno della consultazione clinica favorendo una comunicazione più equa tra medico e paziente: una migliore comprensione della propria condizione di malattia avrebbe il merito di rendere maggiormente condivisa la scelta tra medico e paziente rispetto alle opzioni terapeutiche a disposizione. Un uso appropriato dei social media potrebbe dunque favorire il coinvolgimento attivo del paziente nel percorso di cura (§ 2) e contribuire a favorire un modello di assistenza sanitaria centrato sul paziente (*patient-centred healthcare*) (Markham *et al*., 2017; Rozenblum e Bates, 2013), migliorando gli esiti clinici e le condizioni di salute dei pazienti (Parchman, Zeber e Palmer, 2010).

Non è facile determinare quali fattori specifici e relativi all'uso dei social media incidano sul miglioramento delle condizioni di salute dei pazienti (Househ, Borycki e Kushniruk, 2014). A questo proposito, alcuni studiosi hanno notato che è necessario distinguere più accuratamente all'interno della categoria generica dei social media tra progetti collaborativi come Wikipedia, blog o microbloc come Wordpress e Twitter, piattaforme di contenuti come YouTube, siti di social network come Facebook o tra mondi virtuali come Second Life (Hamm *et al*., 2013). Allo stesso tempo, occorre anche effettuare distinzioni rispetto alle differenti modalità d'uso possibili per ciascuna delle categorie specifiche appena menzionate. Riconoscendo una tale complessità, in questo contributo abbiamo scelto di focalizzare la nostra attenzione su un social network specifico (Facebook) e su una malattia cronica specifica (il diabete). Esploreremo non tanto il modo in cui gli utenti ricercano/condividono informazioni sul web e su Facebook in generale; analizzeremo, più nello specifico, se e come l'utilizzo di un social network come Facebook possa diventare, nel contesto della salute, uno strumento educativo per il coinvolgimento attivo dei pazienti (e delle loro famiglie).

Nel prossimo paragrafo (§ 2), giustifichiamo la scelta di prendere in considerazione l'ambito delle malattie croniche (e del diabete) per esplorare preliminarmente il modo in cui Facebook viene utilizzato nel contesto della salute in Italia. Sottolineamo perché l'educazione e il coinvolgimento attivo sono stati indicati come fattori determinanti per favorire la gestione consapevole e autonoma del percorso di cura

da parte dei pazienti, e come questo abbia ricadute rilevanti sul piano degli esiti clinici.

Nel § 3 introduciamo la metodologia dello studio che abbiamo condotto per determinare in che modo Facebook è utilizzato dai pazienti diabetici e dalle loro famiglie. Abbiamo organizzato lo studio in due fasi: nella prima fase abbiamo somministrato un questionario per comprendere i bisogni informativi che spingono i pazienti con diabete a cercare/condividere informazioni sul web e a discutere questioni legate al diabete all'interno di gruppi su Facebook, comunità on-line frequentate da persone con diabete e dai loro familiari. Inoltre, abbiamo utilizzato il questionario per misurare il livello di coinvolgimento attivo e di alfabetizzazione sanitaria dei pazienti con diabete che usano Facebook. Nella seconda fase abbiamo analizzato i temi e le interazioni generate dai post pubblicati su Facebook all'interno di un gruppo chiuso specifico: "Tutti i Diabetici Uniti in Rete" (TIDUIR).

Nel § 5 presentiamo i risultati del questionario (§ 4) e offriamo un'analisi qualitativa delle interazioni comunicative pubblicate all'interno del gruppo Facebook TIDUIR. Evidenziamo le potenzialità di tali interazioni, basate su una condivisione di informazioni e di esperienze tra utenti che condividono un percorso di malattia analogo. Concludiamo sostenendo che in questo contesto il percorso educativo sembra fondato sul supporto tra pari (*peer support*); il paziente che condivide informazioni e discute on-line agisce in maniera collaborativa (e altruista): diventa un paziente autonomo e impegnato che condivide la propria esperienza e competenza con altri pazienti (e con i loro familiari).

2. Facebook e diabete: l'educazione tra pari nelle comunità on-line

I dati discussi all'interno del report globale sul diabete promosso dall'Organizzazione Mondale della Sanità sono chiari: le diagnosi di diabete sono in aumento, con 422 milioni di adulti nel mondo colpiti da questa malattia cronica (WHO, 2016). I dati sull'incremento delle diagnosi in Italia sono in linea con questa tendenza: negli ultimi trent'anni la diffusione del diabete è quasi raddoppiata, coinvolgendo il 5,3% della popolazione nel 2016 (Italian Diabetes e Obesity Barometer, 2018; WHO, 2016).

Al di là che si abbia a che fare con il tipo 1 o 2, il diabete è una malattia cronica che richiede una gestione complessa (e quotidiana): a titolo puramente esemplificativo, basti ricordare che i pazienti con diabete devono saper misurare la glicemia tramite il glucometro e, per quanto riguarda il diabete di tipo 1 e un sottogruppo di pazienti con diabete di tipo 2, determinare la dose di insulina necessaria sulla base del conteggio dei carboidrati del pasto che si andrà a consumare. Il riferimento a questi esempi è sufficiente per mettere in luce in che senso un buon controllo

del diabete dipende anche da comportamenti quotidiani che, per essere messi in atto in maniera appropriata, richiedono una serie di conoscenze mediche (rispetto a quali debbano essere i valori target delle glicemie o a come debba essere adeguata, di conseguenza, la terapia insulinica) e pratiche (rispetto all'utilizzo di strumenti tecnologici come microinfusori e glucometri). Tali conoscenze devono poi essere regolate allo stile di vita, che varia in relazione alle abitudini individuali dei pazienti, ma anche agli avvenimenti e agli imprevisti che caratterizzano le fasi della vita di ogni persona.

La capacità di gestire in maniera consapevole e autonoma il percorso di cura (*self-management*) deve essere pensata in riferimento a questa complessità. Una complessità che richiede un coinvolgimento attivo e un'educazione continua del paziente. In questo contesto, il ruolo dei social media in genere, e di Facebook per quel che riguarda i temi discussi in questo contributo, deve dunque essere considerato a partire dal ruolo che coinvolgimento attivo e autogestione del paziente rivestono nel miglioramento degli esiti clinici e della qualità di vita del paziente cronico (Islam *et al*., 2019; Merolli, Gray e Martin-Sanchez, 2013).

La nozione di coinvolgimento attivo del paziente è di recente introduzione e di crescente interesse in medicina (Menichetti, Libreri, Lozza e Graffigna, 2016). Coinvolgere il paziente nelle cure è un imperativo etico e una necessità pratica per i sistemi sanitari occidentali (Weil, 2016). Diversi contributi scientifici ribadiscono che la partecipazione attiva del paziente sia un passo necessario per raggiungere una gestione efficace e sostenibile dei servizi sanitari (Bodenheimer, Lorig, Holman e Grumbach, 2002; Hibbard e Greene, 2013; James, 2013). Nella letteratura scientifica l'espressione "coinvolgimento attivo" può fare riferimento:

– alle conoscenze, competenze e risorse per gestire una malattia (*patient activation*) (Hibbard, Stockard, Mahoney e Tusler, 2004);

– alla percezione di acquisire un maggiore controllo sulle decisioni e sulle azioni che riguardano la gestione della propria malattia (*patient empowerment*) (Anderson e Funnell, 2010);

– all'avere un ruolo attivo nelle decisioni cliniche (*patient involvement*) (Thompson, 2007);

– all'intraprendere efficacemente un percorso emotivo, informativo e di buone pratiche comportamentali che gradualmente, passo dopo passo, porta la persona ad acquisire non solo un ruolo attivo nel percorso di cura, ma anche a mettere in campo azioni propositive per migliorare l'esperienza di cura (sua e/o di altre persone) (*patient engagement*) (Graffigna, Barello, Bonanomi e Lozza, 2015).

In questo contributo, adottiamo quest'ultima definizione prendendo in considerazione una visione ampia e complessa di coinvolgimento attivo (*patient engagement*). A partire da un'analisi qualitativa dei bisogni delle persone con diabete e diverse altre malattie croniche in Italia, il modello teorico sviluppato da Graffigna, Barello, Bonanomi e

Lozza (2015) concettualizza la nozione di *patient engagement* come un percorso emotivo, informativo e comportamentale, composto da quattro fasi esperienziali principali: *blackout*, allerta, adesione, ed *engagement* vero e proprio. Segue una breve descrizione di tali fasi, utile per comprendere la prospettiva teorica e i risultati del nostro studio.

Nella fase di *blackout* la persona attraversa un momento di disorientamento informativo, sconforto emotivo e disordine nelle abitudini legate alla gestione della malattia. Generalmente questa fase connota il periodo appena successivo alla diagnosi, o momenti di cambiamento del piano terapeutico o di peggioramento della malattia. Diversa è invece la fase di allerta, dove prevale un senso di allarme e preoccupazione più che di sconforto, e dove la persona intraprende tentativi di raccolta di informazioni e inizia a introdurre nuove pratiche comportamentali di gestione della malattia nelle proprie abitudini quotidiane. Nella fase di adesione la persona acquisisce una visione di sé come paziente, ha già un buon bagaglio informativo e risulta aderente rispetto alle prescrizioni di cura. Questa fase descrive una relativa stabilità e padronanza da parte della persona rispetto al percorso di cura, senza però che riesca a sentirsi risorsa attiva e parte integrante del team di cura, o portatrice di esperienze di malattia capaci di integrare e arricchire la conoscenza medica e di supportare altre persone, che è quello che invece succede nella fase di *engagement*. Infatti, le persone in fase di *engagement* sono finalmente riuscite a dare senso alla malattia e a vederne i risvolti positivi per la propria esistenza in senso lato, diventando addirittura ambasciatori di buone pratiche di cura e salute tra le persone che hanno attorno.

È possibile immaginare che a ciascuna fase corrispondano bisogni diversi, in termini di supporto emotivo e informativo. Un qualsivoglia percorso di educazione, per essere efficace, dovrebbe tener conto delle differenti fasi di *engagement* che abbiamo descritto.

Adattando questo discorso generale al contesto specifico della cura del diabete, si può dunque immaginare che alla ricezione passiva e disordinata di informazioni che potrebbe caratterizzare le fasi di *blackout* o di allerta, possano fare seguito – nelle fasi di adesione ed *engagement* – atteggiamenti proattivi che portano a ricercare (e a interpretare) in autonomia nuove informazioni e nuove evidenze sul controllo del proprio diabete. Questa capacità di interpretare e leggere in maniera critica evidenze e informazioni relative al diabete sembra dunque fondamentale per descrivere una buona gestione della malattia e, insieme a quest'ultima, anche il desiderio di farsi ambasciatore di buone pratiche nella propria comunità (in presenza oppure on-line).

In questo contesto, l'interazione tra pazienti è stata interpretata come un nuovo canale per promuovere la discussione e la condivisione di informazioni sulla gestione della cura (O'Keeffe e Montori, 2016). Al di là che si stia facendo riferimento alle comunità in presenza (come associazioni o gruppi di pazienti) o alle comunità on-line (per esempio,

a gruppi di pazienti su Facebook), l'interazione tra pazienti può essere definita come un tipo di supporto tra pari (*peer support*). Dennis (2003, 329) ne ha dato la seguente definizione:

> peer support, within the health care context, is the provision of emotional, appraisal, and informational assistance by a created social network member who possesses experiential knowledge of a specific behaviour or stressor and similar characteristics as the target population, to address a health-related issue of a potentially or actually stressed focal person.

Il riferimento all'educazione tra pari è importante perché esiste tutta una letteratura che mostra l'impatto di questa tipologia di supporto sul miglioramento della capacità di gestione autonoma del diabete (Joseph, Griffin, Hall e Sullivan, 2001; Piette, Resnicow, Choi e Heisler, 2013; Tang *et al*., 2014; Tang, Afshar, Elliott, Kong e Gill, 2019). Più nello specifico, il supporto da parte di chi condivide la stessa esperienza di malattia può essere utile sul piano della condivisione del vissuto di malattia (piano emotivo), delle informazioni (piano cognitivo) e dei consigli pratici efficaci – perché già usati da altri – per migliorare le pratiche quotidiane di gestione della malattia (piano comportamentale), e può diventare dunque uno strumento per favorire il coinvolgimento attivo del paziente. In questo contesto, i forum e i gruppi di discussione on-line potrebbero essere molto utili come strumenti di supporto tra pari (Kingod, 2018; Merolli *et al*., 2013).

Dalla letteratura scientifica, sappiamo che l'uso di comunità online come Facebook potrebbe favorire il coinvolgimento attivo dei pazienti, e che l'uso dei social media potrebbe avere un grande potenziale in termini di capacità di creare cambiamento rispetto a come pazienti e strutture sanitarie si relazionano vicendevolmente (Rozenblum e Bates, 2013; Smailhodzic *et al*., 2016). Un'analisi della letteratura sull'uso dei social media per favorire il coinvolgimento attivo e l'educazione dei pazienti offre incoraggianti indicazioni in tal senso, ma sottolinea la sostanziale mancanza di studi a riguardo (Househ *et al*., 2014). Nel caso dei pazienti con diabete, uno studio sui partecipanti alle comunità on-line basate sul supporto tra pari offre qualche dato positivo preliminare rispetto all'utilità che si ha nella condivisione di informazioni ed esperienze con altri pazienti diabetici (Gavrila, Garrity, Hirschfeld, Edwards e Lee, 2019). Altri studi descrivono alcuni dei temi predominanti discussi sui social media (Tenderich, Tenderich, Barton e Richards, 2018) o su Facebook all'interno di gruppi dedicati al diabete (Arsand, Bradway e Gabarron, 2019; Yan Zhang, He e Sang, 2013). Confermando altre ricerche sull'affidabilità delle informazioni condivise in gruppi di discussione on-line (Cole, Watkins e Kleine, 2016), White e colleghi (2018) hanno somministrato un questionario e hanno chiesto ai partecipanti di valutare i consigli ricevuti all'interno di alcuni gruppi su Facebook: il 99%

dichiara di non aver avuto alcun problema causato dai consigli ricevuti online, e il 40% di aver trovato quei consigli anche utili ed efficaci.

Per quanto riguarda il contesto italiano, siamo a conoscenza di una sola ricerca che ha affrontato questi temi: Troncone, Cascella e Chianese (2015; 2016) hanno condotto un'analisi testuale computerizzata preliminare per analizzare i nuclei tematici di discussione dei post pubblicati all'interno del gruppo Facebook "Mamma e diabete", un gruppo di supporto per mamme di bambini con diabete di tipo 1.

In attesa di prove più solide, risulta chiaro che il ruolo dei social media per la salute, e di Facebook per quel che riguarda i temi discussi in questo contributo, deve essere considerato a partire dal ruolo che *engagement* del paziente e capacità di gestione autonoma della malattia rivestono nel miglioramento della salute (Islam *et al*., 2019; Merolli *et al*., 2013).

3. Metodologia della ricerca

Lo studio che proponiamo è stato disegnato per capire in che modo:

– la comunicazione su Facebook può contribuire alla diffusione di conoscenze sul funzionamento e la gestione del diabete;

– la condivisione di conoscenze (ed esperienze) su Facebook può promuovere la partecipazione attiva dei pazienti nella gestione del proprio percorso di cura e di salute.

3.1. *Analisi dei bisogni di persone con diabete che usano Facebook*

In una prima fase abbiamo somministrato un questionario strutturato auto-compilato e in forma anonima, volto a comprendere i bisogni informativi delle persone con diabete che fanno uso di Facebook. Oltre a identificare le frequenze d'uso e i principali motivi che spingono queste persone a utilizzare internet e Facebook per cercare e/o condividere informazioni sul diabete, abbiamo misurato il livello di coinvolgimento attivo nella gestione delle cure (*patient engagement*) e il livello di alfabetizzazione sanitaria (*health literacy*).

Il questionario era rivolto a persone adulte, reclutate on-line, con diagnosi di diabete di tipo 1 o 2, e che utilizzano Facebook. Abbiamo articolato il questionario in quattro sezioni principali:

(a) informazioni sociodemografiche e cliniche;

(b) scala di valutazione del livello di coinvolgimento attivo del paziente: Patient Health Engagement Scale (Graffigna *et al*., 2015), e scala di valutazione del livello di alfabetizzazione sanitaria: Brief Health Literacy Screeners (Lorini *et al*., 2017), entrambe validate;

(c) set di item pensati ad hoc e finalizzati a valutare i bisogni informativi legati al diabete su una scala Likert da 1 a 5. Gli item riguardavano: alimentazione, attività fisica, farmaci e terapie, complicanze, ambulatori e servizi;

(d) set di item pensati ad hoc e finalizzati a comprendere la frequenza, l'utilità percepita e le funzioni d'uso di internet e di Facebook.

I dati raccolti sono stati analizzati mediante il software statistico SPSS versione 4.0. Abbiamo condotto analisi descrittive (media, range, deviazione standard) e presentato i dati di tipo categoriale tramite frequenze e statistiche percentuali. Infine, abbiamo effettuato delle correlazioni di Pearson per studiare la relazione tra il livello di *patient engagement*, *health literacy* e i bisogni informativi, e tra queste variabili e l'utilizzo di internet e Facebook per la gestione del diabete.

3.2. *Analisi delle interazioni comunicative del gruppo Facebook "Tutti i Diabetici Uniti in Rete" (TIDUIR)*

Nella seconda fase dello studio abbiamo analizzato un campione di post pubblicati all'interno del gruppo chiuso di Facebook TIDUIR, uno dei gruppi segnalati più frequentemente dai partecipanti al questionario somministrato nella prima fase. Lo scopo dichiarato di TIDUIR è di unire le persone con diabete (e i loro familiari, parenti, amici, ...) in una comunità on-line che ammette la condivisione di esperienze e la pubblicazione di notizie "avvalorate da studi scientifici e dalle linee-guida ufficiali"[1].

Il campione analizzato include tutti i post pubblicati nell'arco del mese di gennaio 2019, salvati manualmente insieme ad altre informazioni relative al numero di "mi piace" e di commenti. Per ciascun post abbiamo annotato se questo includeva la condivisione di testo, immagini, video, link o una loro combinazione. Abbiamo utilizzato queste informazioni per produrre una prima descrizione del campione, che abbiamo considerato essere significativo rispetto al tipo di interazioni comunicative condotte normalmente all'interno di TIDUIR. Abbiamo poi approfondito l'analisi dei post da un punto di vista tematico, attribuendo a ciascun post una funzione principale e distinguendo, più nello specifico, tra post che hanno l'obiettivo principale di:

1. costruire un rapporto con la comunità;
2. chiedere o condividere informazioni;
3. motivare gli altri utenti della comunità.

Abbiamo incluso in una categoria a parte tutti quei post riguardanti la gestione del gruppo e considerati quindi come non rilevanti per descri-

[1] La descrizione delle finalità del gruppo TIDUIR è disponibile al link seguente: http://www.facebook.com/groups/253803764055/about/ (consultato il 19 febbraio 2019).

vere le dinamiche di supporto tra pari di TIDUIR: post di gestione del gruppo non legati al diabete, post di condivisione della nuova immagine del gruppo, post in cui si propone lo scambio o la vendita di farmaci e/o di materiale per il controllo della glicemia. Per ciascuna funzione abbiamo poi riconosciuto alcune sottocategorie:

1. costruire un rapporto con la comunità, tramite post (a) di saluto, (b) relativi alla tecnologia, (c) relativi all'ambito clinico, (d) che descrivono che cosa significa vivere con il diabete, (e) riguardanti aspetti di gestione del gruppo;
2. chiedere o condividere informazioni, tramite post su (a) stile di vita (di cui attività fisica, dieta, ricette), (b) gestione clinica del diabete e delle sue complicanze, (c) gestione delle procedure di accesso ad ambulatori e servizi sanitari, (d) gestione delle tecnologie per il diabete, (e) notizie sul diabete, (f) significato emotivo e vita con il diabete;
3. motivare gli altri utenti della comunità, tramite post su (a) stile di vita.

Abbiamo usato questo schema di classificazione per discutere alcuni casi che consideriamo esemplificativi e che descrivono in che modo il supporto tra pari, nelle sue declinazioni sul piano emotivo e informativo, si realizza nelle interazioni comunicative su TIDUIR.

Abbiamo contattato uno degli amministratori del sito per discutere l'obiettivo dello studio e valutarne la fattibilità sul piano etico. In accordo anche con gli altri amministratori del gruppo abbiamo pubblicato un post per condividere l'obiettivo del nostro studio e della nostra analisi; abbiamo chiesto contestualmente agli utenti di contattarci se reputavano di non voler partecipare allo studio e preferivano che escludessimo i loro post dall'analisi. Nessun utente ha espresso parere negativo, abbiamo quindi incluso nell'analisi tutti i post pubblicati nel mese di gennaio 2019. Per garantire ulteriormente la privacy e il rispetto dei dati sensibili abbiamo anonimizzato tutti gli esempi inclusi in questo contributo.

4. Chi sono le persone con diabete che usano Facebook? Una prima fotografia

4.1. *Sezione (a): caratteristiche sociodemografiche e cliniche*

119 persone con diabete e utenti di Facebook hanno acconsentito a compilare il questionario ed espresso il loro consenso al trattamento dei dati forniti a fini di ricerca, perlopiù donne (67%), con un buon livello di istruzione (54% con diploma superiore), con un lavoro full-time (33%), e coniugate o conviventi (57%). Il 71% dei partecipanti ha riferito di avere una diagnosi di diabete di tipo 1 e il 69% di seguire una terapia insulinica. Per un dettaglio maggiore circa le caratteristiche dei partecipanti coinvolti rimandiamo alla tabella 1.

TAB. 1. *Caratteristiche dei partecipanti (n = 119)*

Età: anni (media, deviazione standard)	18-77 (44, 9)
Genere	
Maschi	39 (33%)
Femmine	80 (67%)
Titolo di studio	
Licenza Elementare	4 (3%)
Licenza Scuola Media Inferiore	18 (15%)
Diploma Scuola Media Superiore	64 (54%)
Laurea	22 (18.5%)
Master/Dottorato	8 (7%)
Altro	3 (2.5%)
Stato civile	
Celibe/nubile	35 (29%)
Sposato/convivente	68 (57%)
Separato/divorziato	7 (6%)
Vedovo/a	3 (2.5%)
N/A	5 (4%)
Stato lavorativo	
Lavoratore full-time	39 (33%)
Lavoratore part-time	12 (10%)
Disoccupato	20 (17%)
Pensionato	13 (11%)
Altro	33 (28%)
Area geografica di residenza	
Nord Italia	46 (39%)
Centro Italia	27 (23%)
Sud Italia e Isole	40 (34%)
Estero	2 (2%)
N/A	4 (3%)
Tipo di diabete	
Diabete di tipo 1	84 (71%)
Diabete di tipo 2	25 (21%)
N/A	10 (8%)
Trattamento con terapia insulinica	82 (69%)
Presenza di comorbidità	47 (39.5%)
Fumatore	30 (24%)
Indice di massa corporea	Media 26.7 kg/m2 (deviazione standard = 22.5)

4.2. *Sezione (b): engagement, health literacy e principali bisogni informativi*

Da un punto di vista descrittivo, un'analisi delle frequenze dei livelli di *engagement* mostra che il 26% dei partecipanti ha importanti bisogni di carattere psicosociale rispetto al diabete e bassi livelli di *engagement*: il 4.2% riferisce di sentirsi in una fase di *blackout* e il 21.8% in una fase di allerta. I partecipanti che percepiscono di sentirsi meno bisognosi e più attivamente coinvolti rispetto al diabete e al percorso di cura sono numerosi (56.3%): il 45.4% riferisce di sentirsi in una fase di adesione

e si attribuisce buone capacità di gestione delle cure, il 10.9% risulta in una fase di *engagement* e si riconosce ottime risorse emotive, cognitive, comportamentali per la gestione del diabete. Questi risultati, con una frequenza maggiore di pazienti in fase di adesione, risultano in linea con quanto emerso in studi precedenti su altre popolazioni di pazienti (Graffigna *et al*., 2015; Yaying Zhang *et al*., 2017).

I risultati della scala di alfabetizzazione sanitaria permettono di attribuire ai partecipanti allo studio buone capacità di comprensione e gestione delle informazioni relative al diabete. Nel dettaglio, su una scala da 1 a 5, i partecipanti riportano un livello di alfabetizzazione sanitaria medio pari a 3.7 (DS = 0.8) e, in particolare, buoni valori per quanto riguarda la capacità di comprendere le informazioni scritte (media = 3.9, DS = 0.9).

Le analisi dei bisogni informativi dei pazienti mostrano un quadro simile a quello già descritto in riferimento ai livelli di *engagement* e *health literacy*. Su una scala Likert da 1 (per nulla) a 5 (del tutto), la maggior parte dei partecipanti dichiara di avere "abbastanza" informazioni utili nella gestione del diabete rispetto a: farmaci e terapie (52.1%), alimentazione (51.5%), complicanze (45.3%), ambulatori e servizi (41.8%), attività fisica (38.7%).

I bisogni informativi maggiori riguardano l'area degli ambulatori e dei servizi, e quella delle complicanze relative al diabete: rispettivamente, il 24.2% e il 21.1% dei partecipanti dichiara di avere nessuna o soltanto poche informazioni relativamente a questi ambiti.

4.3. *Sezione (c): uno zoom descrittivo sull'uso dei social media*

L'84% dei partecipanti riferisce di essere un frequentatore assiduo di Facebook, con diversi accessi al giorno; il 52.2% dichiara di cercare on-line informazioni per il diabete abbastanza o molto spesso. L'87.5% percepisce internet come uno strumento abbastanza (60.2%) o molto utile (27.3%), uno spazio in cui tendenzialmente è possibile trovare le informazioni di cui si ha bisogno per gestire il diabete. In particolare, i partecipanti percepiscono Facebook utile sia per condividere sia per cercare informazioni sul diabete (42.2%), con il 28.9% che vede Facebook funzionale specialmente per condividere informazioni sul diabete. L'area relativa a farmaci e terapie per il diabete è quella che vede i pazienti più attivi sia nel cercare (71%) sia nel condividere (76%) informazioni.

Per quanto riguarda le ragioni che spingono a utilizzare i gruppi su Facebook dedicati a persone con diabete e ai loro familiari, le principali ragioni indicate dai partecipanti sono: rimanere aggiornati sul diabete e su come curarlo (26.9%), sentirsi in contatto con persone che condividono la stessa malattia (19.3%), sentirsi parte attiva di una comunità (9.2%) e, infine, esprimere idee e vissuti sul diabete (7.6%).

4.4. *Patient engagement, health literacy e bisogni informativi: quale relazione?*

Abbiamo poi condotto analisi ulteriori per comprendere la relazione tra livelli di *engagement*, alfabetizzazione sanitaria e bisogni informativi dichiarati dai partecipanti. In altri termini, eravamo interessati a capire se l'espressione di una maggiore necessità informativa fosse in qualche modo legata al livello di *engagement* e di alfabetizzazione sanitaria dei partecipanti. Abbiamo trovato che il livello di *engagement* varia in maniera sinergica e significativa ai bisogni e alle capacità informative. All'aumentare del livello di coinvolgimento attivo nel percorso di gestione del diabete, aumenta anche la percezione di possedere informazioni adeguate sul diabete, e in particolare su alimentazione, attività fisica, complicanze e servizi di cura. Inoltre, livelli di *engagement* più alti sono statisticamente associati con livelli di alfabetizzazione sanitaria generalmente più alti (correlazione di Pearson = .211; Sig = .043); sono associati, più in particolare, a maggiori competenze di comprensione delle informazioni sul diabete.

Non abbiamo trovato alcuna relazione tra livelli maggiori o minori di *engagement* e la capacità di leggere autonomamente le informazioni scritte relative al diabete o di compilare i moduli medici (tab. 2).

TAB. 2. *Relazione tra patient engagement, health literacy e bisogni informativi*

	Media	Correlazione con *Patient Engagement*	Significatività
Bisogni informativi (alimentazione)	3.84	.271	.008*
Bisogni informativi (attività fisica)	3.61	.292	.004**
Bisogni informativi (farmaci e terapie)	3.87	.189	.070
Bisogni informativi (terapie alternative)	1.83	.111	.289
Bisogni informativi (complicanze)	3.59	.260	.011*
Bisogni informativi (ambulatori e servizi)	3.33	.313	.002**
Health literacy (comprensione informazioni)	3.96	.380	.000**
Health literacy (lettura informazioni)	3.66	.038	.717
Health literacy (scrittura informazioni)[1]	2.53	–.026	.807

Analisi statistiche effettuate: Correlazioni di Pearson.
* significatività <.05.
** significatività <.01.
[1] Scoring invertito.

4.5. *Engagement, Ricerca di informazioni e uso di Facebook*

Infine, abbiamo analizzato la relazione tra l'uso di Facebook e i bisogni di diverse tipologie di utenti, in termini di *engagement*, alfabetizzazione sanitaria e bisogni informativi dichiarati. Le analisi che abbiamo effettuato mostrano che all'aumentare dei livelli di *engagement* dei pazienti, decresce in maniera significativa la frequenza con cui internet viene utilizzato per cercare informazioni sul diabete. Questo dato ci permette di sottolineare che pazienti con livelli di *engagement* più alti usano preferenzialmente Facebook per restituire le informazioni di cui dispongono agli altri.

Non emergono invece altre relazioni significative per quanto riguarda le altre variabili d'uso di internet e Facebook (ovvero la frequenza d'uso di Facebook e l'utilità percepita di internet e Facebook per la gestione del diabete). L'utilizzo dei gruppi su Facebook sembra seguire logiche differenti: non cambia in funzione dei livelli di *engagement* dei pazienti, né in termini di frequenza d'uso né di utilità percepita (tab. 3). Abbiamo ottenuto risultati simili anche analizzando le variabili relative all'alfabetizzazione sanitaria e ai bisogni informativi dei partecipanti (analisi non riportate).

TAB. 3. *Relazione tra patient engagement e utilizzo di Facebook e Internet*

	Media	Correlazione con *Patient Engagement*	Significatività
Frequenza d'uso Facebook	1.21	–.019	.858
Frequenza d'uso internet	4.49	–.204	.048*
Utilità percepita internet	3.15	.134	.214
Utilità percepita Facebook	3.14	.117	.287

Analisi statistiche effettuate: Correlazioni di Pearson.
* significatività <.05.

4.6. *Chi sono le persone con diabete che usano Facebook?*

Questa prima fase di ricerca ha permesso di evidenziare alcune informazioni importanti relative alle caratteristiche dei pazienti con diabete che usano Facebook, rispetto ai loro bisogni informativi e rispetto a come essere più o meno coinvolti nelle cure e competenti nel comprendere le informazioni sanitarie possa contribuire all'uso e alla diffusione di conoscenze su internet e Facebook. Riassumendo, emerge che:

– le persone che partecipano a gruppi on-line sul diabete hanno, nel complesso, buoni livelli di *engagement*, buone competenze di alfabetizzazione sanitaria e buone risorse informative a disposizione, e questi tre aspetti sembrano associati tra loro;

– nello specifico dei livelli di *engagement*, il quadro che risulta da questa indagine è comunque abbastanza variegato, con persone ancora poco attivamente coinvolte nel percorso di cura e persone ampiamente autonome;

– vi è una valutazione sostanzialmente positiva delle informazioni a disposizione on-line ed emerge una valutazione positiva di Facebook, soprattutto da chi dichiara di partecipare alla discussione nei gruppi sul diabete;

– i gruppi Facebook sul diabete permettono, nello specifico, di rimanere informati e aggiornati rispetto alla gestione e cura del diabete, e solo in seconda istanza di facilitare un senso di condivisione, comunità e comunanza con altre persone che condividono un percorso di malattia analogo;

– più è alto il livello di *engagement*, più raramente le persone usano internet per cercare informazioni sul diabete;

– la valutazione dei gruppi Facebook sul diabete e il loro utilizzo rimangono invariati a prescindere dalle competenze e risorse informative dei pazienti, così come dal loro livello di *engagement*.

5. Facebook come strumento di educazione e supporto tra pari: il caso di TIDUIR

224 sono stati i post pubblicati su TIDUIR nell'intero mese di gennaio del 2019. I messaggi sono stati pubblicati da 155 utenti differenti, in maggioranza donne (147; 66%). Soltanto 16 utenti hanno pubblicato più di un messaggio durante il periodo di osservazione cui fa riferimento questa analisi, con 42 (19%) post pubblicati da uno degli amministratori del gruppo. La tabella 4 include dati ulteriori relativi sulle reazioni ai post (numero totale di "mi piace" e di commenti), oltre che un conteggio complessivo sul tipo di post pubblicati e cioè sul numero di post contenenti testo, immagini, video, link o una loro combinazione. I post includono principalmente testi (201; 90%).

Tab. 4. Tipi di reazione ai post e tipi di post pubblicati

Tipi di reazione		Tipi di post pubblicati			
Mi piace	Commenti	Testo	Immagine	Video	Link
1843	2151	201	15	11	44

I risultati dell'analisi delle funzioni di supporto tra pari sulla base delle categorie presentate in § 3.2 è illustrato nella figura 1.

111 post (50%) sono utilizzati allo scopo di chiedere o condividere informazioni sul funzionamento del diabete e su come gestirlo. 96 (43%) post riguardano invece più da vicino le dinamiche di costruzione

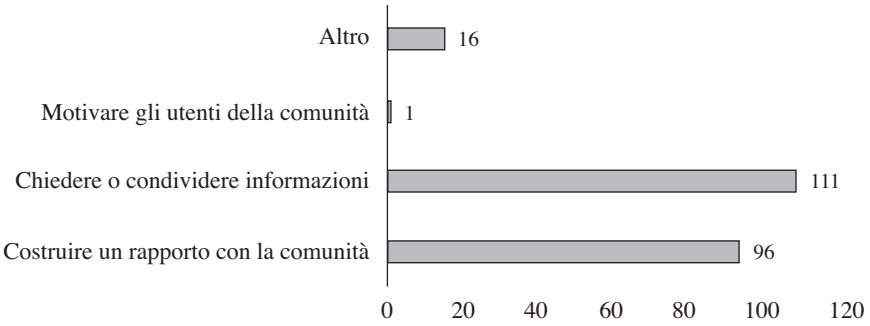

Fig. 1. Funzioni del supporto tra pari su TIDUIR.

dei rapporti della comunità di TIDUIR. Questi risultati riflettono quelli emersi dalla prima fase di ricerca, e supportano ulteriormente il fatto che i gruppi su Facebook abbiano principalmente la funzione di promuovere lo scambio di informazioni, e solo in seconda istanza di costruire relazioni e/o di condividere emozioni e vissuti. Abbiamo poi riconosciuto a un unico post la funzione di motivare altri utenti della comunità; si tratta del messaggio postato da uno degli amministratori del gruppo che ha l'esplicito obiettivo di esortare gli altri membri a condividere le loro informazioni sull'attività fisica svolta e, implicitamente, a fare più attività fisica (Post n. 112).

Post n. 112 (19 "Mi piace", 69 commenti)
ATTENZIONE ATTENZIONE Non ho visto nessun post che ci mostra la vostra attività fisica mattutina!!! Tutti pigri come la sottoscritta???? Non ci credo!!! Forza e coraggio e fateci venire voglia di sgranchirci le ossa!! Chi sarà il primo???

Nella categoria "Altro", abbiamo incluso post che riguardano:
– informazioni sulla gestione del sito non strettamente legati al diabete (7 post);
– scambio o vendita di farmaci e/o materiale per il controllo della glicemia (6 post);
– temi di altro genere e comunque non pertinenti con gli obiettivi dichiarati del gruppo (3 post).

5.1. *Costruire un rapporto con la comunità*

I post catalogati come messaggi che mirano a costruire un rapporto con la comunità riguardano prevalentemente messaggi di saluto da parte dei nuovi membri del gruppo (92%) (fig. 2).

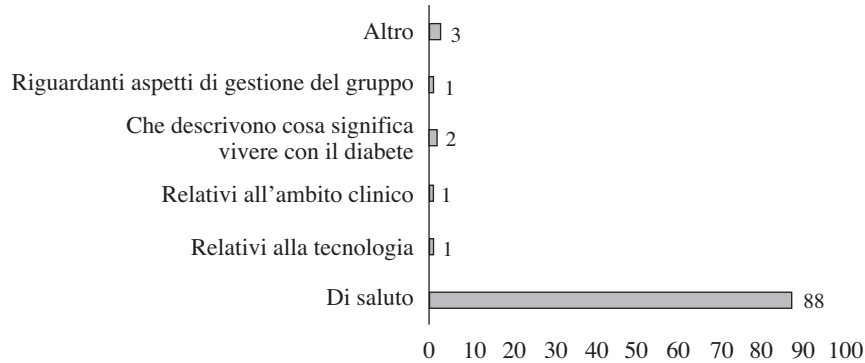

FIG. 2. Costruire un rapporto con la comunità, tramite post.

Questi post non ricevono molta attenzione in termini di "Mi piace" e di commenti, si tratta generalmente di messaggi di ringraziamento per l'inclusione del gruppo e che includono, nel 44% dei casi (42 post), una presentazione clinica – per così dire – di sé (Post n. 8) o del proprio familiare con diagnosi di diabete (Post n. 111) (fig. 3).

Post n. 8 (3 "Mi piace", 1 commento)
Grazie per l'approvazione! T1 da 10 anni
Amministratore: Benvenuto

Post n. 111 (8 "Mi piace", 7 commenti)
Grazie per avermi accolta, sono madre di una ragazza diabetica.

Nella maggioranza dei casi, questi post vengono commentati con semplici messaggi di benvenuto da uno o due altri utenti – si tratta, molto spesso, di uno degli amministratori.

Tuttavia, anche post di questo tipo, all'apparenza poco utili, talvolta aprono a interazioni comunicative più interessanti per discutere il tema del supporto emotivo tra pari. Nello specifico, ci sono casi in cui la comunità si palesa con più forza, soprattutto in termini di numero di commenti, per supportare membri che nel loro messaggio di saluto al gruppo condividono informazioni e storie, per così dire, "particolari".

Ci sono casi in cui la comunità di utenti diventa più attenta, e cioè aumenta il numero di "Mi piace" e di commenti ai messaggi di saluto da parte di un membro che è stato appena incluso nel gruppo (Post n. 17). Il post n. 17 è emblematico a questo proposito. Tra i commenti, si trovano richieste di informazioni su come sono trascorsi tutti quegli anni col diabete rispetto alle complicanze ([5]); ci sono poi utenti che approfittano della conversazione viva (e forse dell'esperienza altrui) per chiedere consigli su come risolvere problemi particolari ([8]). C'è il messaggio di benvenuto dell'amministratore ([2]), ma ci sono anche

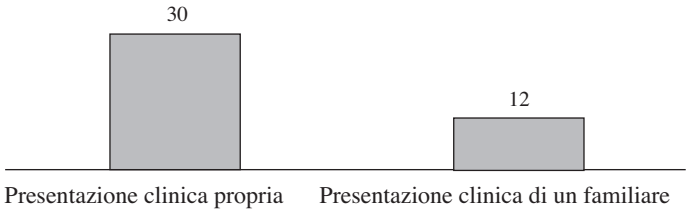

FIG. 3. Post di saluto con presentazione clinica propria o di un familiare con diagnosi di diabete.

altri utenti che fanno i complimenti per la lunga convivenza col diabete, magari aggiungendo informazioni sull'atteggiamento emotivo ([10], nota l'utilizzo dell'espressione "TENIAMO DURO" come esempio di informazione sull'atteggiamento emotivo rispetto al trascorrere una vita "insieme" al diabete).

Accanto a vari messaggi che tratteggiano una sorta di contorno emotivo della comunità ([12]), ci sono poi messaggi che contengono elementi linguistici particolari: in questo caso segnaliamo l'uso di espressioni come "infartino" ([11]) o di messaggi ironici ([13] e [14]).

Il commento [9] è un buon caso di moderazione da parte di uno degli amministratori che esorta l'Utente 2 a ripubblicare il proprio post al di fuori della discussione corrente, come nuovo post, per favorire una maggiore visibilità e dunque l'intervento di un maggior numero di utenti.

Post n. 17 (35 "Mi piace", 24 commenti)

[1] *Utente zero*: Salve sono diabetica tipo 1 da 55 anni

[2] *Amministratore:* Benvenuta 😊

[3] *Utente zero*: Amministratore Daniela Sanna grazie siete un bel gruppo...

[4] *Amministratore:* Utente zero grazie

[5] *Utente 1:* Utente zero lo so è brutto chiederlo e mi scuso in anticipo... hai avuto problemi alla vista? Ai reni? Piede diabetico? Dei sintomi insomma

[6] *Utente zero*: Utente zero no fortunatamente solo cataratte operate...

[7] *Utente zero*: Anche infarto nel 2004 con due stent... ma sto bene...

[...]

[8] *Utente 2:* Diabetico da 46 anni tipo 1, chiedevo un consiglio: circa sei mesi fa il diabetologo mi ha consigliato di cambiare il microinfusore che ho della roche con l'ultimo sempre della roche cioè INSIGHT, tutto ok per 4/5 mesi, adesso comincio ad avere problemi di glicemie alte, come se non facessi insulina, comunque scopro che si tratta dell'ago che più di un giorno è mezzo non dura, mentre prima andava bene per 3 giorni, esattamente quanto dura la cartuccia di insulina, c'è qualcuno di voi che ha notato questo problema? Perché lago Dura solo un giorno e poi si ottura l'ago..?

[9] *Amministratore:* Utente 2 potresti scrivere un post così lo possono leggere tutti ed esserti di aiuto. Qui leggono solo le persone interessate a questo post

[10] *Utente 3*: COMPLIMENTI IO QUACOSINA MENO PERO TENIAMO DURO

[11] *Utente* 0: Utente 3 si infatti a parte un infartino nel 2004... ancora sopravvivo

[12] *Utente 3*: Utente zero IO NEL 2013 CON 5 RICOVERI SEMPRE PER LO STESSO MOTIVO IN UN ANNO ADESSO FACCIAMO A GARA CHI NE A DI PIU COSI PASSIAMO IL TEMPO

[13] *Utente 0*: Mi dispiace quando sento di bimbi col diabete..mi fanno tanta tenerezza...

[14] *Utente 3:* Utente zero O SI TI LEVANO IL CUORE

[...]

[15] *Utente 4*: Ciao io da 44 anni😃 e ho la gastroparesi allo stomaco... ho avuto due iniezioni in un occhio !!!!😃

[16] *Utente 5:* cosa è una gara ahahaha

5.2. *Chiedere o condividere informazioni*

La categoria più interessante, e più frequente, per analizzare il supporto fornito tra i partecipanti al gruppo raccoglie post pubblicati con la funzione principale di chiedere o condividere informazioni.

Il post n. 104 ci permette di commentare un caso interessante di un messaggio pubblicato da un nuovo utente del gruppo. Il nuovo utente ha appena scoperto di essere diabetico e utilizza il primo post per tentare di comprendere alcune indicazioni fornite dall'infermiera relative alla necessità di misurare la glicemia per qualche giorno, così da fornire un "quadro più completo" al medico durante la prima visita specialistica. Confida, dunque, nell'esperienza di pazienti più esperti. Il supporto arriva puntuale, da tre utenti ([2], [4], [5], [7]); in un caso, oltre a fornire informazioni testuali, viene pubblicata anche l'immagine di una pagina di un diario usato per segnare la misurazione delle glicemie seguendo uno schema a coppia (prima e dopo i pasti) ([5]). L'inizio di questa interazione è di estremo interesse, se si pensa che il tema della misurazione delle glicemie è uno dei più discussi all'interno della consultazione diabetologica oltre che uno dei comportamenti di gestione autonoma più importanti nel contesto della cura del diabete (Bigi, 2014). A questo riguardo, uno studio recente su un corpus di visite diabetologiche italiane ha individuato parecchi casi di incomprensioni in contesti comunicativi nei quali l'appropriatezza della misurazione delle glicemie era il punto di discussione problematico principale (Rossi e Macagno, 2019; Macagno e Rossi, 2019). Alla luce di questi risultati, il supporto degli utenti di TIDUIR diventa ancora più interessante e facilita il percorso di educazione terapeutica che l'utente inizierà ad intraprendere ufficialmente all'interno della consultazione clinica con lo specialista. E la discussione non si ferma qui.

L'Utente 3 dichiara di essere anche lei "insulino-resistente" ([8]). Una dichiarazione che è interpretata come una testimonianza ([9]) e che da avvio a una conversazione ulteriore nella quale l'Utente zero

esprime l'esigenza di voler sapere di più e al più presto che cosa significhi essere insulino-resistente nella vita di ogni giorno ([9]). Si crea un contesto nel quale l'Utente zero si sente legittimato a porre domande: all'Utente 3 stesso col quale ha appena scoperto di avere qualcosa in comune ([10]), ma anche alla comunità di membri più ampia, per avere un parere preliminare su come sono i valori delle glicemie che ha appena iniziato a misurare ([23]). Hanno luogo interazioni stimolanti dal punto di vista dell'educazione tra pari: le risposte fornite dall'Utente 3 generano domande e curiosità in altri utenti, per esempio sull'utilità del microinfusore ([16]); osservazioni interessanti emergono anche sulla relazione tra andamento della glicemia e regime alimentare, altro cardine della gestione autonoma (e quotidiana) del diabete ([26]).

Post n. 104 (6 "Mi piace", 28 commenti)

[1] *Utente zero*: Buongiorno a tutti. Ho appena scoperto di essere diabetica e insulino-resistente. Tra qualche giorno ho un appuntamento con uno specialista, nel frattempo però l'infermiera che mi ha fatto i test mi ha consigliato di misurare la glicemia per qualche giorno, così da poter fornire al medico un quadro più completo.

Solo che non mi ha specificato quando. Secondo voi quando devo misurarla? Grazie a tutti.

[2] *Utente 1*: al risveglio a digiuno, se poi vuoi dare informazione completa prima di mangiare e due ore dopo

[3] *Utente zero*: Utente 1 grazie

[4] *Utente 2*: E giusto come ti ha detto Utente 1

[5] *Utente 1*: questo foglio è una pagina del librettino che mi danno al centro se riesci a tagliare e stampare hai una guida [segue immagine di una pagina di un diario utilizzato per segnare la misurazione delle glicemie]

[6] *Utente zero*: Utente 1 grazie!

[7] *Utente 3*: Prima di ogni pasto e due ore dopo ogni pasto

[8] *Utente 3*: Anche io sono insulino resistente..

[9] *Utente zero*: Utente 3 grazie della testimonianza. Io sono totalmente ignorante in maniera, ho cercato in rete quindi ho capito la parte teorica, solo che non ho chiaro di cosa implichi nella vita di tutti i giorni, essere diabetica e insulina resistente. Ecco perché non vedo l'ora di andare dal Diabetologo, perché voglio capire meglio come gestire il tutto.

[10] *Utente zero*: Utente 3 approfitto per farti una domanda. Di solito a metà mattina avverto un malessere, come se avessi un calo di zuccheri, secondo te può dipendere da questo? Potrebbe aver senso misurare la glicemia quando mi succede? Grazie

[11] *Utente 3*: Utente 0 l'unico problema che implica essere resistenti all'insulina è che si è molto più propensi agli sbalzi della glicemia, possiamo dire che non sai quanto il tuo assume di insulina.. Io faccio 16 unità ma non so quanto il mio corpo realmente... ne assume. Per quanto riguarda il tuo malessere.. A me capita spesso.. Molte volte dopo pranzo mi stendo sul divano.. E comincio ad avere tutti j sintomi di una glicemia bassa.. Ma poi la misuro e non ho niente.. Non penso che possa dipendere da quello però

[12] *Utente zero*: Utente 3 capito, grazie

[13] *Utente 3*: Io faccio 16 unità per un piatto di pasta.. Niente pane.. E ne altro.. Faccio tutti i giorni questo tanto.. Adesso sono un paio di giorni che vado bassa.. Pur non avendo cambiato nulla nell'alimentazione

[14] *Utente 3*: Utente 0 di nulla figurati

[15] *Utente 3*: Faccio 16 unità a pasto.. E siccome ho il microinfusore la mia basale è di 0.95 unità a ora

[16] *Utente 4*: Utente 3 come ti trovi con il microinfusore?

[17] *Utente 3*: Io sinceramente mi trovo benissimo 😁

[18] *Utente 4*: Utente 3 si anche io anche xché ho un diabete molto ballerino 😡

[19] *Utente zero*: Grazie mille a tutti! 🖤

[20] *Utente 5*: Ti consiglio di farti un diario alimentare dove puoi annotare cosa mangi e le misurazioni prima di mangiare e due ore

[21] *Utente 6*: Ciao ti consiglio di misurarla al risveglio, prima dei pasti principali e due ore dopo e anche prima di andare a dormire. Sono tante misurazioni ma in questo modo il quadro è completo. Vedrai che ti daranno da fare la conta dei carboidrati e con la terapia adatta tutto ritorna sotto controllo!

[22] *Utente zero*: Rieccomi 😉 visto che siete stati così carini approfitto per un'altra domanda 😉

Ho misurato la glicemia prima di cena ed era 103 e 2 ore dopo cena era 175. Com'è?

[23] *Utente 7*: 103 perfetto

175 due ore dopo non è malvagio ma se fosse stato meno era meglio potrebbe anche dipendere da cosa hai mangiato... in diabetologia la soglia verso l'alto è stimata a da 180 in su

[24] *Utente 7*: Fossi in te la riproverei 1 ora dopo il 175 💪

[25] *Utente zero*: Utente 7 grazie! Stamattina a digiuno invece 125

[26] *Utente zero*: Utente 7 avevo mangiato 7 ravioli (li ho contati 😅) piccoli ricotta e spinaci conditi con burro e parmigiano, circa 50 grammi di salmone affumicato e 2 cucchiai di purè di patate

Dei 111 messaggi che abbiamo individuato in questa categoria, il 44% riguarda la richiesta di informazioni (fig. 4).

La figura 5 illustra più in dettaglio i temi dei post pubblicati allo scopo di chiedere o condividere informazioni su TIDUIR. La categoria

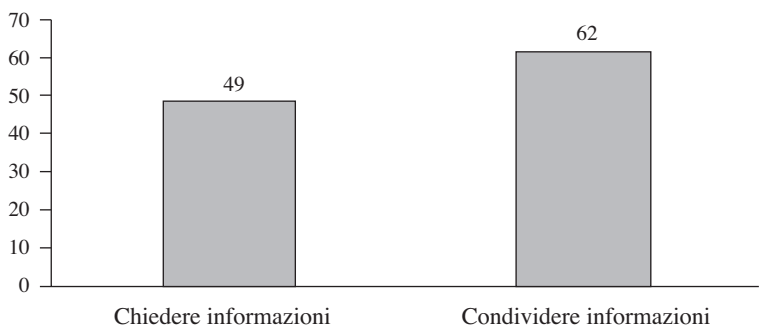

FIG. 4. Chiedere o condividere informazioni.

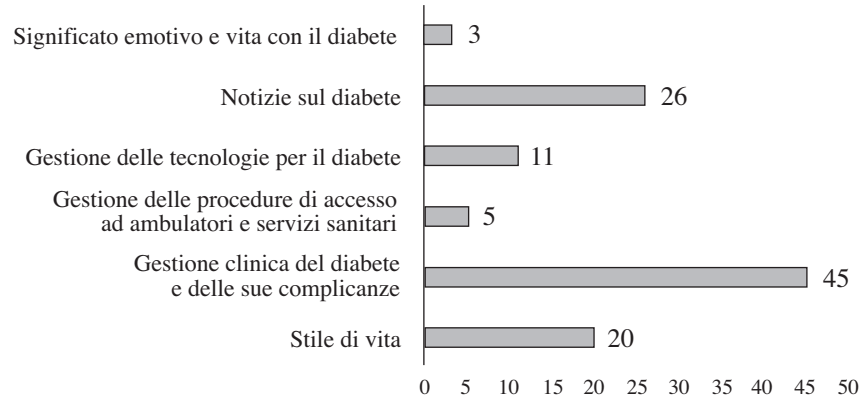

FIG. 5. Temi sui quali si chiedono o condividono informazioni tramite post.

più rappresentata raccoglie post di richiesta o di condivisione di informazioni rispetto alla gestione clinica del diabete e delle sue complicanze (45; 41%); tra questi post, 28 sono includono richieste di informazioni. Questo dato è coerente con quanto emerso nella rilevazione quantitativa, dove l'area delle complicanze relative al diabete era emersa come una delle principali fonti di bisogno informativo.

26 sono i post nei quali vengono condivise notizie sul diabete (23%), 9 sono quelli pubblicati da uno degli amministratori. Infine, 20 messaggi (18%) trattano temi relativi agli stili di vita, importanti per mantenere una buona gestione del diabete; abbiamo catalogato 3 messaggi relativi all'esercizio fisico, 4 relativi alla dieta e 13 relativi alla condivisione di ricette (e pubblicati prevalentemente da uno degli amministratori a cui sono attribuibili 12 post).

L'esperienza degli utenti più esperti può essere utile in contesti molto diversi:

– per l'interpretazione di evidenze, per esempio rispetto alla misurazione della glicemia (Post n. 176);

– per la gestione di emergenze o di errori nella somministrazione della terapia (Post n. 135);

– per problemi legati all'uso e al malfunzionamento di strumenti tecnologici (Post n. 151);

– per dubbi rispetto all'assunzione di medicinali o cibi (Post n. 165).

Post n. 176 (19 "Mi piace, 25 commenti)

Questi i miei valori... [segue immagine con valore della glicemia a 472]

Post n. 135 (8 "Mi piace, 44 commenti)

C'è qualcuno sveglio? avrei un problema urgente. Alla mia bimba per sbaglio abbiamo fatto la basale io e il suo papà. Succede qualcosa? sono in panico

Post n. 151 (5 "Mi piace, 21 commenti)

Buonasera a tutti. Primo giorno di sensore ottimo direi. il valore interstiziale (146) e quello capillare (123) a 10 minuti di distanza sono differenti quindi si tratta di una bella differenza. Per qualcuno potrebbe trattarsi di una unità di insulina in più. A tal proposito come vi comportate qualora non aveste il glucometro a portata di mano? Si cerca di fare una media oppure ci si affida prettamente al valore del sensore? (ovvio che non c'è alternativa). Altra cosa, lavoro in un posto contaminato da farine e zuccheri e sapete che sono prodotti volatili in quanto allo stato polveroso. Mi conviene coprirlo ulteriormente con una fascia oppure basta il solito cerotto? Dove acquistare cerotti e/o fasce coprenti a buon mercato? Scusate la lungaggine del post ma ho bisogno di sapere per prevenire qualsiasi intoppo e grazie alla vostra esperienza potrò gestire i dubbi amletici 😂😂😂 grazie mille a tutti

Post n. 165 (5 "Mi piace, 31 commenti)

Ma il collutorio può far alzare la glicemia?

Nei commenti a questi post abbiamo osservato tutta l'esperienza e la competenza di utenti che si sono già trovati a gestire problemi analoghi e che pertanto possono mettere a disposizione dell'utente specifico che ha richiesto supporto e, più in generale, di tutta la comunità, le loro conoscenze e le loro soluzioni pratiche di cui hanno già potuto sperimentare l'efficacia. Sono pazienti che, tramite la loro esperienza, hanno la possibilità di divenire ambasciatori di buone pratiche di gestione del diabete entro la comunità. Possono dunque avere l'opportunità di sperimentare, consolidare e rafforzare il loro ruolo nel percorso di cura; un ruolo che va al servizio della comunità più ampia di pazienti, alimentando dinamiche positive di coinvolgimento attivo.

Gli utenti più esperti sembrano allora supportare efficacemente utenti meno esperti o momentaneamente in difficoltà: in alcuni casi gli utenti esperti supportano emotivamente i familiari di un bimbo/a con diabete di tipo 1; in altri casi, proteggono la qualità informativa di temi più generali ma ugualmente determinanti per il benessere e la crescita della comunità segnalando la pubblicazione di notizie false e chiedendo che vengano rimosse dal gruppo (Post n. 212).

Post n. 212 (2 "Mi piace", 16 commenti)

[1] *Utente zero:* Questa notizia che alcuni ricercatori abbiano trovato il modo di far rigenerare le cellule del pancreas è vera o no perché nel primo caso saremmo davanti ad una scoperta eccezionale.

[2] *Utente 1:* Non è vera... assolutamente non vera... Non c'è nessun articolo scientifico con il nome di quelle persone, sui siti in inglese le notizia e del 2015, il professore Norman Hook non esiste.

[3] *Utente 1:* Sui vari gruppi si è chiesto a medici importanti ed io personalmente mi sono occupata della veridicità dell'articolo tramite ricercatori all'estero.

Dall'analisi dei post emerge bene come il ruolo di supporto della comunità on-line verso un percorso di *engagement* nella gestione del diabete non si basi e si riduca al solo "fare esperienza comune di una malattia cronica". L'esperienza su cui riflettere riguarda anche la condivisione di un obiettivo di vita più o meno esplicitamente dichiarato: come il vivere il "dolce destino" (commento al Post n. 138), e cioè vivere bene una vita che ci si è trovati a vivere, e che si deve imparare a vivere, insieme al diabete. In tal senso, la comunità on-line diviene il luogo in cui far emergere e condividere un piano più esistenziale che la diagnosi e gestione del diabete va a toccare, offrendo occasioni più allargate di *engagement* verso il percorso di malattia, come percorso di graduale inclusione della malattia nei propri piani di vita e di revisione funzionale dei propri obiettivi in funzione della malattia. In tal senso, lo scambio offerto dalla comunità on-line sembra fornire potenzialità di sostegno allineate con una visione sistemica e allargata di *engagement* (Graffigna *et al*., 2014).

6. Limitazioni

In linea con le ricerche precedenti che abbiamo discusso in questo contributo (§ 2), abbiamo presentato i risultati di uno studio disegnato per mettere in luce le dinamiche di uso e di ricerca di informazioni on-line da parte di persone con diabete, molte delle quali dichiarano di beneficiare delle interazioni con altri pazienti all'interno di comunità on-line su Facebook come TIDUIR. In particolare, abbiamo osservato tramite due angolazioni diverse il ruolo che le comunità online su Facebook svolgono per favorire l'educazione e il coinvolgimento attivo delle persone con diabete nella gestione della malattia. Si tratta di un lavoro preliminare e presenta alcune limitazioni che discutiamo brevemente in questo paragrafo.

Una limitazione riguarda il campione di partecipanti alla rilevazione quantitativa e più in particolare il modo in cui la cosiddetta distorsione di auto-selezione (*self-selection bias*) potrebbe aver influenzato i risultati che abbiamo ottenuto (§ 4). I partecipanti che hanno accettato di aderire allo studio sono stati reclutati su base volontaria e tramite un campionamento di convenienza (sono stati utilizzati canali di divulgazione on-line e sono stati contattati i moderatori di gruppi Facebook o altre pagine online per il diabete per diffondere il link al questionario). Potrebbero dunque rappresentare soprattutto quella parte della popolazione già attivamente coinvolta nella gestione della malattia, nonché nella ricerca di informazioni on-line e nella partecipazione ai gruppi di discussione. Allo stesso tempo, i risultati mostrano che i partecipanti alla presente ricerca hanno caratteristiche nel complesso analoghe – specialmente in termini di coinvolgimento attivo – a quelli di altri studi condotti su campioni

più ampi e rappresentativi (Graffigna *et al*., 2015; Yaying Zhang *et al*., 2017), il che implica che i risultati possono fornire indicazioni utili per la popolazione oggetto dello studio, per quanto vadano letti con cautela e necessitano di essere consolidati in ulteriori studi.

Inoltre, l'obiettivo del presente lavoro è puramente esplorativo e descrittivo, e non offre indicazioni circa l'impatto dell'uso di internet e dei social media sul coinvolgimento attivo nella gestione della malattia da parte delle persone con diabete. Non sappiamo, in altri termini, se e in che modo l'uso dei social media e di Facebook possa effettivamente influire su una gestione migliore della malattia. I risultati dell'analisi qualitativa ottenuti con questo studio forniscono informazioni promettenti che riconoscono una certa potenzialità ai gruppi su Facebook nel favorire l'informazione tra pari e l'*engagement* dei pazienti. Questi risultati, per quanto inseriti in uno studio esplorativo e per quanto possano spiegarsi facilmente in relazione al fatto che i partecipanti condividono obiettivi comuni e dinamiche di supporto ovvie (gli utenti meno esperti sono supportati dall'attività dei moderatori e degli utenti più competenti), non sono da sottovalutare per l'importanza che possono avere nel migliorare la comprensione delle informazioni fornite ai pazienti nelle visite diabetologiche, perché aiutano a calare nelle pratiche di vita quotidiana raccomandazioni cliniche generali la cui applicazione è tutt'altro che scontata. Inoltre, i temi discussi nei post non sono periferici rispetto alla gestione della malattia. Studi che hanno analizzato la comunicazione tra operatori sanitari e pazienti diabetici di tipo 2 hanno individuato molti casi di incomprensione su temi analoghi (Rossi e Macagno, 2019; Macagno e Rossi, 2019). Un esempio su tutti riguarda la misurazione della glicemia: non sempre alla comprensione dell'importanza dell'autocontrollo corrisponde un'adeguata comprensione della modalità appropriata di misurazione (a coppie, prima e dopo i pasti) e della valutazione dei risultati ottenuti. Discutere di questi temi on-line può rappresentare ragionevolmente un'opportunità: invece di aspettare la visita successiva che, nel caso del diabete di tipo 2 viene programmata spesso dopo sei mesi, le persone con diabete che frequentano i gruppi on-line hanno l'opportunità di discutere quotidianamente dubbi e timori legati all'interpretazione delle glicemie e, più in generale, alla gestione del diabete. Ulteriori ricerche su campioni più ampi, anche rispetto ai diversi gruppi di discussione on-line dedicati al diabete, si rendono necessarie per corroborare i risultati di questo studio esplorativo e per valutare se e come l'uso di tali risorse on-line comporti effettivamente dei cambiamenti nelle pratiche di gestione della malattia e nei vissuti dei pazienti. Inoltre, non è scontato che ciò che sembra funzionare nel caso del diabete in termini di appropriatezza del supporto e qualità delle informazioni condivise funzioni altrettanto bene per patologie croniche differenti.

7. Conclusioni

Lo studio che abbiamo presentato suggerisce che i gruppi su Facebook possano essere utilizzati come strumenti efficaci di educazione e supporto tra pari. In particolare, sappiamo che le persone con diabete che hanno risposto al questionario, e che utilizzano Facebook, hanno buoni livelli di *engagement*, buone competenze di alfabetizzazione sanitaria e buone risorse informative a disposizione. A questo proposito, sarebbe interessante condurre studi interventistici per analizzare nel dettaglio in che modo la partecipazione all'interno di gruppi su Facebook possa migliorare il livello di *engagement* dei pazienti e quindi favorire un coinvolgimento attivo e autonomo nella gestione del diabete.

Dall'analisi qualitativa che abbiamo condotto sul campione di post pubblicati all'interno di TIDUIR emergono osservazioni interessanti sulle tipologie di supporto (informativo ed emotivo) che è possibile ottenere da una comunità online su Facebook. Anche in questo caso i nostri risultati non possono che essere considerati preliminari: intanto sarebbe opportuno disegnare studi su campioni più ampi, e trasversali a più gruppi di discussione sul diabete e/o riguardanti altri tipi di malattie croniche; si dovrebbe poi pensare di far valutare i contenuti dei post pubblicati a esperti del settore (nel caso del diabete: diabetologi/he, infermieri/e, nutrizionisti/e). Andando in questa direzione, altri ricercatori hanno raccontato di aver ottenuto valutazioni positive sulla qualità delle informazioni pubblicate on-line in gruppi di discussione analoghi (Cole, Watkins e Kleine, 2016); un risultato che, se ulteriormente confermato, apre a scenari ottimistici sull'utilizzo di Facebook e dei social network come strumenti di educazione atti a favorire la partecipazione del paziente diabetico e, più in generale, del paziente cronico nella gestione autonoma del percorso di malattia e di salute.

Riferimenti bibliografici

Anderson, R.M., Funnell, M.M. (2010). Patient empowerment: myths and misconceptions. *Patient Education and Counseling*, *79*, 3, pp. 277-282, https://doi.org/10.1016/j.pec.2009.07.025.

Arsand, E., Bradway, M., Gabarron, E. (2019). What Are Diabetes Patients Versus Health Care Personnel Discussing on Social Media? *Journal of Diabetes Science and Technology*, 1932296818821649, https://doi.org/10.1177/1932296818821649.

Bigi, S. (2014). Healthy Reasoning: The Role of Effective Argumentation for Enhancing Elderly Patients' Selfmanagement Abilities in Chronic Care. In G. Riva, P. Ajmone Marsan e C. Grassi (a cura di), *Active Ageing and Healthy Living: A Human Centered Approach in Research and Innovation as Source of Quality of Life*. Amsterdam: IOS Press, pp. 193-203.

Bodenheimer, T., Lorig, K., Holman, H., Grumbach, K. (2002). Patient self-ma-

nagement of chronic disease in primary care. *JAMA*, *288*, 19, pp. 2469-2475.
Cole, J., Watkins, C., Kleine, D. (2016). Health Advice from Internet Discussion Forums: How Bad Is Dangerous? *Journal of Medical Internet Research*, *18*, 1, e4, https://doi.org/10.2196/jmir.5051.
Colineau, N., Paris, C. (2010). Talking about your health to strangers : understanding the use of online social networks by patients, *4568*, https://doi.org/10.1080/13614568.2010.496131.
Dennis, C.-L. (2003). Peer support within a health care context: a concept analysis. *International Journal of Nursing Studies*, *40*, 3, pp. 321-332.
Gavrila, V., Garrity, A., Hirschfeld, E., Edwards, B., Lee, J.M. (2019). Peer Support Through a Diabetes Social Media Community. *Journal of Diabetes Science and Technology*, 1932296818818828, https://doi.org/10.1177/1932296818818828.
Gómez-Zúñiga, B., Fernandez-Luque, L., Pousada, M., Hernández-Encuentra, E., Armayones, M. (2012). ePatients on YouTube: analysis of four experiences from the patients' perspective. *Medicine 2.0*, *1*, 1, e1-e1, https://doi.org/10.2196/med2.2039.
Graffigna, G., Barello, S., Libreri, C., Bosio, C.A. (2014). How to engage type-2 diabetic patients in their own health management: implications for clinical practice. *BMC public health*, *14*, 1, p. 648, https://doi.org/10.1186/1471-2458-14-648.
Graffigna, G., Barello, S., Bonanomi, A., Lozza, E. (2015). Measuring patient engagement: development and psychometric properties of the Patient Health Engagement (PHE) Scale. *Frontiers in Psychology*, *6*, 274, https://doi.org/10.3389/fpsyg.2015.00274.
Hamm, M.P., Chisholm, A., Shulhan, J., Milne, A., Scott, S.D., Given, L.M., Hartling, L. (2013). Social media use among patients and caregivers: a scoping review. *BMJ Open*, *3*(5), e002819, https://doi.org/10.1136/bmjopen-2013-002819.
Hibbard, J.H., Greene, J. (2013). What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. *Health Affairs (Project Hope)*, *32*, 2, pp. 207-214, https://doi.org/10.1377/hlthaff.2012.1061.
Hibbard, J.H., Stockard, J., Mahoney, E.R., Tusler, M. (2004). Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. *Health Services Research*, *39* (4 Pt 1), pp. 1005-1026, https://doi.org/10.1111/j.1475-6773.2004.00269.x.
Househ, M., Borycki, E., Kushniruk, A. (2014). Empowering patients through social media: the benefits and challenges. *Health Informatics Journal*, *20*, 1, pp. 50-58, https://doi.org/10.1177/1460458213476969.
Islam, S.M.S., Tabassum, R., Liu, Y., Chen, S., Redfern, J., Kim, S.-Y., ... Chow, C.K. (2019). The role of social media in preventing and managing non-communicable diseases in low-and-middle income countries: Hope or hype? *Health Policy and Technology*, https://doi.org/https://doi.org/10.1016/j.hlpt.2019.01.001.
Italian Diabetes e Obesity Barometer (2018).
James, J. (2013). Patient Engagement. *Health Affairs Health Policy Brief*, https://doi.org/10.1377/hpb20130214.898775.
Joseph, D.H., Griffin, M., Hall, R.F., Sullivan, E.D. (2001). Peer coaching: an

intervention for individuals struggling with diabetes. *The Diabetes Educator*, *27*, 5, pp. 703-710, https://doi.org/10.1177/014572170102700511.

Kingod, N. (2018). The tinkering m-patient: Co-constructing knowledge on how to live with type 1 diabetes through Facebook searching and sharing and offline tinkering with self-care. *Health* (London, England: 1997), 1363459318800140, https://doi.org/10.1177/1363459318800140.

Kofinas, J.D., Varrey, A., Sapra, K.J., Kanj, R.V, Chervenak, F.A., Asfaw, T. (2014). Adjunctive social media for more effective contraceptive counseling: a randomized controlled trial. *Obstetrics and Gynecology*, *123*, 4, pp. 763-770, https://doi.org/10.1097/AOG.0000000000000172.

Lorini, C., Santomauro, F., Grazzini, M., Mantwill, S., Vettori, V., Lastrucci, V., ... Bonaccorsi, G. (2017). Health literacy in Italy: a cross-sectional study protocol to assess the health literacy level in a population-based sample, and to validate health literacy measures in the Italian language. *BMJ Open*, *7*, 11, e017812, https://doi.org/10.1136/bmjopen-2017-017812.

Macagno, F., Rossi, M.G. (2019). Metaphors and problematic understanding in chronic care communication. *Journal of Pragmatics*, *151*, pp. 103-117, https://doi.org/10.1016/j.pragma.2019.03.010.

Markham, M.J., Gentile, D., Graham, D.L. (2017). Social Media for Networking, Professional Development, and Patient Engagement. *American Society of Clinical Oncology Educational Book. American Society of Clinical Oncology. Annual Meeting*, *37*, pp. 782-787, https://doi.org/10.14694/EDBK_180077.

Menichetti, J., Libreri, C., Lozza, E., Graffigna, G. (2016). Giving patients a starring role in their own care: a bibliometric analysis of the on-going literature debate. *Health Expectations : An International Journal of Public Participation in Health Care and Health Policy*, *19*, 3, pp. 516-526, https://doi.org/10.1111/hex.12299.

Merolli, M., Gray, K., Martin-Sanchez, F. (2013). Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. *Journal of Biomedical Informatics*, *46*, 6, pp. 957-969, https://doi.org/10.1016/j.jbi.2013.04.010.

Moorhead, S.A., Hazlett, D.E., Harrison, L., Carroll, J.K., Irwin, A., Hoving, C. (2013). A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of Medical Internet Research*, *15*, 4, e85, https://doi.org/10.2196/jmir.1933.

O'Keeffe, D.T., Montori, V.M. (2016). What's up #DOC? The role of social media in diabetes management. *Diabetic Medicine*, *33*, 6, pp. 853-854, https://doi.org/10.1111/dme.12955.

Parchman, M.L., Zeber, J.E., Palmer, R.F. (2010). Participatory decision making, patient activation, medication adherence, and intermediate clinical outcomes in type 2 diabetes: a STARNet study. *Annals of Family Medicine*, *8*, 5, pp. 410-417, https://doi.org/10.1370/afm.1161.

Piette, J.D., Resnicow, K., Choi, H., Heisler, M. (2013). A diabetes peer support intervention that improved glycemic control: mediators and moderators of intervention effectiveness. *Chronic Illness*, *9*, 4, pp. 258-267, https://doi.org/10.1177/1742395313476522.

Rossi, M.G., Macagno, F. (2019). Coding problematic understanding within patient-provider interactions. *Health Communication*. Advance online publication, https://doi.org/10.1080/10410236.2019.1652384.

Rozenblum, R., Bates, D.W. (2013). Patient-centred healthcare, social media and the internet: the perfect storm? *BMJ Quality eamp;Amp; Safety*, *22*, 3, p. 183 LP-186, https://doi.org/10.1136/bmjqs-2012-001744.

Rupert, D.J., Moultrie, R.R., Read, J.G., Amoozegar, J.B., Bornkessel, A.S., O'Donoghue, A.C., Sullivan, H.W. (2014). Perceived healthcare provider reactions to patient and caregiver use of online health communities. *Patient Education and Counseling*, *96*, 3, pp. 320-326, https://doi.org/10.1016/j.pec.2014.05.015.

Smailhodzic, E., Hooijsma, W., Boonstra, A., Langley, D.J. (2016). Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. *BMC Health Services Research*, *16*, 442, https://doi.org/10.1186/s12913-016-1691-0.

Tang, T.S., Afshar, R., Elliott, T., Kong, J., Gill, S. (2019). Study protocol and baseline sample characteristics: From clinic to community: Using peer support as a transition model for improving long-term diabetes-related health outcomes. *Contemporary Clinical Trials*, https://doi.org/https://doi.org/10.1016/j.cct.2019.02.002.

Tang, T.S., Funnell, M., Sinco, B., Piatt, G., Palmisano, G., Spencer, M.S.,... Heisler, M. (2014). Comparative effectiveness of peer leaders and community health workers in diabetes self-management support: results of a randomized controlled trial. *Diabetes Care*, *37*, 6, pp. 1525-1534, https://doi.org/10.2337/dc13-2161.

Tenderich, A., Tenderich, B., Barton, T., Richards, S.E. (2018). What Are PWDs (People With Diabetes) Doing Online? A Netnographic Analysis. *Journal of Diabetes Science and Technology*, 1932296818813192, https://doi.org/10.1177/1932296818813192.

Thompson, A.G.H. (2007). The meaning of patient involvement and participation in health care consultations: a taxonomy. *Social Science e Medicine (1982)*, *64*, 6, pp. 1297-1310, https://doi.org/10.1016/j.socscimed.2006.11.002.

Troncone, A., Cascella, C., Chianese, A., Iafusco, D. (2015). Using computerized text analysis to assess communication within an Italian type 1 diabetes Facebook group. *Health Psychology Open*, *2*, 2, 2055102915615338, https://doi.org/10.1177/2055102915615338.

Troncone, A., Cascella, C., Chianese, A., Iafusco, D. (2016). What Relatives and Caregivers of Children with Type 1 Diabetes Talk About: Preliminary Results from a Computerized Text Analysis of Messages Posted on the Italian Facebook Diabetes Group. In S. Bassis, A. Esposito, F.C. Morabito e E. Pasero (a cura di), *Advances in Neural Networks*. Cham: Springer International Publishing, pp. 235-242.

Weil, A.R. (2016, April). The Patient Engagement Imperative. *Health Affairs (Project Hope)*. United States, https://doi.org/10.1377/hlthaff.2016.0337.

White, K., Gebremariam, A., Lewis, D., Nordgren, W., Wedding, J., Pasek, J., ... Lee, J.M. (2018). Motivations for Participation in an Online Social Media Community for Diabetes. *Journal of Diabetes Science and Technology*, *12*, 3, pp. 712-718, https://doi.org/10.1177/1932296817749611.

WHO (2016). Global report on diabetes. World Health Organization, Geneva. Retrieved October 3, 2018, from http://www.who.int/diabetes/global-report/en/.

Wicks, P., Massagli, M., Frost, J., Brownstein, C., Okun, S., Vaughan, T.,... Heywood, J. (2010). Sharing health data for better outcomes on Pa-

tientsLikeMe. *Journal of Medical Internet Research*, *12*, 2, e19, https://doi.org/10.2196/jmir.1549.
Zhang, Y., Graffigna, G., Bonanomi, A., Choi, K.-C., Barello, S., Mao, P., Feng, H. (2017). Adaptation and Validation of a Chinese Version of Patient Health Engagement Scale for Patients with Chronic Disease. *Frontiers in Psychology*, *8*, 104, https://doi.org/10.3389/fpsyg.2017.00104.
Zhang, Y., He, D., Sang, Y. (2013). Facebook as a Platform for Health Information and Communication: A Case Study of a Diabetes Group. *Journal of Medical Systems*, *37*, 3, 9942, https://doi.org/10.1007/s10916-013-9942-7.

Health and participation: Facebook as an educational tool for engaging patients

With this study we aimed to explore the role of social media in the healthcare context by analysing how Facebook can represent a tool to foster patient education and engagement within the context of chronic care. More specifically, we have explored how Facebook is used by patients with diabetes (and by their relatives) to share information and/or discuss issues relevant for managing their disease. This is a preliminary explorative study, with a quantitative phase where a survey was administered to 119 patients with diabetes and Facebook users, and qualitative phase where a sample of interactions among members of a Facebook group for patients with diabetes and their relatives was inductively analysed. The qualitative analysis showed how participating in the Facebook group has the main function of sharing information on how to manage diabetes, followed by having the function of emotionally support patients. The results from the survey confirmed these findings and revealed how patients using Facebook groups for diabetes are usually more engaged in their care, with good health literacy levels, and low informational needs. Based on these findings, we highlight the potential relevance of these online groups and communities to promote peer support; they can indeed represent important tools to improve the abilities of patients to self-manage their disease and their motivation in playing an active role in the care process.

Keywords: social media, Facebook, diabetes, peer support, patient engagement, health literacy.

Per lo studio presentato in questo contributo le autrici hanno beneficiato della collaborazione con il centro di ricerca "EngageMinds Hub – Consumer e Health Engagement Center" dell'Università Cattolica di Milano. Ringraziamo anche tutti i partecipanti al gruppo di discussione su Facebook "Tutti i Diabetici Uniti in Rete" per aver partecipato e supportato questa ricerca. Questo studio ha beneficiato dei finanziamenti ricevuti dalla Fundação para a Ciência e a Tecnologia in Portogallo (Grants n. SFRH/BPD/115073/2016 e PTDC/FER-FIL/28278/2017).

Maria Grazia Rossi, ArgLab - Instituto de Filosofia da Nova (IFILNOVA), Universidade Nova de Lisboa, Avenida de Berna 26, 1069-061 Lisbon, Portugal, mgrazia.rossi@fcsh.unl.pt

Julia Menichetti, Institute of Clinical Medicine, University of Oslo, Akershus universitetssykehus, 1478, Lørenskog, j.p.m.delor@medisin.uio.no

FABRIZIO MACAGNO

ANALIZZARE L'ARGOMENTAZIONE SUI SOCIAL MEDIA

Il caso dei tweet di Salvini

Twitter, con i suoi 280 caratteri a disposizione per messaggio (*tweet*), è la piattaforma di microblogging più utilizzata e con le relazioni più complesse con il mondo dell'informazione. A livello quantitativo, la maggior parte dei messaggi diffusi attraverso questo strumento sono di natura informativa/giornalistica (Kwak, Lee, Park e Moon, 2010), vale a dire finalizzati a diffondere notizie e commenti su queste ultime. Inoltre, Twitter è usato per condividere idee, informazioni e opinioni personali (Scanfeld, Scanfeld e Larson, 2010). A causa di questo uso preferenziale, l'interazione col mondo dell'informazione tradizionale è duplice. Da un lato, Twitter è usato per condividere e commentare articoli giornalistici; dall'altro, i *tweet* sono sempre più usati come una fonte di informazione che integra articoli giornalistici convenzionali (Cobianchi, Del Sal e Splendore, 2014; Conway, Kenski e Wang, 2013; Ott, 2017).

La brevità e la relazione fluida con il mondo del giornalismo hanno reso Twitter uno strumento estremamente potente nella comunicazione politica (López-Meri, Marcos-García e Casero-Ripollés, 2017). Tramite brevi *tweet*, i politici possono comunicare con il pubblico in modo rapido, fornendo informazioni riguardanti decisioni o discussioni e coinvolgendolo in un dibattito diretto (Conway *et al*., 2013, 1598). Questo ha portato alla coscienza dei rischi potenziali di una comunicazione "liofilizzata", che rischia di diventare parziale (Scarfone, 2017) e quindi potenzialmente pericolosa. Tuttavia, gli studi che al momento più comunemente si interessano dell'analisi dei *tweet* prendono in considerazione solo i macro-dati (*sentiment analysis*, frequenze dei messaggi ecc.), mentre la dimensione argomentativa è quasi del tutto ignorata, confinata a studi o riflessioni sul linguaggio abusivo o offensivo in genere (Lee e Queal, 2016; Mendes, 2016).

L'importanza, l'uso e il potenziale impatto di Twitter all'interno della comunicazione politica rende l'uso argomentativo dei *tweet* e il profilo argomentativo del loro autore cruciale (Hansen e Walton, 2013). Questo articolo propone una metodologia di analisi basata sulla teoria dell'argomentazione, applicandola a un corpus di *tweet* prodotto dal

rappresentante del precedente governo (un utilizzatore della piattaforma che si definisce col suo profilo istituzionale) più attivo nell'uso di questo strumento, vale a dire l'ex Ministro dell'Interno e Vicepresidente del Consiglio dei Ministri italiano, Matteo Salvini[1]. L'analisi, avente come oggetto un corpus di 843 tweet raccolti in sette mesi e mezzo a partire dalla sua nomina a membro del governo, verrà condotta su tre livelli, identificando dapprima i messaggi argomentativi e quindi esaminando le caratteristiche di questi ultimi, al fine di riconoscere i tipi di argomenti e la loro qualità, le fallacie e le parole chiave che li caratterizzano.

1. Background teorico

L'analisi del profilo argomentativo si basa su tre fondamenti teorici, vale a dire la distinzione tra i diversi tipi di argomento, la qualità degli argomenti – che si determina sulla base della relazione tra premesse e *backing* e soprattutto della valutazione dei presupposti su cui un argomento si basa – e l'uso di parole emotive, che possono individuare strategie ridefinitorie che spesso configurano fallacie.

1.1. *Tipi di argomento*

L'analisi argomentativa di un testo si basa in genere su due fondamenti teorici, vale a dire i concetti di *argomento presuntivo* e di *schema argomentativo*. Un argomento è un supporto fornito a una conclusione sulla base di una regola di inferenza, che corrisponde al concetto di *topos* o *locus* nella tradizione dialettica antica (Hitchcock, 1998; Macagno e Walton, 2014a; Rigotti, 2007) e quello di *warrant* nella tradizione argomentativa moderna (Hitchcock, 2003; Toulmin, 1958). A sostegno delle premesse potenzialmente dubbie sono comunemente fornite prove (o evidenze), chiamate anche *backing*. Un argomento fornisce una ragione presuntiva, in quanto 1) esso deve soddisfare un "onere della prova" fornendo ragioni che modifichino lo *status quo* della conclusione; e 2) la conclusione, qualora la prima condizione sia soddisfatta, può considerarsi accettabile fino a prova contraria, cioè fino a quando non siano presentati argomenti contrari e più forti (Walton, 1995b; 2001). Un argomento tuttavia ha una duplice natura, logica e pragmatica. Un argomento non è solamente una inferenza, una relazione logico-semantica tra due proposizioni, ma un atto linguistico complesso (van Eemeren e Grootendorst, 1984, 40-46) che può avere differenti obiettivi comunicativi (Walton, 1990). Per tale

[1] https://www.socialbakers.com/statistics/twitter/profiles/italy/society/politics/ (consultato il 30 dicembre 2018). In generale, Salvini (e quindi la Lega) è l'attore politico in generale più attivo su Twitter (Cepernich e Bracciale, 2018).

ragione, l'analisi argomentativa deve analizzare il contesto e soprattutto la finalità discorsiva per cui un argomento è comunicato, che può essere di natura persuasiva ma anche deliberativa, negoziale, informativa, oppure finalizzata alla scoperta di nuove spiegazioni (*discovery*), all'analisi delle prove (*inquiry*) o alla costruzione di relazioni personali (*eristic* o *rapport building*) (Macagno e Bigi, 2017; Walton, 1990; 2008).

Questa definizione implica l'esistenza di molte e differenti regole di inferenza e quindi di diversi tipi di argomento, che nella letteratura della teoria dell'argomentazione sono normalmente analizzati per mezzo di schemi argomentativi (Macagno e Walton, 2015; Walton, Reed e Macagno, 2008). Gli schemi argomentativi sono rappresentazioni delle strutture degli argomenti più comuni, raffigurati come una successione di premesse che sostengono una conclusione. A questi "schemi" sono associate specifiche domande critiche, vale a dire un insieme di condizioni che, se non soddisfatte, invalidano l'argomento. In tali circostanze l'argomento perde la sua natura presuntiva, cioè non può essere più utilizzato per fornire una ragione, valida fino a prova contraria, per accettare la conclusione. Queste condizioni sono espresse in forma dialogica come domande, in quanto l'onere della loro risposta può variare a seconda delle circostanze dialogiche, cioè delle regole del dialogo in cui un argomento è prodotto (un controesame di un teste ha regole differenti di un dialogo accademico o tra amici). Un esempio di schema argomentativo è il seguente argomento per ragionamento pratico (Walton *et al*., 2008, 94-98):

Premessa 1	Il mio fine è *A*.
Premessa 2	Il migliore modo per ottenere *A* è effettuare l'azione *B* (invece delle sue possibili alternative B_1, B_2, B_3...).
Conclusione	Quindi, è ragionevole fare *B*.

Domande critiche:

1. Gli altri obiettivi che possono entrare in conflitto con *A* sono stati presi in considerazione?
2. Le altre azioni alternative a *B* (e B_1, B_2, B_3...) che possono avere come effetto A sono state prese in considerazione?
3. Tra queste azioni alternative, quale è la più efficiente e la migliore?
4. Su quali basi posso affermare che è possibile da un punto di vista pratico effettuare *B* in queste circostanze?
5. Quali sono le conseguenze di *B* e sono queste accettabili?

In totale, più di 60 schemi argomentativi sono stati analizzati nella letteratura (Macagno e Walton, 2015; Walton *et al*., 2008). Tuttavia, gli

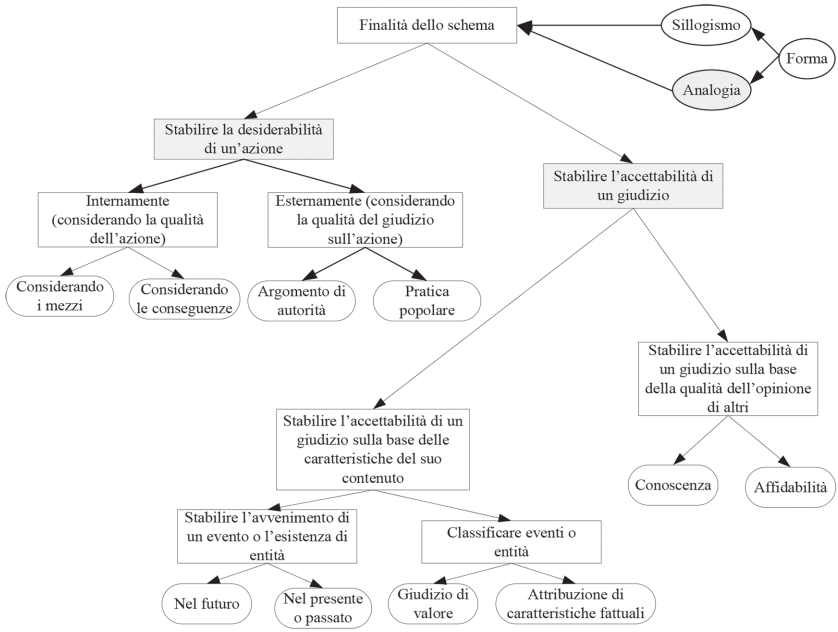

Fig. 1. Tipi di schemi argomentativi.

schemi che costituiscono le componenti di altri schemi più complessi possono essere ridotti alle categorie pragmatiche e semantiche rappresentate nella figura 1.

A queste categorie si aggiungono due altri schemi caratterizzati dalla loro forma, vale a dire l'argomento per analogia e quello per esempio. Il primo esprime una relazione semantica rappresentata dagli schemi sopra classificati tramite un paragone con uno stato di cose eterogeneo a quello espresso dalla conclusione (Macagno, 2017), mentre il secondo per mezzo di una illustrazione, cioè uno stato di cose appartenente alla stessa categoria (lessicalmente definita) della conclusione. Questo schema di annotazione è stato utilizzato e validato in precedenti pubblicazioni in altre aree di studio (Konstantinidou e Macagno, 2013; Macagno e Konstantinidou, 2013).

1.2. *Qualità degli argomenti*

La definizione di argomento sopra esposta unisce un criterio strutturale a uno pragmatico, distinguendosi dalla nozione puramente logica o persino logico-deduttiva che caratterizza praticamente tutti gli approcci tradizionali (Walton, 1990). Questa prospettava ha implicazioni cruciali sui criteri per la valutazione della qualità argomentativa e soprattutto

per la definizione e l'individuazione delle tattiche ingannevoli, comunemente chiamate fallacie. In particolare, questo approccio permette di giustificare la qualità di un argomento tramite considerazioni strutturali e pragmatiche, cioè senza entrare nel merito della valutazione del contenuto espresso dall'argomento, che porrebbe l'analista sullo stesso piano di un partecipante al dialogo.

Da un punto di vista *pragmatico*, un argomento è considerato come uno *strumento verbale* e *sociale* per risolvere una differenza di opinioni o un dubbio, sia di natura cognitiva che pratica. Per tale motivo, un argomento deve essere pertinente al dialogo, al discorso, o alla finalità comunicativa del testo in cui esso è usato (Macagno, 2008; 2018b; Walton, 2004b; Walton e Macagno, 2016), contribuendo allo scopo del discorso (Dascal, 2003, 37-42; Leech, 1983, 99; Macagno, 2018b; Schank *et al*., 1982). Un argomento dialogicamente irrilevante (per esempio un attacco eristico all'oratore in un contesto di discussione critica o di esame delle prove) diventa in tale prospettiva presuntivamente fallace, in quanto modifica il dialogo in essere tra gli interlocutori manipolandone l'obiettivo che questi hanno tacitamente accettato come condiviso (Walton, 2004a). Un argomento può anche essere irrilevante relativamente alla conclusione che si presume che esso debba sostenere (Walton, 1995a, 2008), come per esempio nel caso di argomenti diretti contro un punto di vista che non corrisponde a quello espresso dall'interlocutore.

La prospettiva *pragmatica* si fonde con una considerazione *dialettica* e *strutturale* quando ci si sposta da una valutazione della pertinenza a quella della possibile accettabilità delle premesse. Un argomento, in quanto discorso, si fonda su presupposizioni, proposizioni che il parlante dà per scontate in quanto parte del *common ground*, cioè con quanto è comunemente accettato dalla comunità a cui si rivolge (Atlas, 2008; Stalnaker, 1998). Tuttavia, quando le presupposizioni sono in contraddizione o conflitto con quanto è comunemente condiviso, l'argomento che le veicola diventa manipolatorio (Macagno e Walton, 2017). Infine, un argomento si deve basare su regole specifiche di inferenza (analizzate nella tradizione classica sotto il nome di "luoghi specifici" o "massime;" per un approfondimento sul tema, si veda Macagno e Walton, 2014; Rigotti e Greco-Morasso, 2019; Rubinelli, 2009) e premesse accettate dall'interlocutore. Qualora tali premesse non fossero evidenti, in quanto non comunemente condivise o presuntivamente estranee all'auditorio (come nel caso di premesse riguardanti fenomeni specifici e non conosciuti dai più), esse devono essere fondate su evidenze.

Questo comporta il ruolo fondamentale del criterio *strutturale*, vale a dire l'uso dei *backing* o prove (Kuhn, 1993; 2010, 817). In teoria dell'argomentazione e nell'area di *argumentation and education*, l'assenza dell'uso di evidenze è normalmente considerata come indicazione di una limitata capacità argomentativa (Kuhn, 2010). Nel nostro caso, l'assenza di prove a sostegno di premesse dubbie non può essere pre-

suntivamente addotta a una finalità manipolatoria (come invece nel caso delle irrilevanze argomentative che configurano una fallacia); tuttavia un argomento privo dei *backing* necessari è necessariamente incompleto e presuntivamente inaccettabile (Walton, 2010).

1.3. *Parole emotive e ridefinizioni*

In teoria dell'argomentazione, il concetto di parola emotiva si riferisce all'uso argomentativo di determinati termini per giustificare un giudizio di valore (che a sua volta può essere usato esplicitamente o implicitamente per giustificare una scelta o un'azione) senza fornire ulteriori ragioni (Stevenson, 1937; 1938; 1944). La parola emotiva è tale quando rappresenta il fondamento unico di un giudizio di valore; il problema è che spesso tali termini sono usati in modo fallace, in quando manipolano la conoscenza condivisa dagli interlocutori.

Si possono distinguere tre differenti usi manipolatori delle parole emotive: l'epiteto circolare, la definizione persuasiva e la quasi-definizione. Gli *epiteti circolari* (*loaded words* o *question begging epithets*) (Bentham, 1824, 213-220; Macagno e Walton, 2014b) si riferiscono all'uso indebito di un termine emotivo – il parlante usa una *loaded word* presupponendo alcune caratteristiche del referente che l'interlocutore non conosce o che non può accettare. Le *definizioni persuasive* consistono in ridefinizioni e per questo sono normalmente segnalate dalla strategia della "dissociazione" (Halldén, 1960; Schiappa, 2003; Van Rees, 2008), vale a dire la distinzione tra un significato nuovo del termine (spesso indicato come quello "vero" o "reale") e uno corrispondente all'uso comune (presupposto quindi come "falso"). Le *quasi-definizioni*, infine, consistono nella modificazione di quello che comunemente si descrive come un tipo di "connotazione" (Kecskes, 2003; Kerbrat-Orecchioni, 1977), vale a dire le inferenze valutative (che possono determinare specifiche risposte emotive e affettive) che sono comunemente associate a un termine come conseguenza del contesto in cui è normalmente usato[2].

[2] Il concetto di connotazione è estremamente complesso e spesso controverso (per una panoramica sul concetto, oltre al sopracitato studio di Kerbrat-Orecchioni, si veda Garza-Cuarón, 1991). Nel presente articolo si considera solamente la connotazione in quanto significato distinto dalla definizione lessicale (Mel'čuk, 2015, 283; Mel'čuk e Iordanskaja, 2009):

"A meaning 'σ' is a lexical connotation of an LU L of language L if and only if 'σ' satisfies simultaneously Conditions 1-2:

'σ' is associated by language L with the denotation of L and has observable linguistic manifestations in L.

'σ' is not part of L's lexicographic definition."

Tuttavia, il criterio 1 è stato limitato ai significati che possano essere provati – cioè quelli derivante dai precedenti contesti d'uso frequente (Kecskes, 2008; 2013; Kecskes e Zhang, 2009). Chiaramente tale approccio è limitativo, in quanto esclude la dimensione

Nello specifico, le parole emotive sono state sottoposte a una duplice analisi, semantica e connotativa, prendendo in considerazione sia la loro definizione, e quindi il raffronto tra le condizioni codificate del loro uso e il loro effettivo utilizzo nel testo in esame, sia il loro "significato emotivo" (Macagno e Walton, 2014b; Stevenson, 1937; 1944), cioè le inferenze valutative che tali termini innescano automaticamente come risultato del loro uso preferenziale. Tale significato emotivo è stato ricostruito sia in base a indicatori morfologici (come nel caso di derivazioni create a tale scopo, si veda le sezioni 5.4 e 6.4) sia, nel caso di dubbio, studiando i contesti tipici d'uso del termine analizzato, attraverso strumenti automatici di analisi di corpora (in particolare il software SketchEngine). Le tre tipologie di usi fallaci delle parole emotive sono descritte nella tabella 1 (la parola emotiva è indicata come PE).

Tab. 1. *Strategie per l'uso delle parole emotive*

Strategia basata su parole emotive	Definizione	Esempio
Epiteti circolari	L'uso di PE presuppone uno stato di cose *x* che non è stato precedentemente provato o che non è accettato.	Il suo comportamento è solamente *sincero* e perfettamente *naturale*, privo di ogni *ipocrisia* che caratterizza l'uomo moderno (parlando di un *adultero*; tratto da Artsybashev, 1915, 27)
Definizioni persuasive	La definizione di PE è stata modificata per poter designare uno stato di cose *x* che normalmente non sarebbe ricaduto dentro la sua definizione, e per poter giustificare una conclusione valutativa su di esso.	Se chiami la *prigione* "*vera libertà*," le persone ne verranno attratte (Huxley, 2010, 91, traduzione mia)
Quasi-definizioni	Il significato emotivo di PE, o meglio le inferenze valutative automaticamente innescate dall'uso di PE, sono modificate tramite la ricontestualizzazione del termine, cioè il suo ripetuto uso in contesti che associno ad essa specifiche conclusioni valutative.	Ho sempre disprezzato il pregiudizio che conferisce un odioso significato al nome di spia: questo nome non suona male che alle orecchie di chi non ama il Governo: *uno spione non è altro che un amico del bene dello stato, il flagello dei delinquenti, il fedel suddito del suo Principe* (Casanova, 1911, 112)

soggettiva e affettiva che spesso è analizzata in termini appunto di connotazione. Tuttavia essa permette di poter predire determinate inferenze a partire dall'uso di specifiche parole e soprattutto poter confermare tali ipotesi predittive per mezzo dell'analisi di *corpora* (Macagno, *submitted*).

Questi fondamenti teorici sono stati utilizzati sia direttamente (analisi dei tipi di argomento, per cui esiste già una procedura usata in precedenti lavori) sia attraverso la loro declinazione in una procedura per annotare un corpus di messaggi pubblicati su Twitter.

2. Metodologia: costruzione del corpus argomentativo

Al fine di delineare un profilo comunicativo – e nello specifico argomentativo – dell'utilizzatore della piattaforma, è stato raccolto un corpus rappresentativo di *tweet*. Il periodo è stato determinato considerando il profilo pubblico dell'utilizzatore, vale a dire il momento in cui Matteo Salvini è stato nominato Ministro dell'Interno (1° giugno 2018). A partire da questa data, l'utilizzatore si identifica pubblicamente sul suo profilo di Twitter come "Ministro dell'Interno e Vicepresidente del Consiglio;" quindi si presume che le sue comunicazioni, seppure pubblicate attraverso il suo profilo personale e non istituzionale, riflettano il suo ruolo di membro del governo (peraltro ribadito nella maggior parte dei *tweet*) e non quello di leader politico della Lega. I *tweet* sono stati raccolti durante un arco temporale in cui il profilo pubblicamente denunciato si è mantenuto distinto (seppure con dovute eccezioni) da quello di capo di uno specifico partito politico, vale a dire fino alla data del 12 gennaio, data in cui i *tweet* iniziano a esprimere esplicitamente messaggi propagandistici ai fini delle elezioni europee (e in cui si invitano i cittadini in quanto *elettori* a votare la Lega). Questo periodo (che ammonta a 225 giorni) è piuttosto omogeneo dal punto di vista del ruolo dell'utilizzatore, o perlomeno non è esplicitamente ambiguo come nel caso dei *tweet* successivi, in cui diventa complesso o impossibile distinguere Salvini in quanto ministro dal Salvini in quanto politico in permanente campagna elettorale (Blumenthal, 1982).

La costruzione del corpus è stata facilitata dal programma Chorus, un software per la raccolta e analisi dei dati di Twitter (Brooker, Barnett e Cribbin, 2016). Chorus estrae automaticamente i *tweet* sia secondo modalità temporali (in un periodo prescelto) sia per parole chiave, elencando i dati raccolti in una schermata riproducente i primi 100 caratteri di ciascun messaggio. Considerando le limitazioni programma (che visualizza un numero limitato di dati), i *tweet* relativi al periodo mancante (1° giugno-19 giugno) sono stati estratti manualmente. In totale, i messaggi raccolti ammontano a 3327 (circa 15,5 *tweet* al giorno).

Questo corpus iniziale è stato sottoposto a un'analisi preliminare basata sui primi 100 caratteri visualizzati dal software, che ha portato a una riduzione della base di dati ai soli messaggi di natura argomentativa. La selezione dei *tweet* argomentativi si è basata sui seguenti criteri negativi:

1. Criterio formale 1. I *tweet* che propongono unicamente rimandi ad articoli o contenuti scritti da terzi sono stati esclusi, in quanto non propongono esplicitamente un argomento.

2. Criterio formale 2. I *tweet* che riproducono nella totalità o in parte sostanziale messaggi precedenti sono stati esclusi, in quanto non riflettono un supporto originale a una opinione o conclusione, ma piuttosto una finalità comunicativo-retorica di ribadire o rinforzare una ragione.

3. Criterio contenutistico-pragmatico. I *tweet* finalizzati a esprimere sensazioni, emozioni, valutazioni oppure informazioni sia di natura personale che pubblica sono stati esclusi in quanto non presuntivamente argomentativi (Macagno e Bigi, 2017, validato da Macagno e Bigi, forthcoming).

4. Criterio contenutistico-strutturale. I *tweet* che non rispecchiano la struttura argomentativa base (completa o parziale) delineata da Toulmin (vale a dire *conclusioni* sostenute da una o più *premesse* e da una regola di inferenza o *warrant*, Toulmin, 1958) sono stati esclusi in quanto non presuntivamente argomentativi (Dusmanu, Cabrio e Villata, 2017)

A questi quattro criteri si aggiunge un quinto criterio, questa volta positivo:

5. Criterio pragmatico-strutturale. Sono considerati argomentativi i *tweet* che: a) riportano opinioni o informazioni fattuali, quando queste sono usate come premesse o conclusioni; oppure b) esprimono conclusioni espresse come domande retoriche (presuntivamente finalizzate a persuadere) (Bosc, Cabrio e Villata, 2016). Il concetto di "informazione fattuale" si riferisce a un contenuto proposizionale potenzialmente verificabile o a un discorso diretto o indiretto.

L'analisi basata su questi cinque criteri ha portato alla costruzione di un corpus di *tweet* argomentativi costituito da 843 messaggi originali. Il metodo di raccolta e selezione è rappresentato graficamente nella figura 2.

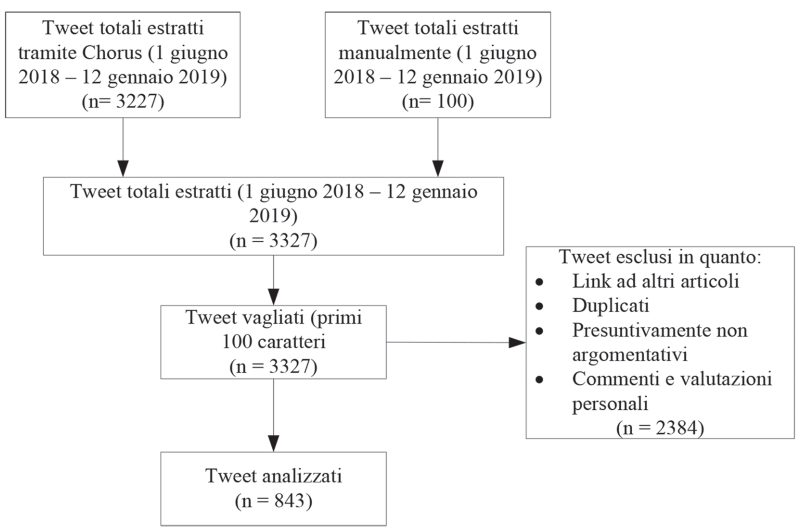

FIG. 2. Costruzione del corpus.

I messaggi di natura argomentativa rappresentano il 25% circa di tutti i *tweet*; il restante 75% è stato escluso in quanto di natura puramente espressiva o informativa (condivisione di opinioni, sentimenti, o dati e fatti di natura personale o istituzionale). I *tweet* che costituiscono il corpus argomentativo di riferimento non sempre sono di natura meramente verbale. Alcuni includono elementi visuali o multimediali. Tuttavia, per le finalità di questo studio, solamente la componente verbale è stata analizzata; la componente non verbale è stata considerata solamente ai fini della disambiguazione dei messaggi (per esempio, nei casi di riferimenti diretti alle immagini riportate) o della valutazione della correttezza di eventuali citazioni o comunicazioni di opinioni di altri (per esempio, quando il *tweet* commenta i contenuti espressi in un video).

3. Metodo: annotazione della qualità degli argomenti

Mentre gli schemi argomentativi sono usati per annotare *corpora* (si veda per esempio Konstantinidou e Macagno, 2013a), nella letteratura in teoria dell'argomentazione la procedura per l'analisi della qualità degli argomenti è lasciata implicita. Per determinare la posizione di un

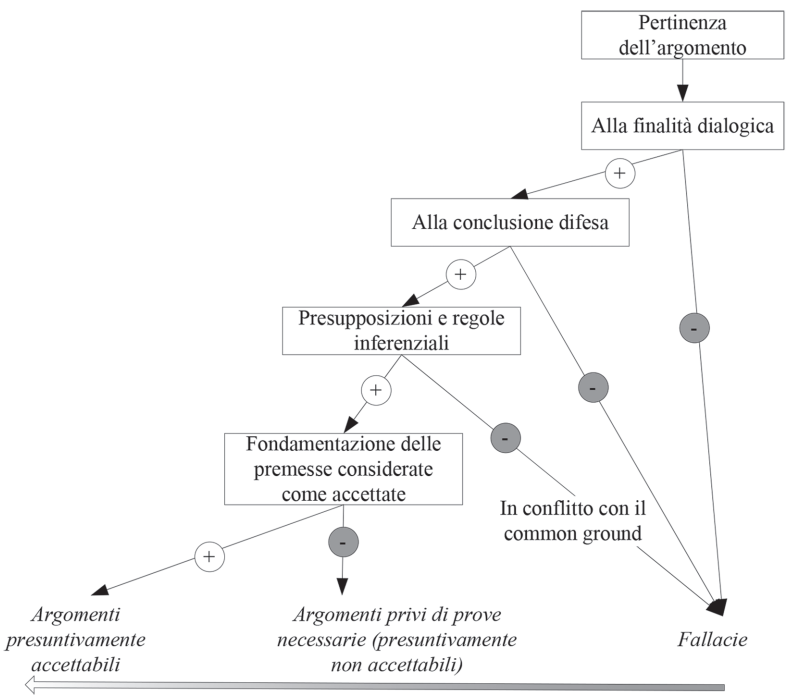

Fig. 3. Valutazione degli argomenti.

argomento nel *continuum* menzionato nella sezione 1.2, è stata quindi delineata una successione di valutazioni che, a seconda della risposta, conducono a una fase di analisi ulteriore o a un giudizio. Questa procedura è rappresentata nella figura 3 alla pagina precedente.

Mentre gli argomenti privi di prove (cioè sostegni o *backing*) segnalano necessariamente l'estrema debolezza di un argomento (Erduran, Simon e Osborne, 2004; McNeill e Krajcik, 2008) – e che quindi *può* essere a sua volta spiegato come tentativo di manipolare gli interlocutori – le fallacie sono *forte indizio* di una intenzione manipolatoria. I tre passaggi di analisi sopra indicati individuano le seguenti fallacie:

TAB. 2. *Fallacie e evidenze testuali*

Evidenza testuale	Fallacia	Esempio
Irrilevanza pragmatica (finalità del dialogo)	*Ad hominem* (un attacco personale usato per confutare un argomento o rifiutare un punto di vista)	Sei una persona immorale (incoerente), quindi quel che dici non può essere accettato.
Irrilevanza logico-semantica (conclusione difesa)	*Straw man* (o manipolazione del punto di vista dell'interlocutore per poterlo attaccare più facilmente)	Luigi mi ha detto che sono incoerente. Queste minacce di morte non sono accettabili.
Presupposizioni fattuali in conflitto col *common ground*	Falsa dicotomia (concetti contrari o azioni alternative presupposte come contraddittorie)	Dobbiamo costruire nuovi carceri, se non vogliamo mettere tutti in libertà.
Regole di inferenza presupposte in conflitto col *common ground*	*Post hoc ergo propter hoc* (giustificazione di un nesso causale tramite una successione temporale o spaziale)	La principessa Diana è morta in un luogo che anticamente era sacrificale. Quindi è morta per un sacrificio.
	Generalizzazione indebita	Questo animale è un gatto e ha il pelo nero; quindi tutti i gatti hanno il pelo nero.
	Secundum quid (ignorare le qualificazioni necessarie per inferire una conclusione da una generalizzazione)	Tutti hanno il diritto alla difesa personale. Quindi anche Bob, che è un bambino, ha il diritto di comprare una pistola.

Queste procedure e questi indizi testuali sono usati per annotare e individuare le mosse che sono più strettamente legate a intenti manipolatori, mentre gli argomenti privi delle prove (*backing*) necessarie sono stati contrassegnati con un codice distinto e quindi conteggiati a parte.

4. Analisi linguistica

Il corpus è stato sottoposto a un trattamento automatico lessicale e a un'annotazione manuale. L'analisi automatica si è limitata unicamente alla dimensione delle parole chiave, vale a dire quelle parole che, ai sensi della definizione fornita dalla *corpus linguistics*, sono caratterizzate da una ricorrenza in un dato corpus significativamente maggiore rispetto a quella nel corpus considerato di riferimento (cioè il corpus che rispecchia l'uso comune) (Kilgarriff, 2012). Al fine di identificare le parole chiave utilizzate da Salvini nei suoi *tweet*, il corpus descritto nella sezione 1.1 è stato analizzato tramite il programma Sketch Engine.7 (Kilgarriff *et al*., 2014). Le *keyword* sono state automaticamente calcolate considerando il valore denominato *keynessness*, che si riferisce al rapporto tra la frequenza normalizzata di un lemma nel corpus target (il nostro corpus) e nel corpus di riferimento (in questo caso il corpus italiano precostituito nel programma, chiamato *Italian Web 2016* e contenente oltre 200 milioni di parole).

Un'annotazione più dettagliata e specifica è stata effettuata in un secondo momento, considerando le proprietà semantiche e le inferenze comunemente giustificate dall'uso di specifiche parole. Le parole più frequentemente usate per sostenere conclusioni sono state esaminate distinguendo la loro specifica funzione e isolando quelle utilizzate per giustificare giudizi di valore o decisioni (le parole emotive).

La procedura usata per l'identificazione e l'analisi di una parola emotiva si è basata su 3 passaggi. In primo luogo, viene determinata la natura pragmatica e la struttura argomentativa del *tweet*. Nella quasi totalità dei casi, le parole emotive ricorrono in strutture costituite da una opinione (*x* è accettabile/deve essere fatto) giustificata tramite un giudizio morale (*x* è buono da un punto di vista etico/logico...). Il secondo passo consiste nell'individuazione del fondamento di tale giudizio morale. Quando il giudizio si basa unicamente sull'uso di specifici termini che generano inferenze a livello di giudizi di valore, questi possono essere considerati "termini emotivi."

L'ultimo passo consiste nell'analisi della giustificazione dell'uso delle parole emotive, che coinvolge 3 fattori. In primo luogo, si esamina se lo stato di cose a cui la parola emotiva si riferisce è condiviso universalmente dal pubblico – vale a dire se ci sono prove che esso non lo sia (per esempio informazioni terze, dichiarazioni nel corpus, etc.). In secondo luogo, qualora si riscontri una differenza tra quanto presupposto dall'uso del termine e il *common ground*, si analizza la definizione da dizionario del termine, che è confrontata con quella che emerge nel caso specifico. Nel caso in cui emerga una discrepanza, si valuta se le condizioni che legittimano una classificazione come *definizione persuasiva* (assenza di una ridefinizione esplicita nel messaggio; condivisione dello stato di cose designato da parte del parlante e del

pubblico) o come *epiteto circolare* sono soddisfatte. Come ultimo passo, qualora i primi due test risultassero negativi, si analizza l'inferenza e si determina se la conclusione valutativa è comunemente associata all'uso del termine o se invece quest'ultimo è stato quasi-definito. A tale fine si cercano i migliori esempi (da dizionario) nel corpus di riferimento in Sketch Engine (valore determinato automaticamente nella piattaforma) e si analizzano le collocazioni (vale a dire il contesto in cui il termine è utilizzato). Qualora i giudizi normalmente inferiti sono differenti da quelli riscontrati nel messaggio in analisi o polarmente incompatibili con questi, si valuta la possibilità di una *quasi-definizione*.

5. Risultati: analisi argomentativa

I risultati quantitativi forniscono una prospettiva generale sulle caratteristiche argomentative e le strategie più frequenti usate dall'ex Ministro dell'Interno nelle sue comunicazioni su Twitter. Questa sezione prende in considerazione due dimensioni interconnesse dello stile argomentativo, vale a dire a) i tipi di argomento usati, assieme alla loro frequenza e alla percentuale di argomenti completi, vale a dire fondati sulle evidenze necessarie; e b) le fallacie commesse.

5.1. *Analisi quantitativa: tipi di argomento*

L'analisi argomentativa del corpus mostra specifiche tendenze. In primo luogo, gli argomenti identificati ammontano a 800, comprendenti sia gli argomenti valutati come privi di evidenze necessarie (466), sia quelli che non si configurano come incompleti (334). Si riporta nella tabella 3 la classificazione degli argomenti più usati con a fianco l'indicazione della loro frequenza.

Tab. 3. *Tipi di argomento utilizzat*i

Tipo di argomento	Occorrenze	Frequenza (%)
Argomento per ragionamento pratico	175	22.0
Argomento per conseguenze	114	14.2
Argomento per valori	109	13.6
Argomento per segno (carattere)	106	13.2
Argomento per classificazione per conseguenze	82	10.2
Argomento per commitment	75	9.4
Argomento per classificazione	41	5.1
Argomento per esempio	21	2.6
Argomento per opinione popolare	20	2.5
Argomento per regole	19	2.4
Argomento per causa-effetto	15	1.8
Argomento per autorità	13	1.6
Altri	10	1.2

Da questa tabella si nota come gli argomenti più frequenti sono quelli finalizzati alla giustificazione di una decisione (ragionamento pratico, conseguenze, regole, *commitment*), che nel loro insieme costituiscono quasi il 50% del totale degli argomenti. Gli argomenti finalizzati a giudicare uno stato di cose come positivo o negativo, cioè l'argomento per valori e per classificazione per conseguenze, raggiungono quasi il 25% del totale. In particolare, questi due argomenti sono perlopiù usati come argomenti a sostegno degli argomenti decisionali, in quanto hanno come conclusione la valutazione di uno stato di cose che viene poi identificato come obiettivo o come conseguenza da evitare o perseguire. Un dato estremamente interessante riguarda l'uso dell'argomento per segno, che è usato nel corpus quasi solamente per giustificare una qualità del carattere e in oltre la metà dei casi (60 occorrenze) il carattere positivo di Salvini stesso. A questi argomenti si devono aggiungere quelli per *commitment* che non si limitano a giustificare un'azione sulla base dell'impegno preso, ma evidenziano le qualità positive dell'impegno di Salvini (66 occorrenze), giustificando una valutazione positiva del suo carattere. Nell'insieme, questi due usi degli argomenti per la costruzione retorica del carattere del parlante ammontano a circa il 16% degli argomenti. Chiaramente questi dati si riferiscono unicamente agli usi argomentativi dei *tweet*, escludendo tutti i casi in cui gli slogan ("dalle parole ai fatti" etc.) possono avere questa funzione valutativa, ma compaiono isolatamente e sono presuntivamente finalizzati a commentare su contenuti veicolati in modo non testuale, ipertestuale o extra-testuale (collegamenti ipertestuali, immagini, etc.).

5.2. *Analisi qualitativa: tipi di argomento*

La strategia argomentativa in senso stretto che risulta più comune nel corpus consiste nell'associazione di due tipi di schemi argomentativi, quelli deliberativi (per conseguenze o per ragionamento pratico) e quelli classificatori (per valori o per valutazione). Da un punto di vista strutturale, Salvini tende a non fornire in perlomeno uno di questi due argomenti i fondamenti che renderebbero ragionevole l'argomento complesso risultante. Un esempio tipico è il seguente:

Esempio 1
Se per dare un Futuro ai nostri figli, costretti a scappare all'estero, dovrò ignorare uno "zerovirgola" imposto da Bruxelles, per me quel numero vale poco. Prima viene la felicità dei nostri giovani.

In questo caso, Salvini usa un ragionamento pratico (fine-mezzo) in cui una decisione (ignorare il tetto massimo del rapporto deficit/PIL) è giustificata tramite un fine valutato positivamente da tutti (felicità

dei giovani; dare un futuro). Il duplice argomento presenta un'assenza completa dei fondamenti necessari in quanto né gli effetti, né l'unicità della possibile azione da intraprendere (ignorare lo "zerovirgola"), né tantomeno il giudizio di valore sulle condizioni ("zerovirgola," che designa un accordo preso su un tema economico) sono giustificati.

5.3. *Analisi quantitativa: fallacie*

L'analisi delle fallacie commesse rivela un numero di argomenti incorrettamente utilizzati (e quindi presuntivamente manipolatori) quasi pari a quello dei *tweet* (806). Considerando anche il fatto che in alcuni casi più di una fallacia è commessa all'interno dello stesso messaggio, l'incidenza degli "argomenti" ingannevoli nei messaggi rimane estremamente alta, vicina al 90% dei *tweet*. I dati sono riportati nella tabella 4.

TAB. 4. *Tipi di fallacie commesse*

Fallacia	Occorrenze	Frequenza nel corpus (%)
Ad hominem	350	41.5
Straw man	145	17.2
Falsa dicotomia	133	15.8
Generalizzazione indebita	98	11.6
Post hoc	50	5.9
Secundum quid	30	3.5

Questi numeri mostrano come la fallacia più frequente sia quella *ad hominem*, dato che rispecchia un atteggiamento argomentativo estremamente aggressivo. Tuttavia, se si analizza più nel dettaglio questa mossa, possiamo identificare specifiche strategie. Salvini non usa solo attacchi diretti, vale a dire attacchi personali basati su epiteti peggiorativi, che comunque rappresentano più della metà degli *ad hominem*. A questi egli associa tre tattiche che sono finalizzate a perseguire una strategia duplice, vale a dire la critica del carattere della vittima dell'attacco e la caratterizzazione del carattere del parlante come positivo. Le tre strategie sono:

a) la critica di incoerenza o parzialità (che è contrapposta alla coerenza e integrità del parlante);

b) la vittimizzazione (che presuppone che il parlante non abbia le colpe e i difetti di cui è accusato, e anzi è innocente, tale da suscitare pietà e simpatia); e infine

c) la ridicolizzazione (che manifesta la superiorità del parlante) (Macagno, 2013; Walton, 1998).

Le distribuzioni di questi attacchi personali sono rappresentate nella tabella 5.

Tab. 5. *Tipi di ad hominem*

Tipi di *ad hominem*			
Diretto	Accusa di incoerenza e parzialità	Ridicolizzazione	Vittimizzazione
191	74	59	26
54.6%	21.1%	16.8%	7.4%

Questa distribuzione mostra una tendenza strategica alla difesa, soprattutto se si analizzano gli epiteti e le accuse che caratterizzano gli *ad hominem*. Nel 45% dei casi Salvini attacca gli opponenti in nome di valori che non possono non essere condivisi dagli interlocutori, vale a dire *coerenza* e *integrità* (seconda categoria), *intelligenza* (terza categoria) e *innocenza* (ultima categoria). Tuttavia, se andiamo ad analizzare le parole emotive usate negli attacchi diretti personali, notiamo che gli insulti veri e propri sono in totale 27 e perlopiù riferiti agli autori di atti e o dichiarazioni a loro volta offensivi nei confronti dell'ex Ministro. Tali epiteti quindi sono giustificabili in quanto inseriti in un contesto eristico, in cui Salvini risponde con insulti a insulti. Il problema da un punto argomentativo è il resto.

Gli *attacchi personali* fallaci sono basati su valori condivisi e condivisibili da tutti i lettori. Per esempio troviamo spesso accuse implicite di superbia (neologismi tramite suffissazione in *-one* – 13 occorrenze), parole e neologismi che indicano assenza di qualità morali ("centri asociali" – 6; "buonisti" – 21; "amici di <categorie moralmente disprezzate>" – 5; "radical/figli di papà" – 9; "benpensanti" o persone "perbene" – 13), etico-professionali ("chiacchieroni" – 11; categorie come "miliardari," "burocrati," "banchieri," "speculatori," ricchi in genere – 34), o logiche (tra cui epiteti come "incompetente," "fenomeno," "incapace," "ignorante" – 58). La strategia comune è quella di un attacco, ma in nome di valori per cui è accettata l'indignazione. Tali *ad hominem* sono soprattutto usati per attaccare categorie di cui Salvini stesso si presenta implicitamente come l'alternativa. In questo modo, il suo attacco personale permette una inferenza implicita a sostegno della bontà del suo carattere, vale a dire moralità, integrità, coerenza, umiltà, e intelligenza. Mentre nessuna prova è addotta per giustificare tali caratteristiche del suo *ethos*, l'attacco agli oppositori in nome di questi valori etici presuppone (nel caso della ridicolizzazione) o implica (nel caso di attacchi personale agli oppositori) il loro possesso da parte di chi propone le idee che li incarnano (la giusta indignazione, secondo Tommaso D'Aquino, può essere buona in quanto avviene in nome della giustizia).

5.4. *Analisi qualitativa: fallacie*

L'analisi qualitativa delle fallacie identifica alcuni usi preferenziali delle mosse manipolatorie più frequenti. Iniziando dalla fallacia più frequente, l'*ad hominem* abusivo, possiamo fornire il seguente *tweet* che illustra una strategia tipica:

Esempio 2
I giornali di sinistra sostengono che avrei usato parole di... Hitler. Ma vergognatevi, cretini!

Questa strategia è estremamente comune nel corpus, e consiste nell'attacco personale (in questo caso tramite un insulto diretto e una espressione di scandalo) che vuole rispondere e confutare un giudizio basato su informazioni fattuali e verificabili (Walton, 1998). Salvini, invece di attaccare i fondamenti del giudizio, ne attacca gli autori. Questo attacco è totalmente irrilevante, in quanto gli autori non si erano posti come le autorità su cui si basava il fondamento del giudizio. L'effetto è quello di rispondere a una mossa dialogica di tipo persuasivo (fondata su ragioni) con una mossa eristica, in cui l'obiettivo è attaccare l'interlocutore. L'inosservanza del criterio pragmatico giustifica la classificazione di questo *tweet* come fallace (*ad hominem*).

Questo esempio nasconde un'ulteriore strategia che consiste nell'uso inappropriato delle presupposizioni (per uno studio delle presupposizioni nei messaggi su Twitter, si veda Brocca, Garassino e Masia, 2016). Il termine "vergognarsi" presuppone una valutazione negativa dell'agente (Ben-Ze'ev, 2000, 525): l'agente deve aver commesso un'azione (in questo caso diffamare una persona) che è comunemente giudicata come immorale. Questo giudizio tuttavia è ingiustificato, dal momento che Salvini non fornisce nessuna prova a fondamento né della falsità dell'affermazione (I giornali hanno davvero mentito? Cosa ha veramente affermato Salvini?). La presupposizione ha come effetto inserire la proposizione presupposta nel *common ground* degli interlocutori (Macagno, 2018a; Macagno e Walton, 2017), i quali si trovano ingiustificatamente ad avere accettato un giudizio su cui mai si sono espressi. Questa strategia pragmatica ricade nell'ambito delle strategie argomentative fallaci senza qualificarsi come fallacia per sé, essendo essa il risultato o il fondamento di altri argomenti ingannevoli. Essa è quindi genericamente classificata come "presupposizione indebita." A livello quantitativo, le presupposizioni indebite caratterizzano il 28% dei *tweet* argomentativi di Salvini (240 occorrenze totali). In pratica, in poco meno di un terzo dei suoi messaggi argomentativi, l'ex Ministro dell'Interno manipola le conoscenze condivise dai suoi interlocutori, dando per scontate informazioni infondate o false, che il suo pubblico non conosce e che non potrebbe accettare.

Una fallacia che caratterizza il profilo argomentativo di Salvini e che illustra l'uso della presupposizione indebita è la fallacia di *falsa*

dicotomia, commessa in oltre il 15% dei *tweet* argomentativi. Tale mossa consiste nel presupporre la contraddittorietà di due concetti che possono essere solamente contrari, al fine di giustificare o una decisione o un giudizio di valore (Macagno e Walton, 2011). Un uso prototipico di questa strategia presupposizionale è il seguente:

Esempio 3

Ancora sciocchezze da Bruxelles. Perché la signora commissaria non viene a visitare un Campo Rom a Roma o Milano, fra armi, illegalità, bambini sfruttati e oggetti rubati? Così magari si sveglia e cambia idea. Io voglio ordine e regole. Punto.

Questo messaggio presenta due chiari esempi di presupposizione indebita. Il primo è rappresentato dall'uso dell'avverbio "ancora" e dall'organizzazione della struttura informativa della frase (Atlas, 1991; 2004; Beaver, 2010; Gundel e Fretheim, 2004, 177; Strawson, 1971). In questo caso, l'avverbio attiva la presupposizione che l'evento descritto si sia verificato nel passato e che l'evento abbia le caratteristiche usate per caratterizzarlo. In questo caso, Salvini considera che i suoi lettori condividano il "fatto" che a Bruxelles abbiamo commesso atti qualificati come "sciocchezze" e che il fatto in questione (una posizione riguardante le iniziative riguardanti i Rom) sia una "sciocchezza." Visto che non esistono prove di tale condivisione globale di opinione (e quindi non si può escludere una opinione contraria) e considerato il fatto che il giudizio di valore sull'ultima opinione non può essere condiviso (in quanto Salvini apparentemente informa il suo pubblico di tale presa di posizione), tale presupposizione è da considerarsi come indebita[3]. Simile analisi si applica al predicato "svegliarsi," che genera una presupposizione (la condizione di "sonno <della ragione>" di Bruxelles) non giustificata né universalmente condivisa.

La presupposizione è utilizzata anche come mezzo per introdurre implicitamente generalizzazioni che altrimenti sarebbero difficilmente accettabili. Un esempio è il seguente *tweet*, in cui si combinano diverse strategie manipolatorie:

Esempio 4

ROBA DA MATTI! Ci sono 119mila toscani in condizioni di povertà assoluta, ma il governatore sinistro annuncia un "ricorso" contro il #DecretoSicurezza, una legge che dà più legalità, risorse e strumenti agli amministratori locali. Lui pensa ai clandestini, noi agli italiani!

[3] Tale analisi è valida se si considera il *tweet* come un messaggio autonomo. Salvini usa costantemente auto-rimandi e auto-citazioni e tale presupposto potrebbe essere meno problematico se inserito nel complesso dell'intera attività sulla piattaforma social. In tale ottica, l'utilizzatore avrebbe potuto riferirsi a precedenti messaggi propri – ipotesi che tuttavia rimane da verificare, come d'altra parte la correttezza dei presupposti di questi ultimi.

Questo esempio illustra tre fenomeni argomentativi distinti. Il più evidente consiste in una replica irrilevante, in quanto a un giudizio basato su ragioni (ricorso), Salvini replica con un attacco personale sia implicito (lui pensa ai clandestini – quindi non pensa agli italiani) che esplicito (attraverso il giudizio valutativo ingiustificato "da matti" e l'aggettivo valutativo "sinistro"), vale a dire come un duplice *ad hominem*.

Il secondo fenomeno è più complesso, in quanto si basa sull'uso indebito della presupposizione. Salvini ricostruisce il pensiero del governatore toscano come basato su due categorie, gli "italiani" e i "clandestini:" secondo Salvini, dal momento che il governatore ha come fine la protezione dei clandestini, si oppone al decreto che tutela gli italiani e quindi rinuncia e anzi si oppone alla protezione degli interessi dei toscani "poveri." Questa implicita dicotomia è pragmaticamente ambigua, in quanto il "punto di vista" o la caratteristica che accomuna i due predicati è lasciato all'interpretazione del lettore. Due ricostruzioni sono possibili. Da un lato, gli "italiani" e i "clandestini" sono presentati come incompatibili dal punto di vista dei diritti e soprattutto del decreto sicurezza a cui il governatore si oppone: chi si oppone al decreto, implica Salvini, protegge i clandestini, e quindi va contro gli interessi degli italiani. Questa dicotomia è presupposta ma non può essere condivisa o persino condivisibile (il decreto pregiudica gli interessi e i diritti non solo degli immigrati irregolari, ma anche dei richiedenti asilo e di coloro che sono in possesso di un regolare permesso di soggiorno). In tal senso, la categoria dei clandestini è generalizzata in modo manipolatorio così da coincidere con quella degli immigrati. Tuttavia, questa contrapposizione non è l'unica che l'ex Ministro prende come comunemente accettata dal pubblico. Egli contrappone l'opposizione al decreto (e quindi la tutela dei "clandestini") agli interessi dei "toscani poveri" (senza specificare quanti immigrati rientrano nella categoria di "toscano" ai fini di tali statistiche), come se il rifiuto di ridurre i diritti di una parte della popolazione avesse come effetto automatico la rinuncia o l'impossibilità di aiutare una differente categoria di persone.

Il terzo fenomeno argomentativo è l'argomento per *causa ad effetto*. Quando Salvini argomenta che il decreto sicurezza è "una legge che dà più legalità, risorse e strumenti agli amministratori locali," la pertinenza della frase relativa è giustificabile solo attraverso la sua funzione argomentativa, in quanto dovrebbe fornire una ragione a sostegno dalla contrapposizione tra il decreto e le persone in povertà assoluta. Tuttavia, l'ex Ministro dell'Interno non fornisce prove o ragioni che giustifichino come un aumento della *legalità* (oltre che dei mal definiti "strumenti" e "risorse") possa contrapporsi alla *povertà assoluta*. In tal caso, l'argomento è privo di una sua componente essenziale, la giustificazione di un nesso causale che non può essere dato per scontato. Nello spettro della valutazione degli argomenti, tale fenomeno rappresenta un argomento incompleto. Mentre non lo si può considerare come presuntivamente

fallace (e quindi manipolatorio), in quanto la determinazione di questo giudizio comporterebbe considerazioni al di là dell'analisi del testo in esame (come l'esame delle comunicazioni previe e di tutti gli argomenti a cui l'hashtag #DecretoSicurezza può rimandare), non si può assolutamente valutarlo come un argomento accettabile.

Nel nostro Esempio 4, possiamo notare come Salvini non descriva fedelmente l'opinione del suo opponente, ma anzi la distorce e la manipoli per poterla poi classificare e attaccare come una "cosa da matti." Tale mossa manipolatoria, chiamata *straw man* in argomentazione (Macagno e Walton, 2017), è tanto frequente nella "retorica" di Salvini da esserne una delle sue caratteristiche più evidenti. La distorsione, come l'attacco, configurano tale fallacia come una tattica estremamente ingannevole e aggressiva, soprattutto in condizioni non dialogiche che quindi non contemplano una riposta della parte attaccata. La fallacia di *straw man* si può riscontrare nel 17% dei messaggi totali, indicando una chiara tendenza dell'autore dei post a rappresentare in modo non fedele i punti di vista che intende criticare. Spesso lo *straw man* è commesso rappresentando erroneamente la finalità pragmatica delle opinioni discordanti (qualificati come "attacchi," "minacce," "fastidio," "colpevolizzazione"), o banalizzandone il contenuto ("sollevare un polverone;" "se a Bruxelles capiscono che non abbiamo messo numeri a caso"), o distorcendolo o eliminando elementi essenziali dell'opinione riportata ("Dove sta scritto che io devo togliere 4 miliardi di euro dalle tasche degli italiani perché un commissario europeo mi ha detto di fare così???"). La tattica di *straw man* più frequente rimane la prima, che più genericamente si può generalizzare come una descrizione dell'opinione contraria tramite parole emotive usate ingiustificatamente. Un esempio è il seguente:

Esempio 5

Non ho parole, un comizio durante la Messa? Manderò a questo prete il testo della #LeggeSalvini perché capisca che è un passo in avanti nel nome del rispetto, delle regole, della sicurezza, della vera integrazione. Amen.

Lo *straw man* in questo caso consiste nella descrizione della critica alle politiche immigratorie di Salvini come un "comizio," presupponendo la finalità politica della omelia in questione e la faziosità politica del prete, fattori che non sono comprovati da evidenze. In questo caso è utile segnalare come le presupposizioni indebite siano la chiave dello *straw man*.

L'ultimo tipo di fallacia commessa che caratterizza in modo significativo il corpus è la generalizzazione indebita, che assieme all'altra fallacia di generalizzazione, il *secundum quid*, occorre nel 15% dei post argomentativi totali. Tale fallacia si riscontra molto frequentemente usata per estendere la categoria "italiani" dalla percentuale che ha sostenuto la Lega nelle elezioni (poco più del 17%) alla totalità dei sostenitori dell'at-

tuale governo o addirittura alla totalità dei cittadini. Un uso differente consiste nell'uso ingannevole dell'esempio, vale a dire il caso singolo generalizzato a una tendenza, come per esempio un caso di violenza commesso da stranieri mostrato come evidenza di un comportamento generalizzato ("e poi dicono siamo stati 'troppo cattivi'"). La generalizzazione indebita può essere anche usata implicitamente, suggerendo una mentalità condivisa da una categoria di persone:

Esempio 6
ASCOLTATE! "Vogliamo il permesso di soggiorno SUBITO, ADESSO e ORA!" "E CASA per tutti!". Capito? Lo PRETENDONO, ce lo ORDINANO. Gli italiani ci hanno mandato al governo per affermare questo principio: chi non ha il diritto di rimanere in Italia se ne torna al suo Paese!

Questo messaggio, oltre a distorcere i contenuti di una manifestazione di protesta (che non ha certamente il fine di "ordinare" nulla), presuppone che gli italiani che hanno votato Lega (e quindi le politiche anti-immigratorie) rappresentino la totalità degli italiani. Inoltre, implicitamente suggerisce che la categoria complessiva di individui "che non ha diritto di rimanere in Italia" – e non solamente coloro che hanno partecipato alla manifestazione (una minoranza di tale gruppo) – condivida le opinioni riportate.

6. Risultati: parole chiave e parole emotive

6.1. *Analisi quantitativa*

In totale nel corpus possiamo notare un'altissima frequenza di parole emotive usate abusivamente (1236, corrispondente a circa 1,5 parola emotiva per *tweet*). La strategia più comune è la definizione persuasiva (780 occorrenze, pari al 63%), seguita dall'epiteto circolare (305 occorrenze, pari al 24%) e dalla quasi definizione (151 occorrenze, pari al 13%). Questi risultati dell'analisi manuale linguistico-argomentativa sono stati incrociati con quelli derivanti dal trattamento automatico del corpus.

L'analisi automatica del corpus mostra come i messaggi di Salvini siano caratterizzati da alcune parole chiave specifiche. Il termine che presenta un valore di *keynessness* più alto, vale a dire al cui frequenza nel corpus nostro risulti più differente da quello di riferimento, è "buonsenso" (valore: 573; 47 occorrenze), seguito da "scafista," "pacchia," "buonista," "perbene," "delinquente," "clandestino," "trafficante," "sbarco" e "rosiconi." Delle 10 parole chiave individuate, 5 sono dispregiative e di queste 2 ("buonista" e "rosiconi") sono epiteti circolari, in quanto implicano un giudizio di valore negativo senza che vi siano condizioni predefinite per il loro uso (cioè non è possibile giustificarne oggettivamente l'uso per classificare una entità). Le parole "scafista,"

"delinquente," "clandestino," e "trafficante" sono usate come peggiorativi per implicare un giudizio di valore sulla pratica dell'immigrazione non controllata. La loro connotazione è spesso modificata tramite la quasi-definizione, vale a dire la collocazione frequente in contesti in cui il giudizio negativo non può essere frainteso ("trafficanti di esseri umani," "mangiatoia dell'immigrazione clandestina," "bastardi scafisti," "protezione delle frontiere <contro scafisti e clandestini>"). L'aggettivo "perbene" è usato per fini dissociativi ("italiani perbene" è il predicato nominale con indice di *keynessess* più alto) per distinguere comportamenti approvati e indicati come positivi da quelli contrari. Chiaramente questo uso è circolare, visto che nessuna ragione per la predicazione del termine chiave è fornita. Il problema ulteriore che emerge è una sorta di minaccia implicita: chi non approva il comportamento riferito all'"italiano perbene" è automaticamente classificato come appartenente all'altro estremo della falsa dicotomia (certamente non troppo apprezzabile).

6.2. *Analisi qualitativa: epiteti circolari*

Da un punto di vista di analisi qualitativa, possiamo notare che la parola chiave più frequente ("buonsenso") è anche quella argomentativamente più potente, considerando il numero di inferenze che essa suggerisce. "Buonsenso" è comunemente definito come la capacità di "valutare e distinguere il logico dall'illogico, l'opportuno dall'inopportuno, e di comportarsi in modo giusto, saggio ed equilibrato, in funzione dei risultati pratici da conseguire"[4]. L'uso di questo termine ha un duplice effetto argomentativo. Da una parte esso giustifica una conclusione in modo circolare (la conclusione è accettabile perché logica, opportuna, etc.). Dall'altra, designando una capacità umana, implica un giudizio sia su chi accetta che su chi rifiuta l'opinione in questione (chi non è d'accordo non è in grado di ragionare in modo logico, equilibrato, etc.). Un chiaro esempio è il seguente:

Esempio 7
Bene! Non servono comunque i sondaggi per capire che la grande maggioranza degli italiani è d'accordo con noi: si tratta di semplice BUONSENSO. Sbaglio? STOP Ong, STOP scafisti, io non cambio idea, non si molla!

Questo esempio illustra una strategia frequentemente usata in modo analogo in molti *tweet*. Un punto di vista controverso (la grande maggioranza degli italiani è a favore delle politiche migratorie di Salvini) è sostenuto da due argomenti. Il primo consiste in una generalizzazione basata sui sondaggi relativi alla Lega (che non dimostrano in ogni caso

[4] https://dizionari.repubblica.it/Italiano/B/buonsenso.html (consultato il 01 febbraio 2019).

una maggioranza, né un appoggio alla politica in questione), che illustra una fallacia sia di generalizzazione indebita che di *secundum quid*. Il secondo corrisponde all'uso del termine chiave, che permette una inferenza estremamente problematica: la politica migratoria è opportuna e saggia (presupposizione); la maggior parte degli italiani sceglie e sostiene ciò che è saggio e opportuno; quindi la maggior parte degli italiani sostiene la politica migratoria. Questo ragionamento è manipolatorio per due ragioni. Da una parte, Salvini presuppone una qualità della politica che sarebbe autocontraddittorio definire come condivisa (visto che si riferisce pure lui a una "maggioranza" e non alla *totalità* degli italiani), manipolando le presupposizioni. Dall'altra, divide le posizioni riguardanti tale politica in due categorie, quella favorevole e basata sul retto e logico pensiero e quella contraria, che quindi non può qualificarsi come espressa da persone di "buonsenso." Gli oppositori si trovano quindi a essere automaticamente accusati di non sapere distinguere il logico dall'illogico e l'opportuno dall'inopportuno, accusa che da sola è sufficiente a escluderli dal dibattito. In quanto incapaci di valutare, non possono essere considerati come interlocutori credibili, né le loro ragioni possono essere viste come degne di considerazione.

6.3. *Analisi qualitativa: definizioni persuasive*

Come abbiamo evidenziato nella descrizione della metodologia di analisi, la definizione persuasiva può essere facilmente distinguibile quando gli indicatori di una ridefinizione (*dissociation markers*) sono presenti, come gli aggettivi "vero" o "reale." Un esempio è il seguente:

Esempio 8
L'Africa non si aspetta carità, ma si aspetta AIUTO VERO per studiare, curarsi, crescere, lavorare, ognuno nella sua terra senza sradicare popoli e culture.

In questo messaggio, il concetto di "aiuto" è cruciale in quanto è utilizzato per sostenere la conclusione che specifiche politiche di accoglienza devono essere rielaborate. Tuttavia, il termine "aiuto" è comunemente usato per riferirsi proprio ai comportamenti che l'ex Ministro dell'Interno intende limitare. In questo caso emerge come strategico l'uso della definizione persuasiva: il termine "aiuto" è ridefinito in modo da assumere come significato "vero" *solamente* l'insieme delle azioni finalizzate a sostenere i popoli a distanza. Il significato comunemente condiviso (opera di soccorso o assistenza fornita a chi si trova in difficoltà o in pericolo), che include necessariamente l'accoglienza di persone che fuggono dal loro paese (e si trovano in difficoltà), è quindi escluso come "falso," e descritto (quasi-definito) tramite altre parole emotive – a loro volta quasi-definite – che giustificano giudizi negativi ("carità," "sradicare").

6.4. *Analisi qualitativa: quasi-definizioni*

La quasi-definizione è il fenomeno più complesso da giustificare oggettivamente, in quando comporta una ricontestualizzazione di un termine emotivo. Tale strategia può essere individuata tramite l'uso di specifici indicatori linguistici come l'uso inappropriato di suffissi (come l'accrescitivo "*-one*," che esprime una qualità della base percepita come negativa, normalmente un eccesso o una grandezza esagerata e inopportuna, tali da sfiorare il ridicolo o da suscitare il disprezzo, cfr. Lo Duca, 2004, 211) o la segnalazione inequivocabile dell'inferenza pretesa dall'autore (per esempio, l'uso di punti esclamativi, *emoticon*, puntini di sospensione). Un esempio è il seguente:

Esempio 9
Secondo Repubblica e qualche professorone di sinistra non dovrei condividere con voi sui Social ciò che faccio, mangio o bevo...!! Io non cambio, stavo con Voi prima di fare il ministro e continuo a farlo oggi! Bacioni e Maalox ai rosiconi.

Il termine "professorone" è usato qui come parola emotiva per giustificare la conclusione che "il consiglio di non condividere sui Social ciò che faccio... è assurda (ridicola, etc.)." L'ex Ministro dell'Interno non spiega le ragioni a sostegno della sua conclusione; piuttosto egli usa un termine che normalmente permette inferenze valutative positive, persino quando suffissato in *-one*. In questo caso, analizzando il corpus di riferimento, il termine è prevalentemente associato a epiteti o concetti positivi come "qualità," "inarrivabile," "onnisciente" o contrastato con il concetto di errore ("persino il professorone...;" "anche i professoroni"). La strategia utilizzata da Salvini in questo esempio è modificare il contesto prototipico associando il termine a presunta faziosità politica ("di sinistra") e segnalando la ridicolosità dei loro comportamenti tramite epiteti connotati negativamente ("rosiconi"). "Professorone" non è ridefinito, ma piuttosto manipolato dal punto di vista delle inferenze che possono essere ricavate dal suo uso (è un professorone, quindi un ridicolo superbo).

La quasi-definizione può essere identificata tramite il contrasto tra l'uso di un termine comunemente considerato neutro e un contesto fortemente negativo. Un esempio è il seguente:

Esempio 10
Col Pd caos e clandestini, con la Lega ordine e rispetto. Certi sindaci rimpiangono i bei tempi andati sull'immigrazione, ma anche per loro è finita la pacchia!

Salvini usa due termini, "immigrazione" e "clandestini" in un contesto fortemente negativo. "Clandestino" è contrastato con "rispetto,"

mentre il termine "immigrazione" è usato per designare pratiche che suppostamente hanno generato la "pacchia" di alcune persone, vale a dire la loro condizione fortunata. Si noti che "pacchia" è usata presupponendo una situazione di sfruttamento di determinate condizioni a scapito dei cittadini, condizione che sicuramente non è giustificata né universalmente condivisa.

La quasi definizione, infine, è spesso associata all'uso figurato di termini. Si consideri nuovamente l'*Esempio 1* analizzato nella sezione 5.2:

> Se per dare un Futuro ai nostri figli, costretti a scappare all'estero, dovrò ignorare uno "zerovirgola" imposto da Bruxelles, per me quel numero vale poco. Prima viene la felicità dei nostri giovani.

La sineddoche "zerovirgola" è usata per indicare la correzione della legge di bilancio. Tuttavia, il *tweet* intende giustificare il giudizio morale negativo di tale correzione (e quindi la decisione di non rispettarla). L'unica ragione che Salvini adduce è l'uso di questo termine in un contesto al tempo stesso nuovo e specifico, in cui "zerovirgola" è contrastato con il "futuro" e la "felicità" dei "nostri" figli, e al contempo associato a termini valutati negativamente come "imposto." I termini "futuro" e "felicità" sono più problematici, in quanto la manovra non si può dire finalizzata alla creazione di futuro e felicità, ma di certe *condizioni* che Salvini descrive come tali senza fornire prove o dettagli. In entrambi i casi, i termini sono definiti persuasivamente.

7. Conclusioni

Twitter è uno strumento estremamente diffuso per condividere informazioni e aggiornamenti di natura personale. Il problema si pone quando un messaggio di 280 caratteri ha come fine difendere una posizione o persino una decisione politica. Il rischio è che gli argomenti diventino senza ragioni nel vero senso del termine, vale a dire si trasformino in conclusioni fondate su premesse non condivise o nei casi più estremi, irrilevanti o non condivisibili. Questa analisi è stata effettuata per esplorare questa ipotesi tramite evidenze ricavate tramite schemi di annotazione. A tale fine, abbiamo sottoposto il corpus di *tweet* argomentativi di Salvini a una triplice analisi argomentativa, condotta tramite l'annotazione dei tipi di argomento e delle fallacie, la valutazione della struttura degli argomenti, e l'identificazione delle strategie ridefinitorie. Nonostante le limitazioni legate alla complessità della metodologia di annotazione (tre differenti codici) e l'estensione del corpus (limitato a un periodo specifico e non comparato con corpora simili), è possibile trarre alcune conclusioni preliminari dai dati emersi.

Nel corpus dei *tweet* argomentativi prodotti dall'ex Ministro dell'Interno italiano, solo il 21% degli argomenti ha le caratteristiche formali

di un argomento completo, in cui un punto vista è fondato su ragioni ed evidenze sufficienti alla sua valutazione. Se si confrontano questi dati con la qualità degli argomenti prodotti da studenti (Erduran *et al*., 2004; Mayweg-Paus e Macagno, 2016; Mayweg-Paus, Macagno e Kuhn, 2016; McNeill e Krajcik, 2011), il quadro è preoccupante. Salvini non solo non fornisce (o non interpreta) le evidenze a sostegno delle sue conclusioni nella grande maggioranza dei casi, ma la metà degli argomenti (in senso ampio) che egli usa si configurano come manipolatori.

Dal punto di vista del tipo di argomenti usati, l'analisi mostra come Salvini tenda a usare insistentemente argomenti finalizzati a difendere un giudizio di valore, sia sulle proprie politiche e sulla propria persona, che sulle altrui opinioni, posizioni e sull'altrui *ethos*. Gli argomenti per conseguenza e per ragionamento pratico sono fondati su argomenti per valori e per classificazione (cioè finalizzati a supportare valutazioni di stati di cose). Gli argomenti per cui sono fornite sufficienti evidenze sono nella maggior parte dei casi argomenti per *commitment*, in cui la relazione tra promesse e future decisioni o giudizi di valori non richiede altri fondamenti che il richiamo a precedenti promesse. La maggior parte delle strategie argomentative è priva delle evidenze necessarie oppure si basa su premesse non condivise o contraddittorie.

Le fallacie sono state determinate e quindi annotate solo sulla base di criteri oggettivi (un'analisi basata su considerazioni di natura giuridica o politica, tendendo in conto la correttezza delle affermazioni rivelerebbe numeri probabilmente ben superiori). Tuttavia, anche usando questi criteri estremamente rigidi e limitativi l'incidenza degli "argomenti" ingannevoli è impressionante. L'alto numero di attacchi personali fallaci mostra una forte preoccupazione per le opinioni contrarie, che vengono sottoposte a critiche ingannevoli e irrilevanti che si estendono a coloro che difendano tali posizioni. Tale sforzo si unisce a uno per frequenza analogo finalizzato a rinforzare il buon carattere dell'ex Ministro dell'Interno. Queste due tendenze hanno come denominatore comune il richiamo a valori condivisi o condivisibili dal pubblico, la cui applicazione agli stati di cose descritti non è tuttavia fondata su ragioni ed evidenze.

Gli *ad hominem* diretti presentano insulti solo quando il giudizio può essere giustificato da una "giusta indignazione" e quindi condiviso dal pubblico. Nei rimanenti casi, l'attacco è condotto in nome di valori come la coerenza, l'integrità, la ragionevolezza, la giustizia, la semplicità contro gli "interessi," i "ricchi," gli "invidiosi," etc. Come tali valori siano istanziati è un problema che è risolto tramite falsi presupposti.

Sono proprio le presupposizioni indebite una caratteristica linguistica saliente del corpus. Salvini tende a manipolare le presupposizioni in diversi modi, non solo tramite l'uso insistito di strategie ingannevoli come lo *straw man* e le false dicotomie, ma anche tramite ridefinizioni ed epiteti circolari. Il linguaggio è manipolato sia a livello di definizioni – e

quindi di significato comune – che a livello di "connotazione" e quindi di inferenze prototipiche. Il risultato è un lessico solo apparentemente condiviso, che è gradualmente costruito artificialmente modificando parole e usi comuni per creare all'interno della comunità di Twitter ragioni preconfezionate e sintetizzate. In questo modo, l'uso di una specifica parola emotiva costituisce da sola una ragione a favore o contro un punto di vista o una persona, senza che siano addotti argomenti o prove ulteriori o perlomeno a sostegno di tale giudizio di valore implicito.

Riferimenti bibliografici

Artsybashev, M. (1915). *Sanine*. New York, NY: Huebsch.

Atlas, J.D. (1991). Topic/comment, presupposition, logical form and focus stress implicatures: The case of focal particles only and also. *Journal of Semantics*, *8*, 1-2, pp. 127-147, https://doi.org/10.1093/jos/8.1-2.127.

Atlas, J.D. (2004). Descriptions, Linguistic Topic/Comment, and Negative Existentials: A Case Study in the Application of Linguistic Theory to Problems in the Philosophy of Language. In M. Reimer e A. Bezuidenhout (a cura di), *Descriptions and Beyond*. Oxford: Oxford University Press, pp. 342-360.

Atlas, J.D. (2008). Presupposition. In L. Horn e G. Ward (a cura di), *The Handbook of Pragmatics*. Oxford, UK: Blackwell Publishing Ltd., pp. 29-52, https://doi.org/10.1002/9780470756959.ch2.

Beaver, D. (2010). Have you noticed that your belly button lint colour is related to the colour of your clothing. In R. Bäuerle, U. Reyle e T. Zimmerman (a cura di), *Presuppositions and Discourse: Essays Offered to Hans Kamp*. Oxford, UK: Elsevier, pp. 65-99.

Ben-Ze'ev, A. (2000). *The subtlety of emotions*. Cambridge, MA: MIT Press.

Bentham, J. (1824). *The book of fallacies*. London, UK: John and H.L. Hunt.

Blumenthal, S. (1982). *The permanent campaign*. Boston, MA: Beacon Press.

Bosc, T., Cabrio, E., Villata, S. (2016). Tweeties Squabbling: Positive and Negative Results in Applying Argument Mining on Social Media. In P. Baroni, T. Gordon, T. Scheffler e M. Stede (a cura di), *COMMA*. Amsterdam: IOS Press, pp. 21-32

Brocca, N., Garassino, D., Masia, V. (2016). Politici nella rete o nella rete dei politici? L'implicito nella comunicazione politica italiana su Twitter. *PhiNBeiheft*, *11*, pp. 66-79.

Brooker, P., Barnett, J., Cribbin, T. (2016). Doing social media analytics. *Big Data e Society*, *3*, 2, pp. 1-12.

Casanova, G. (1911). *Historia della mia fuga dalle prigioni della republica di Venezia dette "li Piombi"*. Milano: Alfieri e Lacroix.

Cepernich, C., Bracciale, R. (2018). Hybrid 2018 campaigning: Italian political leaders and parties social media habits. *Italian Political Science*, *13*, 1, pp. 36-50.

Cobianchi, V., Del Sal, G., Splendore, S. (2014). Nuove forme per le news e (vecchio) giornalismo: i direttori italiani e l'uso di Twitter. *Problemi dell'informazione*, *39*, 2, pp. 100-217.

Conway, B.A., Kenski, K.Wang, D. (2013). Twitter use by presidential primary

candidates during the 2012 campaign. *American Behavioral Scientist*, *57*, 11, pp. 1596-1610, https://doi.org/10.1177/0002764213489014.

Dascal, M. (2003). *Interpretation and understanding*. Amsterdam: John Benjamins Publishing Company.

Dusmanu, M., Cabrio, E., Villata, S. (2017). Argument mining on Twitter: Arguments, facts and sources. In M. Palmer, R. Hwa e S. Riedel (a cura di), *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen: Association for Computational Linguistics, pp. 2317-2322.

Erduran, S., Simon, S., Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education*, *88*, 6, pp. 915-933, https://doi.org/10.1002/sce.20012.

Garza-Cuarón, B. (1991). *Connotation and meaning*. Berlin: Walter de Gruyter.

Gundel, J., Fretheim, T. (2004). Topic and Focus. In L. Horn e G. Ward (a cura di), *The Handbook of Pragmatics*. London: Blackwell, pp. 175-196 https://doi.org/10.1002/9780470756959.ch8.

Halldén, S. (1960). *True love, true humour and true religion: a semantic study*. Lund: Gleerlup.

Hansen, H., Walton, D. (2013). Argument kinds and argument roles in the Ontario provincial election. *Journal of Argumentation in Context*, *2*, 2, pp. 226-258, https://doi.org/10.1075/jaic.2.2.03han.

Hitchcock, D. (1998). Does the traditional treatment of enthymemes rest on a mistake? *Argumentation*, *12*, 1, pp. 15-37, https://doi.org/10.1023/A:1007738519694.

Hitchcock, D. (2003). Toulmin's Warrants. In F. Van Eemeren, A. Blair, C. Willard e F. Snoeck-Henkemans (a cura di), *Anyone Who Has a View: Theoretical Contributions to the Study of Argumentation* (pp. 69-82). Dordrecht: Springer Netherlands, https://doi.org/10.1007/978-94-007-1078-8_6.

Huxley, A. (2010). *Eyeless in Gaza*. London: Random House.

Kecskes, I. (2003). *Situation-bound Utterances in L1 and L2*. Berlin-New York: Mouton de Gruyter.

Kecskes, I. (2008). Dueling contexts: A dynamic model of meaning. *Journal of Pragmatics*, *40*, 3, pp. 385-406, https://doi.org/10.1016/j.pragma.2007.12.004.

Kecskes, I. (2013). *Intercultural pragmatics*. Oxford: Oxford University Press.

Kecskes, I., Zhang, F. (2009). Activating, seeking, and creating common ground: A socio-cognitive approach. *Pragmatics & Cognition*, *17*, 2, pp. 331-355, https://doi.org/10.1075/pc.17.2.06kec.

Kerbrat-Orecchioni, C. (1977). *La connotation*. Lyon: Presses Universitaires de Lyon.

Kilgarriff, A. (2012). Getting to know your corpus. In P. Sojka, P. Aleš, H. Ivan e K. Karel (a cura di), *International conference on text, speech and dialogue*. Heidelberg: Springer, pp. 3-15.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., ... Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, *1*, 1, pp. 7-36, https://doi.org/10.1007/s40607-014-0009-9.

Konstantinidou, A., Macagno, F. (2013). Understanding students' reasoning: argumentation schemes as an interpretation method in science education.

Science e Education, *22*, 5, pp. 1069-1087, https://doi.org/10.1007/s11191-012-9564-3.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, *77*, 3, pp. 319-337, https://doi.org/10.1002/sce.3730770306.

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, *94*, 5, pp. 810-824, https://doi.org/10.1002/sce.20395.

Kwak, H., Lee, C., Park, H., Moon, S. (2010). What is Twitter, a social network or a news media? In M. Rappa, P. Jones, J. Freire e S. Chakrabarti (a cura di), *Proceedings of the 19th international conference on World wide web*. Raleigh, NC: ACM Press, pp. 591-600.

Lee, J., Queal, K. (2016, January 28). Introducing the Upshot's Encyclopedia of Donald Trump's Twitter Insults. *The New York Times*, https://www.nytimes.com/2016/01/29/upshot/introducing-the-upshots-encyclopedia-of-donald-trumps-twitter-insults.html (ultima consultazione 18/07/2019).

Leech, G. (1983). *Principles of pragmatics*. London: Longman.

Lo Duca, M. (2004). Nomi di agente. In M. Grossmann e F. Rainer (a cura di), *La formazione delle parole in italiano*. Tübingen: Niemeyer, pp. 191-217.

López-Meri, A., Marcos-García, S., Casero-Ripollés, A. (2017). ¿Qué hacen los políticos en Twitter? Funciones y estrategias comunicativas en la campaña electoral española de 2016. *El Profesional de La Información*, *26*, 5, pp. 1699-2407.

Macagno, F. (n.d.). How can metaphors communicate arguments?.

Macagno, F. (2008). Dialectical relevance and dialogical context in Walton's pragmatic theory. *Informal Logic*, *28*, 2, pp. 102-128, https://doi.org/10.22329/il.v28i2.542.

Macagno, F. (2013). Strategies of character attack. *Argumentation*, *27*, 4, pp. 369-401, https://doi.org/10.1007/s10503-013-9291-1.

Macagno, F. (2017). The logical and pragmatic structure of arguments from analogy. *Logique et Analyse*, *60*, 240, pp. 465-490, https://doi.org/10.2143/LEA.240.0.3254093.

Macagno, F. (2018a). A dialectical approach to presupposition. *Intercultural Pragmatics*, *15*, 2, pp. 291-313, https://doi.org/10.1515/ip-2018-0008.

Macagno, F. (2018b). Assessing relevance. *Lingua*, *210-211*, pp. 42-64, https://doi.org/10.1016/j.lingua.2018.04.007.

Macagno, F., Bigi, S. (2017). Analyzing the pragmatic structure of dialogues. *Discourse Studies*, *19*, 2, pp. 148-168, https://doi.org/10.1177/1461445617691702.

Macagno, F., Bigi, S. (2019). Analyzing dialogue moves in chronic care communication. Dialogical intentions and customization of recommendations for the assessment of medical deliberation. *Journal of Argumentation in Context*, Advance online publication.

Macagno, F., Konstantinidou, A. (2013). What students' arguments can tell us: Using argumentation schemes in science education. *Argumentation*, *27*, 3, pp. 225-243, https://doi.org/10.1007/s10503-012-9284-5.

Macagno, F., Walton, D. (2011). Reasoning from paradigms and negative evidence. *Pragmatics & Cognition*, *19*, 1, pp. 92-116, https://doi.org/10.1075/pc.19.1.04mac.

Macagno, F., Walton, D. (2014a). Argumentation schemes and topical relations.

In G. Gobber e A. Rocci (a cura di), *Language, reason and education*. Bern: Peter Lang, pp. 185-216.

Macagno, F., Walton, D. (2014b). *Emotive Language in Argumentation*. Cambridge: Cambridge University Press, https://doi.org/10.1017/CBO9781139565776.

Macagno, F., Walton, D. (2015). Classifying the patterns of natural arguments. *Philosophy and Rhetoric*, *48*, 1, pp. 26-53, https://doi.org/10.1353/par.2015.0005.

Macagno, F., Walton, D. (2017). *Interpreting straw man argumentation. The pragmatics of quotation and reporting*. Amsterdam: Springer.

Mayweg-Paus, E., Macagno, F. (2016). How dialogic settings influence evidence use in adolescent students. *Zeitschrift Für Pädagogische Psychologie*, *30*, 2-3, pp. 121-132, https://doi.org/10.1024/1010-0652/a000171.

Mayweg-Paus, E., Macagno, F., Kuhn, D. (2016). Developing argumentation strategies in electronic dialogs: Is modeling effective? *Discourse Processes*, *53*, 4, pp. 280-297, https://doi.org/10.1080/0163853X.2015.1040323.

McNeill, K., Krajcik, J. (2008). Inquiry and scientific explanations: Helping students use evidence and reasoning. In J. Luft, R. Bell e J. Gess-Newsome (a cura di), *Science as inquiry in the secondary setting*. Arlington, VA: National Science Teachers Association Press, pp. 121-134.

McNeill, K., Krajcik, J. (2011). *Supporting Grade 5-8 Students in Constructing Explanations in Science: The Claim, Evidence, and Reasoning Framework for Talk and Writing*. New York: Pearson Allyn e Bacon.

Mel'čuk, I. (2015). *Semantics: From meaning to text* (Vol. 3). Amsterdam-Philadelphia, PA: John Benjamins Publishing Company.

Mel'čuk, I., Iordanskaja, L. (2009). Connotation (in linguistic semantics). In S. Kempgen, P. Kosta, T. Berger e K. Gutschmidt (a cura di), *Die slavischen Sprachen (Ein internationales Handbuch zu ihrer Struktur, ihrer Geschichte und ihrer Erforschung)*. Berlin-New York: Walter de Gruyter, pp. 875-882.

Mendes, A.E. (2016). Digital Demagogue: The Critical Candidacy of Donald J. Trump. *Journal of Contemporary Rhetoric*, *6*, 3-4, pp. 62-73.

Ott, B.L. (2017). The age of Twitter: Donald J. Trump and the politics of debasement. *Critical Studies in Media Communication*, *34*, 1, pp. 59-68, https://doi.org/10.1080/15295036.2016.1266686.

Rigotti, E. (2007). Relevance of context-bound loci to topical potential in the argumentation stage. *Argumentation*, *20*, 4, pp. 519-540, https://doi.org/10.1007/s10503-007-9034-2.

Rigotti, E., Greco-Morasso, S. (2019). *Inference in argumentation: A topics-based approach to argument schemes*. Amsterdam: Springer.

Rubinelli, S. (2009). *Ars Topica: The classical technique of constructing arguments from Aristotle to Cicero* (Vol. 15). Amsterdam: Springer.

Scanfeld, D., Scanfeld, V., Larson, E.L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control*, *38*, 3, pp. 182-188, https://doi.org/10.1016/j.ajic.2009.11.004.

Scarfone, G. (2017). Giornalismo e social network: un'analisi linguistica. *Lingue e Culture Dei Media*, *1*, 1, pp. 44-89.

Schank, R., Collins, G., Davis, E., Johnson, P., Lytinen, S. Reiser, B. (1982). What's the point? *Cognitive Science*, *6*, 3, pp. 255-275, https://doi.org/10.1207/s15516709cog0603_2.

Schiappa, E. (2003). *Defining reality. Definitions and the politics of meaning*. Carbondale and Edwardsville, IL: Southern Illinois University Press.
Stalnaker, R. (1998). On the representation of context. *Journal of Logic, Language and Information*, *7*, 1, pp. 3-19, https://doi.org/10.1023/A:1008254815298.
Stevenson, C. (1937). The emotive meaning of ethical terms. *Mind*, *XLVI*, *181*, pp. 14-31, https://doi.org/10.1093/mind/XLVI.181.14.
Stevenson, C. (1938). Persuasive definitions. *Mind*, *47*, pp. 331-350.
Stevenson, C. (1944). *Ethics and Language*. New Haven: Yale University Press.
Strawson, P. (1971). Identifying Reference and Truth-Values. In P. Strawson (a cura di), *Logico-Linguistic Papers*. London: Methuen, pp. 75-95.
Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
van Eemeren, F., Grootendorst, R. (1984). *Speech acts in argumentative discussions: A theoretical model for the analysis of discussions directed towards solving conflicts of opinion*. Dordrech: Floris Publications.
Van Rees, A. (2008). *Dissociation in argumentative discussions: A pragma-dialectical perspective*. Amsterdam: Springer.
Walton, D. (1990). What is reasoning? What is an argument? *Journal of Philosophy*, *87*, pp. 399-419, https://doi.org/10.2307/2026735.
Walton, D. (1995a). *A pragmatic theory of fallacy*. Tuscaloosa, AL: University of Alabama Press.
Walton, D. (1995b). *Argumentation Schemes for Presumptive Reasoning*. Mahwah, NJ: Routledge, https://doi.org/10.4324/9780203811160.
Walton, D. (1998). *Ad Hominem Arguments*. Tuscaloosa, AL: University of Alabama Press.
Walton, D. (2001). Abductive, presumptive and plausible arguments. *Informal Logic*, *21*, 2, 141-169, https://doi.org/10.22329/il.v21i2.2241.
Walton, D. (2004a). Classification of fallacies of relevance. *Informal Logic*, *24*, 1, pp. 183-185, https://doi.org/10.22329/il.v24i1.2133.
Walton, D. (2004b). *Relevance in argumentation*. Amsterdam-Philadelphia, PA: Routledge.
Walton, D. (2008). *Informal logic: a pragmatic approach*. New York: Cambridge University Press.
Walton, D. (2010). Why fallacies appear to be better arguments than they are. *Informal Logic*, *30*, 2, pp. 159-184, https://doi.org/10.22329/il.v30i2.2868.
Walton, D., Macagno, F. (2016). Profiles of dialogue for relevance. *Informal Logic*, *36*, 4, pp. 523-556, https://doi.org/10.22329/il.v36i4.4586.
Walton, D., Reed, C., Macagno, F. (2008). *Argumentation schemes*. New York: Cambridge University Press, https://doi.org/10.1017/CBO9780511802034.

Analyzing the argumentation on social media. The tweets of Salvini

Twitter is an instrument used not only for sharing public or personal information, but also for persuading the audience. While specific platforms and software have been developed for analyzing macro-analytical data, and specific studies have focused on the linguistic dimension of the tweets, the argumentative dimension of the latter is unexplored to this date. This paper intends to propose a method grounded on the tools advanced in argumentation theory for capturing, coding,

and assessing the different argumentative dimensions of the messages posted on Twitter, focusing on the types of argument communicated, the quality of their premises, and the fallacies committed – including the use of unshared presuppositions and emotive words. This method is applied to a corpus of 843 tweets published by the Italian Minister of the Interior, Mr. Matteo Salvini, from the date of his appointment to the beginning of his campaign for the European elections. The quantitative data provide general indications for detecting the strategies that characterize the argumentative profile of Salvini, which are then analyzed qualitatively.

Keywords: argumentation, emotive words, Twitter, Salvini, fallacies, presupposition.

L'autore ringrazia la Fundação para a Ciência e a Tecnologia per i finanziamenti SFRH/BPD/115073/2016, PTDC/FER-FIL/28278/2017, e PTDC/MHC-FIL/0521/2014 che hanno reso possibile questa ricerca.

Fabrizio Macagno, ArgLab - Instituto de Filosofia da Nova (IFILNOVA), Universidade Nova de Lisboa, Avenida de Berna 26, 1069-061 Lisbon, Portugal, fabrizio.macagno@fcsh.unl.pt