

Le inclinazioni umane non sono *bias*

di Francesco Varanini

Articolo apparso su Agenda Digitale il 24 giugno 2022.

Bias Is Coded Into Our Technology: lo si sente dire sempre più frequentemente.

Le intelligenze artificiali e gli algoritmi applicati all'educazione, alla medicina, al giornalismo, e ad ogni altro servizio offerto per via digitale ai cittadini sono condizionati da *bias*: caso di frequente citato è il mancato rispetto dei diritti dei cittadini di colore.

Le tecnologie riproducono il modello di azione ed usano le conoscenze di insegnanti, medici, giornalisti. Di conseguenza, intelligenze artificiali e gli algoritmi replicano i *bias* degli umani che li creano, e degli umani con le cui conoscenze sono alimentati. Qualcuno cerca di mitigare il problema con l'*algorithmic fairness*: si può tentare di insegnare ad un algoritmo ad essere equo e giusto, o progettare un algoritmo che corregga altri algoritmi. Ma anche in questo caso la soluzione sarà condizionata dai bias del progettista. Se poi magari si immagina un algoritmo capace di migliorare autonomamente la propria qualità etica, il problema resta presente: dietro la sbandierata qualità etica può nascondersi qualche *bias*.

Errori nella codifica digitale, errori umani

Dunque, appare abbastanza vano limitarsi a stanare e correggere i *bias* dei programmi digitali. I *bias* ai quali vale la pena di guardare sono i bias studiati dalle scienze cognitive: *bias* che condizionano l'agire umano.

Ci troviamo allora a prendere in considerazione le *List of Cognitive Biases*. Liste prodotte non da imbonitori di folle, ma da eminenti scienziati. Leggiamo per esempio: "An emotional bias is a distortion in cognition". Gli affetti, si sostiene, sono pericolosi, perché le connotazioni affettive alterano il processo razionale di presa di decisioni. Ma la nostra vita si riduce ad una sequenza di decisioni razionali? In cosa consiste il *bias*?

Propongo di intendere in prima battuta l'inglese *bias* come *severe and systematic errors*. Quante volte sentiamo dire: questo specifico comportamento della tal persona è causato da *bias*, *pregiudizi*; o più in generale sentiamo dire, come cosa assodata, che esistono umani modi di essere che possono essere chiaramente, definitivamente, considerati frutto di *bias*, errori. E siamo invitati a buttar via, come pericolosi, fuorvianti errori, ogni manifestazione di amore, i nostri sogni, i desideri.

Qualcuno pretende, in nome della scienza, di giudicare il mio o il tuo modo di pensare, e si arroga quindi il diritto di dire: sei in errore, stai sbagliando. Dobbiamo chiederci come siamo arrivati a questa situazione.

Provo un gran dispiacere -imbarazzo, turbamento, delusione, senso di pericolo- quando mi trovo ad ascoltare amici scienziati, tecnici, 'esperti' usare questa bacchetta magica: dire *bias cognitivo* per collocarsi sopra, fuori, e rifiutare il pensiero, la posizione altrui, con il motivo che essa è frutto di un modo sbagliato di pensare. L'inquietudine è per me ancora più forte quando a chiamare in causa i *biases cognitivi* sono informatici o computer scientist. Come gli altri scienziati, tecnici ed 'esperti', parlano come se esistessero tavole che descrivono universalmente il comportamento umano 'esatto' ed il comportamento deviante, il comportamento sano e il comportamento malato; ed anche: il pensiero sano e il pensiero malato. L'inquietudine nel sentir parlare in questa maniera il computer scientist è più forte, perché è subito evidente che, mosso dalla propria competenza professionale, sta applicando agli umani le categorie di giudizio buone per giudicare il buon funzionamento di un certo tipo di macchina, detta computer.

Ma esiste, a motivare la mia inquietudine, un'altra più problematica questione: il computer scientist non è un mero osservatore. Scrivendo il codice, è lui stesso ad istruire la macchina, a dirle: queste sono le tavole della legge. E' il computer scientist a stabilire il modo in cui il computer si

comporterà: lui scrive programmi ed algoritmi. E' sempre lui -se accettiamo la possibilità che il computer in qualche modo pensi- a stabilire come il computer dovrà ragionare. Quindi, in fondo, è lui a decidere quale comportamento e quale pensiero dovranno essere considerati viziati da *bias*. E soprattutto: dispone degli strumenti e dell'autorità per correggere gli errori di programmazione. Il computer scientist è quindi, fino in fondo, l'esempio al quale si attengono tutti gli scienziati, i tecnici e gli 'esperti' che si pongono come obiettivo il *debiasing* degli esseri umani.

E' facile infatti ricordare che il concetto di *bias cognitivo* è frutto di studi e ricerche di altre discipline: psicologia, scienze cognitive, neuroscienze. Basta però ripercorrere brevemente la storia per ricordare che tutte queste discipline si fondano sull'attribuire agli umani il modello di funzionamento proprio dei computer. Il concetto di *bias* riguarda il origine il funzionamento della macchina. E' applicato anche agli esseri umani perché gli esseri umani sono considerati fondamentalmente macchine.

La presenza sulla scena del computer legittima così una nuova forma di controllo sociale.

Preferire il computer a sé stessi

La figura di Alan Turing è circonfusa di gloria. E' il fondatore, oggetto di celebrazioni e apologie. Ma il culto occulta la storia. Seguaci ed eredi non hanno motivo di andare alle radici: le basi della disciplina non devono essere messe in discussione.

Seguaci ed eredi, del resto, sono logici formali, ingegneri, matematici, cognitivistici: l'approccio umanistico e la psicologia del profondo sono per loro approcci irrilevanti.

Alan Turing, però, era un bambino, un adolescente, un giovane adulto solo e incompreso. Si sentiva vittima di pregiudizi. Suo padre considerava inaccettabile, per il suo status di funzionario imperiale, tenere con sé in India il bambino, che crebbe quindi in Inghilterra presso tutori. Vedeva i genitori solo in momenti di vacanza. Né la madre, né il fratello, di pochi anni maggiore, vollero accettare la sua omosessualità - forse anzi addirittura scelsero di non vederla.

Il giovane Alan si appassiona alla matematica, alla crittografia. Pensa: non trovo affetti, comprensione, rispetto negli umani. Io stesso fatico a provare autostima. Preferisco agli artefatti linguaggi umani il puro linguaggio della matematica. Preferisco stare in relazione con macchine. Preferisco immaginarmi come macchina tra macchine. Preferisco considerare i pregiudizi di cui sono vittima come errori sistematici, errori di una macchina.

Poco più che ventenne, nel 1936, descrive nell'articolo *On computable numbers*¹ la *computing machine*. Macchina che *computa*. Turing impone la propria definizione: *computazione*. Spiega che la sua macchina non è in grado di eseguire ogni calcolo. Ma che è in grado di eseguire senza errori quel particolare tipo di calcolo che sono le computazioni. La computazione si fonda sull'uso di simboli dal significato inequivocabile, e sull'esistenza di un programma. La computing machine esegue il programma senza cadere in errore.

In ultima analisi, dunque, sono computabili i problemi risolvibili con la computing machine di Turing.

La vita è sostituita dalla computazione. L'agire è inteso come esecuzione di ciò che sta scritto in un Libro delle Regole. La Computer Science è fondata.

Reti neurali

Spinti anche dalle esigenze immediate della guerra, la *computer machine* immaginata da Turing, viene effettivamente costruita. Funziona.

Nel 1943 esce l'articolo *A logical calculus of the ideas immanent in nervous activity*². Ne sono

1 A.M. [Alan Mathison] Turing, "On computable numbers, with an application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society*, 1936-1937, series 2, vol. 42, pp. 230-265 (Received 28 May, 1936. Read 12 November, 1936). <http://www.inf.unibz.it/~nutt/Teaching/FDBs1617/FDBsPapers/turing-36.pdf>

2 Warren S. McCulloch, Walter Pitts, "A logical calculus of the ideas immanent in nervous activity", *The Bulletin of Mathematical Biophysics*, December 1943, Volume 5, Issue 4, pp 115-133. <https://www.cs.cmu.edu/~epxing/Class/10715/reading/McCulloch.and.Pitts.pdf>

autori Warren McCulloch, maturo psichiatra e filosofo, e Walter Pitts, giovanissimo, geniale matematico.

L'articolo descrive il funzionamento della mente, basandosi sul presupposto che al posto dei neuroni del cervello umano stiano circuiti elettronici. Nasce qui il concetto di rete neurale. E al contempo si apre la strada alle scienze cognitive.

L'ipotesi si situa perfettamente nella scia di Turing: l'attività mentale è vista capacità di computare: manipolare simboli numerici sulla base di un sistema di regole. La via di Cartesio -mente senza corpo- è così percorsa fino alle estreme conseguenze.

Paradossalmente, né McCulloch, né Pitts seguono in realtà la via di Cartesio e di Turing. Per loro l'articolo del 1943 è un esercizio di logica formale. Ma sono lunghi dal considerare possibile descrivere in toto la vita tramite la logica. McCulloch non sopravvaluta il cervello e non rimuove il corpo: resta filosofo e medico; anche poeta; intitola il libro che raccoglie i suoi scritti *Embodiments of Mind*³: si rifiuta di concepire una mente senza corpo. Pitts è un matematico creativo: di fronte a problemi diversi, inventa di volta in volta sistemi di notazione diversi. Giovane infelice, come Turing, non cerca salvezza nel paradiso della computazione.

Eppure l'articolo di McCulloch e Pitts diventa canonico. Punto di riferimento ineludibile per i cultori delle nuove discipline legate alla computazione, perché completa e porta alle estreme conseguenze la tesi esposta da Turing: computazione al posto del pensiero umano; mente ridotta a cervello; macchina digitale come sostituto dell'umano. Si apre così la strada non solo alla Computer Science, ma anche ad altre scienze fondate su una precisa ipotesi di lavoro: studiare l'umano attraverso lo studio del computer.

Io spero

Nel 1950 Turing completa il proprio disegno con un secondo articolo: *Computing Machinery and Intelligence*. Non a caso esce sulla rivista *Mind*.⁴ Si chiede esplicitamente: possono le macchine pensare? Esamina le opinioni contrarie per arrivare ad affermare che sì, le macchine possono pensare. Il legame con il suo primo articolo è evidente: Turing invita a preferire le macchine a sé stessi. Nella conclusione esplicita la motivazione personale della sua ricerca, della sua speranza. Scrive proprio: *We may hope, I hope*, spero che le macchine pensino al posto degli umani. Se gli umani hanno pregiudizi, possiamo sperare che le macchine digitali, i computer, non ne abbiano. L'articolo è un testamento. Nel 1954 Turing, disperato, vilipeso, tradito, deluso dagli umani, muore suicida. La macchina è l'immagine simbolica di noi stessi. Le macchine siamo noi stessi come potremmo essere.⁵

Artificial Intelligence e Cognitive Science

Due anni dopo il percorso è compiuto: nel 1956, in incontri tenuti a Dartmouth e a Cambridge, appaiono, strettamente congiunte tra di loro, due nuovi concetti: *Artificial Intelligence* e *Cognitive Science*.

Lo studio dell'Intelligenza Artificiale "procede sulla base della congettura che ogni aspetto dell'apprendimento o di qualsiasi altra caratteristica dell'intelligenza possa essere in linea di principio descritto in modo così preciso da poter essere simulato da una macchina".⁶

Le *scienze cognitive* sono "il riconoscimento di un insieme fondamentale di preoccupazioni comuni alle discipline della psicologia, dell'informatica, della linguistica, dell'economia, dell'epistemologia

3 Warren McCulloch, *Embodiments of Mind*, The MIT Press, Cambridge, Ma., 1965.
<https://mitpress.mit.edu/books/embodiments-mind>

4 Alan M. Turing, "Computing Machinery and Intelligence", *Mind*, October 1950, vol. 59, n. 236, pp. 433-460.
<https://phil415.pbworks.com/f/TuringComputing.pdf>

5 Francesco Varanini, *Le Cinque Leggi Bronzee dell'Era Artificiale*, Guerini e Associati, 2020. Terza Legge.

6 John McCarthy, Marvin L. Minsky, Nathaniel Rochester, Claude E. Shannon, *Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, August 31, 1955.
<https://www.aaai.org/ojs/index.php/aimagazine/article/view/1904/1802>

e delle scienze sociali in generale. Tutte queste discipline si occupano di sistemi di elaborazione". "I sistemi possono essere descritti come 'artificiali', perché quando l'ambiente cambia, ci si può aspettare che cambino anche loro".⁷

"L'intuizione centrale del cognitivismo è che l'intelligenza - compresa quella umana - assomiglia così tanto alla computazione nelle sue caratteristiche essenziali che la cognizione può effettivamente essere definita come computazione di rappresentazioni simboliche".⁸

A questo punto l'esperienza, la conoscenza, l'apprendimento sono sottratte agli esseri umani, separate dalla storia. Ciò che i nuovi scienziati vedono è nient'altro che il funzionamento di una macchina.

Le scienze cognitive separano la mente dal corpo. Considerano il corpo umano una propaggine indegna di particolare attenzione.

Oltre Kant

Scienziati cognitivi e Computer Scientist, seguendo Kant, si sporgono su terreni mai percorsi da Kant stesso. Dobbiamo a Kant la nozione di filosofia come tribunale della Ragione Pura. Oggi è l'accoppiata Scienze Cognitive-Computer Science ad essere tribunale della ragione. Un tribunale che definisce la conoscenza come informazione elaborata da una macchina.

Kant ci offre un'altra immagine: la mente come gran specchio. La mente, concetto inafferrabile, è ora comodamente sostituita dal computer. Il computer è specchio tramite il quale oggi dovremmo conoscere noi stessi.

Kant intendeva fornire una teoria generale della rappresentazione: descrivere il mondo attraverso simboli. Questo è il compito affidato oggi alla computazione. Il pensiero rappresentativo porta con sé la figura dello spettatore. Il Computer Scientist e lo scienziato cognitivistico si considerano spettatori. Osservano il mondo in base a modelli che non possono e non vogliono mettere in discussione. Lo spettatore è sempre innocente, mai individualmente responsabile delle immagini del mondo offerta ai cittadini: e dell'immagine di sé stesso offerta ad ogni cittadino.

Infatti oggi i Computer Scientist dicono ai cittadini: specchiatevi nel vostro Gemello Digitale. Siete le persone che appaiono attraverso i dati che vi descrivono. Gli psicologi cognitivi aggiungono e precisano: sei una macchina, come ogni macchina sei viziato da questo e da quest'altro *bias*.

La novità digitale non è così nuova

Esistono una scienza ed una filosofia che pretendono di descrivere la stessa vita biologica alla luce del funzionamento del computer. Basta ricordare Richard Dawkins⁹ e Daniel Dennet¹⁰. Esistono studiosi che pretendono di descrivere la capacità umana di fare esperienza, di conoscere e di narrare prendendo a modello il modo in cui il computer tratta le informazioni.

Esistono ovviamente anche voci che, già a partire dagli Anni Settanta del Secolo scorso, criticano questi approcci. Voci di filosofi, già Rorty¹¹, Varela¹², Searle¹³. Esistono voci di Computer Scientist

7 Herbert A. Simon, "Cognitive science: The newest science of the artificial", *Cognitive Science*, Volume 4, Issue 1, 1980, pp. 33-46. <https://www.sciencedirect.com/science/article/abs/pii/S0364021381800031>

8 Francisco Varela, Evan Thompson, Eleanor Rosch, *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press., Cambridge, Ma., 1991. https://monoskop.org/images/2/21/Varela_Thompson_Rosch_The_Embodied_Mind_Cognitive_Science_and_Human_Experience_1991.pdf

9 Richard Dawkins, *The Blind Watchmaker*, W.W. Norton & Company, New York 1986,

10 Daniel C. Dennett, *Consciousness Explained*, Little, Brown and Co., New York 1991.

11 Richard Rorty, *Philosophy and the Mirror of Nature*, Princeton University Press, Princeton, NY, <http://ronliskey.com/docs/rorty-mirroir-of-nature.pdf>

12 Francisco Varela, Evan Thompson, Eleanor Rosch, *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press., Cambridge, Ma., 1991. https://monoskop.org/images/2/21/Varela_Thompson_Rosch_The_Embodied_Mind_Cognitive_Science_and_Human_Experience_1991.pdf

13 John R. Searle, "What Your Computer Can't Know", *The New York Review of Books*, October 9, 2014.

che già attorno al 2000 -basta citare Bill Joy¹⁴, Jaron Lanier¹⁵- mettevano in guardia contro supina accettazione del computer come metafora che tutto spiega.

Oltretutto, bisogna ricordare che -nonostante la grande enfasi posta sulla Rivoluzione Digitale- negli ultimi vent'anni su questa scena non è successo nulla di veramente significativo da un punto di vista tecnologico. Le tecnologie, i progetti, le ipotesi di lavoro non sono cambiati granché negli ultimi vent'anni. Ciò che è cambiato si riduce in fondo a pochi punti.

Basta citarne tre. L'enorme aumento della capienza dei supporti destinati a conservare dati.

L'enorme aumento della potenza di calcolo: Machine Learning e Deep Learning rendono praticamente possibile un trattamento massivo dei dati che prima poteva solo essere immaginato. La pervasiva presenza di reti e piattaforme digitali, attraverso le quali transita ormai in ogni suo aspetto la vita psicologica, sociale, economica della comunità umana.

Il terzo punto mostra come la tecnologia si trasforma in politica. Sembra impossibile oggi, purtroppo, per ogni cittadino del pianeta, vivere senza passare attraverso l'uso di *device* digitali.

E' qui che possiamo osservare il salto di qualità delle scienze cognitive: appare evidente negli ultimi venti anni il nuovo pressante ruolo della psicologia cognitiva.

Tversky e Kahneman: un'articolo di cinquant'anni fa

In realtà, possiamo osservare come la storia delle scienze cognitive discende pienamente dalla teoria della computazione di Turing. Ne è anzi il necessario complemento. Turing voleva che la macchina insegnasse agli umani. Voleva mostrare agli, tramite il funzionamento della macchina rispettosa delle regole logico-formali, le regole del corretto comportamento. Sperava in una regola chiara che dicesse agli umani: qui stai sbagliando. Ma non solo: sperava che macchina mostrasse come correggere gli errori: una correzione automatica, sistemica, esogena, funzionante senza che l'essere umano dovesse impegnarsi personalmente a conoscersi e a migliorare sé stesso.

Proprio questo offrono le scienze cognitive.

A questo servono la tesi della natura computazionale della cognizione e la tesi del carattere astratto delle computazioni: l'assunto è che ciò che vale nella teoria matematica degli algoritmi vale anche per i processi cognitivi degli umani.

Per comprendere Turing è importante leggere da sé i due suoi articoli, solo così possiamo cogliere il senso del suo progetto; ed evitiamo di cadere preda delle letture scolastiche che celebrano acriticamente la computazione.

Allo stesso modo, per cogliere il senso della psicologia cognitiva conviene, e basta, leggere *Judgment under Uncertainty: Heuristics and Biases*, di Amos Tversky e Daniel Kahneman, articolo più volte rimaneggiato, la versione a cui far riferimento resta quella del 1974.¹⁶ Psicologi israeliani, avevano avuto nei complessi anni '50 e '60 esperienze di guerra. La riflessione su come decidere in condizioni di incertezza nasce da questa vita vissuta.

Allo stesso tempo erano nutriti di psicologia comportamentista: forte l'influenza di Kurt Lewin, con la sua idea del campo di forza che agisce sull'individuo dall'esterno, spingendolo e tirandolo, suo malgrado, in varie direzioni. Ed erano allo stesso tempo educati al formalismo logico, alla matematica. Il punto di incontro di questa formazione non poteva che essere la computazione.

L'articolo è frutto di studi e ricerche degli Anni Sessanta, trascorsi presso l'Università Ebraica di Gerusalemme. Ma entrambi si trasferiscono negli Anni Settanta negli Stati Uniti. Dove torna

<https://static.trogu.com/documents/articles/palgrave/references/searle%20What%20Your%20Computer%20Can%20Know%20by%20John%20R.%20Searle%20%7C%20The%20New%20York%20Review%20of%20Books.pdf>

14 Bill Joy, "Why the Future Doesn't Need Us", *Wired* 8.04 , April 1, 2000. <https://www.wired.com/2000/04/joy-2/>

15 Jaron Lanier, "One Half a Manifesto", *Edge*, November 10, 2000.

16 Amos Tversky, Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases", *Science*, New Series, September 27, 1974, vol. 185, No. 4157, pp. 1124-1131.

<https://www2.psych.ubc.ca/~schaller/Psyc590Readings/TverskyKahneman1974.pdf>.

<https://apps.dtic.mil/sti/pdfs/AD0767426.pdf>

centrale il tema della decisione nei teatri di guerra: la ricerca è finanziata dal Dipartimento della Difesa.

Ma al volgere tra gli Anni Settanta ed Ottanta cresce l'ondata neoliberista, che finisce per imporre alla pubblica attenzione la nuova figura sociale ideale: lo speculatore finanziario. Sarà questo il campo d'azione al quale le ricerche di Tversky e Kahneman resteranno indissolubilmente associate. Tversky, il più geniale tra i due, muore prematuramente verso il termine del secolo – ma già da tempo i due non lavoravano più insieme.

Molte decisioni, notano Tversky e Kahneman aprendo l'articolo, si basano su convinzioni relative alla probabilità [likelihood] di eventi incerti: l'esito di un'elezione, la colpevolezza di un imputato, il valore futuro del dollaro. Queste convinzioni sono di solito espresse in affermazioni come "penso che ...", "è probabile che...", "è improbabile che..." e così via. Come si valuta la probabilità di un evento incerto o il valore di una quantità incerta? Tversky e Kahneman sostengono che le persone si affidano a un numero limitato di criteri di giudizio fondati sul ridurre i *complex tasks* di valutazione delle probabilità e delle previsione in *judgmental operations* più semplici. In generale, notano Tversky e Kahneman, queste semplificazioni sono abbastanza utili, ma a volte *portano a errori gravi e sistematici*. Nel paragrafo successivo questi errori sistematici sono ribattezzati *bias*.

Tutto il castello della psicologia cognitivistica si riduce a questo. Noi umani, in situazioni di incertezza decidiamo in base a all'intuito, al sentimento, ai sogni, ai desideri, all'inconscio. Più siamo in sintonia con noi stessi, più compiamo scelte senza sapere di preciso perché. Eppure sappiamo, sentiamo profondamente che per noi, in quell'istante, stiamo compiendo la scelta più giusta per noi.

Per Tversky e Kahneman, invece, si tratta di scelte sbagliate. Perché si allontanano dalla *computazione*, dal comportamento della *computer machine* di Turing, che esegue meccanicamente, senza deflettere, la procedura, compiendo passo dopo passo il predefinito numero finito di operazioni previste dal programma, esaminando senza scorciatoie ognuna delle probabilità della tabella che descrive, in astratto, il corretto comportamento di un generico agente razionale. Appunto: un generico agente razionale è posto a modello dell'essere umano.

Biases

Il successivo lavoro di Kahneman e di adepti e seguaci pochissimo o nulla aggiunge. Gli elenchi di *bias*, lungo l'arco di cinquant'anni ripetuti e allungati, con scarsissime sostanziali variazioni, non sono che una ripresa delle situazioni-tipo descritte in quell'articolo seminale.

Esistono liste di sei, dieci, cento *bias*. Liste di banalità, che riducono a ridicolo formalizzazioni atteggiamenti infinite volte già descritti, lungo l'intera storia umana, da filosofia, letteratura, arte. L'unica differenza è che ora, esistendo la *computer machine* come modello di perfezione ed esattezza, gli atteggiamenti sono battezzati come errore e difetto.

Pomposamente si parla *availability heuristic* [aka "What You See Is All There Is" or WYSIATI]. E' un errore, *bias*, basarsi su esempi che vengono in mente a una persona. Si dice che è un errore l'*Anchoring*, l'eccessivo affidamento ad informazioni scarse. L'antica arte della prudenza è ribattezzata *Loss aversion*: è un errore, perché ci allontana da guadagni possibili.

In fondo, a dimostrazione di questa pochezza, basta citare *Attribute substitution*. E' il *bias* esemplare, perché ripete nel modo più preciso cioè che Tversky e Kahneman sostenevano nel '74. "Si verifica quando un individuo deve esprimere un giudizio su qualcosa di *computationally complex*", computazionalmente complesso, "e vi sostituisce qualcosa di *more easily calculated*", più facilmente calcolabile. Cioè: non possiamo scegliere il modo di pensare, il modo di calcolare. Dobbiamo attenerci alle regole auree della computazione. Ciò che è buono per la *computer machine* deve essere buono per noi umani.

Storia economica e politica

La fortuna pubblica della psicologia cognitiva è dovuta ad un passaggio: dal tipo-ideale 'soldato' al tipo ideale 'speculatore finanziario'. L'uomo-che-guadagna-giocando-in-Borsa è l'eroe dell'ultimo ventennio del Ventesimo Secolo.¹⁷

L'agire umano è mosso da solidarietà, cooperazione, benevolenza, orientamento all'interesse comune. Ma dalle stesse origini dell'economia come disciplina autonoma -inizio del 1800- il peso della dottrina utilitarista è prevalente. L'essere umano, si dice, è mosso dal pragmatico orientamento a diminuire la pena e ad incrementare il piacere. La vita si riduce a calcolo costi benefici. Le scelte economiche si riducono a diminuzione delle perdite ed incremento dei profitti. Il profitto finanziario è il più puro ed auspicabile: evita egoisticamente la fatica del lavoro materiale, evita l'onerosa trasformazione delle materie prime in prodotti finiti e le difficoltà implicite nell'organizzazione delle attività produttive.

L'ascesa dell'economia neoliberista, a partire dagli Anni Settanta del secolo scorso, è il trionfo della finanza speculativa a scapito dell'economia produttiva. Non a caso la psicologia cognitiva nasce e si afferma nello stesso arco temporale. Gli schemi di comportamento personali, le strategie d'azione messe in opera da ogni persona -Tversky e Kahneman le chiamano *euristiche*- sono ridotte dal cognitivismo alla fredda logica dello speculatore. E' questa la via per la quale Tversky, Kahneman e sodali considerano emozioni, empatia, affetti pericolosi *bias*: "L'euristica degli affetti è un tipo di scorciatoia mentale in base alla quale le persone prendono decisioni fortemente influenzate dalle emozioni del momento".

Questa educazione del popolo tesa a considerare modello sociale l'uomo-che-guadagna-giocando-in-Borsa trova definitiva affermazione nel premio Nobel riconosciuto a Kahneman nel 2002. Premio Nobel, si dadi bene, per l'economia, "per aver integrato gli *insights* della ricerca psicologica nella scienza economica, in particolare per quanto riguarda il giudizio umano e il processo decisionale in condizioni di incertezza". Tversky, che probabilmente più di Kahneman meritava il riconoscimento, era scomparso sei anni prima.

Il premio appare postumo omaggio alla Presidenza di Bill Clinton. Degno completamento della liberalizzazione del settore finanziario promossa dal Financial Services Modernization Act, entrato in vigore nel 1999, e della globalizzazione culminata nel 2001 con l'ingresso della Cina nella World Trade Organization. Il liberismo porta con sé la crescente divaricazione tra ricchezza e povertà. Si spaccia per razionalità, proposta come atteggiamento sociale ideale ad ogni cittadino, il comportamento degli appartenenti ad una élite sempre più ristretta di aristocrazia del denaro. Il passo successivo è sancito dal premio Nobel dell'Economia, riconosciuto nel 2017 a Richard Thaler. Premio, in questo caso, alla Presidenza Obama.

Le campagne elettorali di Obama, in entrambi i suoi mandati, sono sostenute dalla lobby di Silicon Valley. La sua presidenza coincide con l'affermazione di una nuova scena sociale. La ristretta cerchia dei magnati digitali influenza ed anzi indirizza le politiche pubbliche. La vita sociale dei cittadini si svolge su piattaforme digitali. Il dibattito pubblico si appoggia sui Social Network. Eric Schmidt, CEO di Google, consigliere e finanziatore di Obama, afferma: "la maggior parte delle persone non desidera che Google risponda alle loro domande. Vogliono che Google dica loro cosa fare". Ecco il senso degli algoritmi in base ai quali funzionano i motori di ricerca. Ecco il senso degli algoritmi che pilotano sui Social Network l'opinione pubblica. Ecco il senso della linguistica computazionale: rendere indistinguibile il messaggio scritto da un *bot* dal messaggio scritto da un umano.

Thaler, seguace di Kahneman, appartenente alla ristretta cerchia dei consulenti di Obama, porta a sintesi le ricerche della psicologia cognitiva e ne trae lo strumento finale della propaganda, della persuasione occulta: il *nudge*. *Nudge*: gomitatina, spintarella.¹⁸ Siccome conosco i tuoi *bias*, sono in

17 Dillon Jacobs, *What Daniel Kahneman Thinks Investors Should Know*, Finmaster, March 11, 2022.
<https://finmasters.com/daniel-kahneman-investors/>

18 Richard H. Thaler, Cass R. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness*, Yale University Press, New Haven, 2008.

grado di condizionare i tuoi comportamenti. Sul Web e tramite le app: invito capzioso a schiacciare quel bottone. Notifica ammiccante: 'nel tuo interesse, fai questo'.

I sostenitori delle democrazie liberali e della società aperta cessano di credere nella cittadinanza attiva, nell'educazione diffusa, nel voto popolare, nella partecipazione sociale, nelle scelte civiche ed elettorali. Le azioni degli esseri umani, anche le azioni orientate alla partecipazione democratica, al civismo, si manifesteranno solo come risposte ad uno stimolo dei detentori del potere.

La democrazia è sostituita dal paternalismo. Di paternalismo parlano infatti Kahneman e Thaler. Se poi aggiungono l'aggettivo 'gentile', nulla cambia.

E il quadro si completa. L'iniziale figura che rappresenta l'essere umano razionale è, negli Anni Sessanta del secolo scorso, il soldato. Negli ultimi Venti Anni l'essere umano razionale è lo 'speculatore finanziario'. Nei primi venti anni del nuovo secolo, con il trionfo della cultura digitale, un'altra più misera figura è oggetto d'attenzione degli psicologi cognitivi. Nella polarizzazione sociale si colloca all'opposto dello speculatore finanziario, esponente simbolo dell'élite. E' il cittadino-massa¹⁹, ridotto a utente di servizi digitali, educato tramite *nudge*.

Inclinazione e pregiudizio

E' sempre bene chiedersi come si potrebbero tradurre in italiano le parole dietro le quali si nasconde il potere delle élite dominanti - élite politiche, tecniche o scientifiche.

Il termine bias può essere tradotto con due parole: *inclinazione e pregiudizio*.

Kahneman e Thaler sostengono che l'agire degli umani è viziato da *severe and systematic errors: bias*, appunto.

Si dice ad ogni umano: *sei viziato da pregiudizi*. A ben guardare, è la scoperta dell'acqua calda. Nel momento in cui, qui ed ora, di fronte ad una situazione nuova, mai prima sperimentata, non dispongo di nient'altro che pre-giudizi. Giudizi formulati prima, in tempi precedenti, frutto di mie o altrui esperienze che hanno avuto luogo in un tempo passato.

Qualsiasi lettura di un presente che appare in questo istante ai miei occhi, qualsiasi osservazione di un terreno sconosciuto che calpesto per la prima volta sarà parziale, insufficiente ad offrirmi regole certe in grado di guidare l'azione.

Il punto è che questa ovvia constatazione viene usata dagli psicologi cognitivistici per svalutare ai miei stessi occhi la mia capacità di leggere il contesto, e di adattare l'esperienza alla situazione presente. Mi si dice: il tuo comportamento sarà in ogni caso viziato da *bias*. Mi si colpevolizza e demotiva, per dirmi in realtà: ecco la regola -da altri definita- alla quale ti devi attenere. Le liste di bias sono in realtà liste di comportamenti attesi dall'élite dominante. Articolazioni di un progetto di controllo sociale.

Si deprime così la fiducia in sé stessi, il coraggio, la ricerca di consapevolezza, la motivazione, lo spirito di iniziativa. Al desiderio, al sogno, alla speranza di miglioramento, alla costruzione di futuro che ogni essere umano porta in sé, si sostituisce la descrizione di ognuno in base ad uno o un altro *bias*. Alla faticosa ricerca di libertà, si sostituisce il *nudge*: l'invito sottilmente proposto ad attenermi alle regole. Gli altri *pregiudizi* nei miei confronti mi sono imposti di autorità.

Si invita così, nei tempi digitali, il cittadino a ridursi ad utente di servizi preconfezionati. Non c'è poi da meravigliarsi se la cittadinanza attiva scompare. E se scompare anche il senso del lavoro: spogliato dagli aspetti costruttivi, il lavoro si riduce a penosa esecuzione di un compito assegnato.

L'altra traduzione di *bias* è *inclinazione*. Noi esseri umani siamo imperfetti, deboli, viziati da difetti. Ma non chiamiamoli per favore *bias*, errori di sistema. Chiamiamoli inclinazioni. Ogni essere umano si discosta dalla norma. Il comportamento di qualsiasi essere umano non sarà mai del tutto conforme al comportamento ideale descritto per via logica, previsto in base ad una astratta ragione. Ogni essere umano si contraddistingue per il proprio *carattere*. Carattere vuol dire impronta; ognuno lascia la propria diversa impronta. Ogni essere umano ha una propria idiosincrasia: vuol dire

19 Francesco Varanini. *Macchine per pensare*, Guerini e Associati, 2016.

'temperamento particolare.

Siamo storti, claudicanti. Ma in questo sta la nostra ricchezza. Ciò che l'astratta ragione vede come deficit, alla luce della saggezza appare fertile differenza. Sappiamo che potremo sbagliarci, ma sappiamo anche di essere in grado di apprendere dagli errori. Al momento di prendere decisioni, l'esperienza, sia pure limitata, sia pure maturata in un passato diverso dal presente, ci soccorre. L'euristica, il modo di procedere di ogni umano, non è difettoso perché si scosta da uno standard. All'opposto, il valore dell'essere umano sta nello scostamento dallo standard. Ognuno ha i propri talenti. Accettando di lasciarci giudicare in base a questo o quest'altro *bias*, svalutiamo noi stessi, e spieghiamo l'irripetibile contributo che potremmo dare all'organizzazione per la quale lavoriamo, alla società.

Il pensiero umano non è un caso particolare di computazione, è qualcosa di più e di sempre differente.

Responsabilità

Possiamo sempre chiederci: chi ha il diritto di scrivere la tavola del corretto comportamento al quale devono attenersi i cittadini, il lavoratori, gli umani tutti? Chi è tanto saggio, tanto distaccato dai propri immediati interessi, da poter svolgere questo ruolo?

Due comunità professionali si legittimano a vicenda, si offrono l'un l'altra strumenti, condividono percorsi di ricerca e successi. I computer scientist e i tecnici digitali sono debitori degli scienziati cognitivi, e viceversa.

Machine learning e deep learning, i modelli di linguaggio che simulano i linguaggi umani, le tecniche usate nella scrittura di algoritmi, il modo in cui vengono via via evoluti i motori di ricerca, interfacce uomo-macchina e User Experience Design: tutti questi sviluppi della computer science discendono da ricerche di scienze cognitive. Le scienze cognitive, però, non sarebbero nate se non fosse esistita la computer science.

Le due famiglie professionali, così, condividono la responsabilità implicita nel progettare e nel gestire sistemi che tolgonon libertà agli esseri umani.

Ma -se mai è possibile stabilire in questo campo un ordinamento- il peso etico più duro grava sulle spalle dei computer scientist e dei tecnici digitali. Essi offrono la metafora senza la quale le scienze cognitive non esisterebbero: la computer machine. Macchina che nasce dal considerare l'essere umano difettoso, bisognoso di modelli esterni, superiori. Così si apre la strada a questa ambigua psicologia secondo la quale le inclinazioni umane errori di sistema.

Meglio evitare, parlando di esseri umani, la parola *bias*.