

Tabelline
Quelli che
finiscono in clinica
ragionando
sugli insiemi infiniti

PIERGIORGIO ODIFREDDI

Nell'aprile del 1963, cinquant'anni fa, Paul Cohen annunciò di aver risolto quello che era considerato il più importante problema aperto della matematica: anche più importante del più famoso "ultimo teorema di Fermat", che in fondo riguardava soltanto una curiosità numerica, e traeva il suo interesse principalmente dal fatto di essere facile da enunciare, ma perveracamente difficile da dimostrare. Il problema risolto da Cohen era invece il primo nella lista che il grande matematico David Hilbert aveva stilato nel 1900, come sfida per il nuovo

secolo che stava iniziando. Divenne infatti noto come il "primo problema di Hilbert", e chiedeva se esistevano insiemi infiniti che fossero più grandi dell'insieme dei numeri naturali (come 1 o 2013), e più piccoli dell'insieme dei numeri reali (come la radice di 2 o pi greco). Naturalmente, le espressioni "più grande" e "più piccolo", riferite agli insiemi infiniti, dovevano essere prese con le pinze: senza una loro definizione precisa, si sarebbe soltanto parlato a vuoto. Ma questo era stato fatto, negli ultimi tre decenni dell'Ottocento, da Georg Cantor, che le aveva definite per bene, dimostrando ad esempio

che l'insieme dei numeri reali era appunto "più grande" di quello dei numeri interi. Già lui si era posto il problema poi richiamato da Hilbert, ma non era riuscito a risolverlo, nonostante ci si fosse impegnato talmente, da finire in clinica per l'esaurimento nervoso. Cohen evitò la clinica, e vinse invece la medaglia Fields: l'unico logico ad aver raggiunto il traguardo, finora. Quanto alla sua soluzione, fu che il problema non ha soluzione! E questa è appunto una soluzione, come anche la politica dovrebbe saper riconoscere.

© RIPRODUZIONE RISERVATA

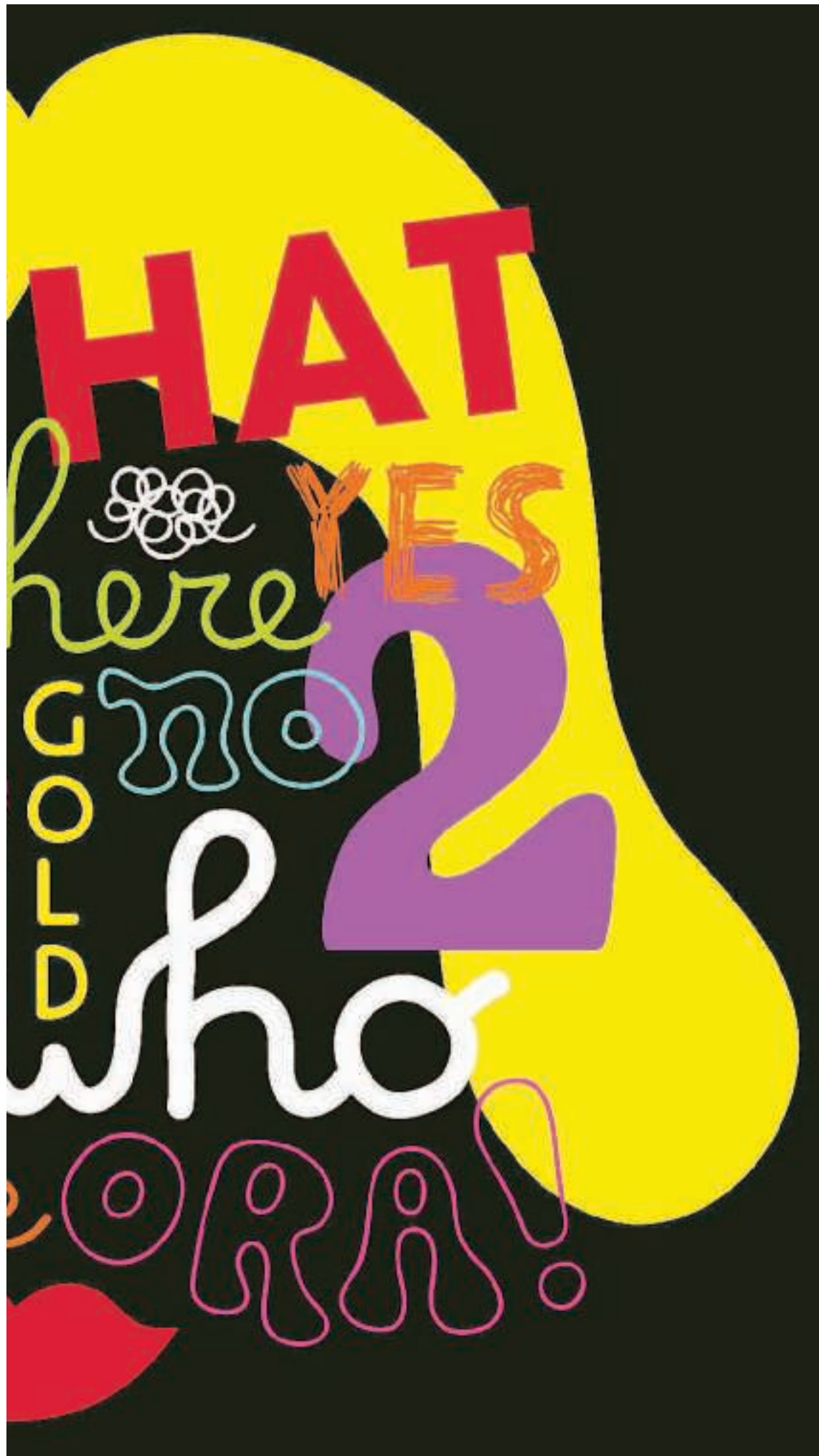


ILLUSTRAZIONE DI OLIMPIA ZAGNOLI

che non siamo in grado di spiegare. Non è lontano il momento in cui il nostro medico ci convocherà per dirci che, visto che i mandorli in città fioriscono in ritardo, noi stiamo per avere un'infezione al rene, anche se lui non sa perché.

Un'altra vittima della rivoluzione di Big Data sono le tecniche di campionamento e di sondaggio. Perché affannarsi a individuare e interrogare un campione di clienti del supermercato, quando, grazie alle telecamere di sorveglianza, si è in grado di esaminare e verificare il comportamento dell'intero universo dei clienti? Poco importa sapere perché, dopo aver comprato i peperoni, il grosso dei clienti cerca un detersivo. Tanto vale spostare lo scaffale dei detersivi non lontano dai peperoni. Con

queste tecniche, dice McKinsey, grandi catene al dettaglio sono arrivate a risparmiare fino al 17 per cento sullo stoccaggio dei prodotti.

Ci sono modi più creativi di fare soldi con Big Data, anche se per ora largamente confinati all'analisi - con i problemi etici e di privacy che comporta - del comportamento e della psicologia dei consumatori. Soprattutto, dicono Mayer-Schoenberger e Cukier, per chi sarà capace di riutilizzare, in modo nuovo, dati raccolti da altri, per scopi diversi. Per ora, però, il potere dei dati è nelle mani di chi li ha. E qualcuno comincia anche a chiedersi se non sia in grado di dare un valore, anche in bilancio, a questi dati. Quanto valgono, insomma, i dati? Si può fare il caso di

Facebook. Al momento della quotazione era stata valutata 104 miliardi di dollari. Un anno prima, l'ultima stima contabile parlava però di soli 6,3 miliardi, il corrispettivo degli investimenti in macchinari, server, computer eccetera. Insomma, gli attivi materiali. Il resto, quasi 100 miliardi di dollari non contabilizzati, possiamo dire che erano dati, il business effettivo di Facebook. Cento miliardi, a fronte di un po' più di due miliardi di pezzi di "contenuto monetizzabile", messi insieme fra il 2009 e il 2011 nelle sue bacheche. Sono i like, i commenti, i post. Vale a dire che, nel mondo di Big Data, ogni *I like* che io clicco vale 5 centesimi di dollaro. Per Mark Zuckerberg, naturalmente.

© RIPRODUZIONE RISERVATA

L'intervento

I numeri non sono tutto contano valori e sentimenti

L'analisi computerizzata delle cifre valuta la quantità, non la qualità. Ma soprattutto nel campo sociale le decisioni dipendono da altri fattori

DAVID BROOKS

Non molto tempo fa mi trovavo a cena con l'amministratore delegato di una grande banca. Aveva appena dovuto decidere se, considerata la debolezza dell'economia e la minaccia di una possibile crisi dell'euro, non fosse meglio abbandonare l'Italia. I suoi economisti avevano prospettato una serie di scenari negativi e calcolato le conseguenze che avrebbero comportato per la sua società. Alla fine, però, il Ceo ha preso la sua decisione sulla base di alcuni valori. La banca che dirige è in Italia da decenni, ed egli non voleva che gli italiani la considerassero opportunistica. Non voleva che le persone che lavorano all'interno della società pensassero che, in tempi difficili, la banca avesse abbandonato il campo. Nel prendere la sua decisione, egli non era ignaro dei dati, ma alla fine ciò che lo ha guidato è stata una visione differente. E, naturalmente, ha fatto bene. Il commercio dipende dalla fiducia. La fiducia è reciprocità rivestita di emozioni. Gli uomini e le aziende che si comportano correttamente nei momenti difficili conquistano un affetto e un'autostima che sono estremamente preziosi, anche se è difficile rilevarlo nei dati.

Racconto questa storia perché ci fa capire i punti di forza e i limiti delle analisi basate sui dati. La grande novità di questo momento storico è che oggi le

nostre vite sono mediate dalla raccolta dati dei computer. In questo mondo, i dati possono essere utilizzati per venire a capo di situazioni sorprendentemente complesse. I dati possono compensare l'eccessiva fiducia nelle nostre intuizioni e ridurre la misura in cui ciò che desideriamo falsifica le nostre percezioni. Ma ci sono molte cose che i dati non sanno fare. Vediamone rapidamente alcune.

I dati faticano a convivere con il sociale. Il nostro cervello è scarso quando ha a che fare con la matematica (svelti, qual è la radice quadrata di 437?), ma se la cava egregiamente nel campo della cognizione sociale. La gente è bravissima a riconoscere i reciproci stati emotivi, a individuare un comportamento non collaborativo e nell'attribuire valore alle cose attraverso le emozioni. L'analisi computerizzata dei dati è eccellente nel misurare la quantità delle interazioni sociali, ma non nel valutare la qualità. Gli scienziati informatici possono rilevare le nostre interazioni con i sei collaboratori con cui trascorriamo la giornata, ma non sono in grado di cogliere il nostro attaccamento agli amici di infanzia che vediamo due volte all'anno, per non parlare dell'amore di Dante per Beatrice, da lui incontrata due sole volte. Perciò, quando prendiamo decisioni che riguardano le relazioni sociali, è sciocco scambiare la sorprendente macchina che abbiamo dentro il cranio per la nuda macchina che

sta sulla nostra scrivania.

I dati faticano a convivere con il contesto. Le decisioni umane non sono eventi isolati. Esse sono inserite in sequenze e contesti. Il cervello umano si è evoluto tenendo conto di questa realtà. Siamo molto bravi a raccontare storie in cui si intrecciano molteplici cause e molteplici contesti. L'analisi dei dati non brilla quanto a pensiero narrativo ed emergente e non è in grado di eguagliare nemmeno la duttilità esplicativa di un romanzo mediocre.

I dati creano pagliai più grandi. Questo è un aspetto che Nassim Taleb, autore di *Antifragile*, ha già sottolineato. Via via che acquisiamo un numero sempre maggiore di dati, possiamo individuare correlazioni statisticamente significative sempre più numerose. La maggior parte di tali correlazioni sono false e, quando tentiamo di comprendere una situazione, esse ci traggono in inganno. Più raccogliamo dati, più l'inattendibilità aumenta in modo esponenziale. Il pagliaio diventa sempre più grande ma l'ago che stiamo cercando è sepolto sempre più profondamente. Una caratteristica dell'era dei grandi dati è il numero di conclusioni «significative» che, come direbbe Nate Silver, non riproducono l'espansione del rapporto rumore/segnale.

I grandi dati hanno difficoltà con i grandi problemi. Se staccando di capire quali email producono maggiori contributi a una campagna potete fare l'esperimento di un controllo casuale. Ma immaginiamo che, invece, stiate tentando di stimolare l'economia durante una recessione. Non avete una società alternativa da usare come gruppo di controllo. Si è discusso molto, ad esempio, su quali siano i migliori stimoli economici e si sono utilizzati montagne di dati ma, per quanto ne sappia, nessuno dei principali protagonisti di questo dibattito è stato persuaso dai dati a cambiare posizione.

I dati fanno dimenticare i valori. Di recente ho visto un libro accademico dall'ottimo titolo *"Nudi dati" è un ossimoro*. Uno degli argomenti sostenuti era che i dati non sono mai "nudi"; sono sempre strutturati secondo le predisposizioni e i valori di qualcuno. Il risultato finale sembra neutrale ma, in realtà, le scelte di valore attraversano tutto il processo, dalla formulazione all'interpretazione.

Tutto ciò non ha lo scopo di affermare che i grandi dati non rappresentano un grande strumento. È solo che, come tutti gli strumenti, anche i dati sono adatti ad alcune cose e inadatti ad altre. Come ha spiegato il professore di Yale Edward Tufte, «il mondo è molto più interessante di qualunque disciplina».

Traduzione Antonella Cesarini

© RIPRODUZIONE RISERVATA

