

IL TEST DI TURING

STORIA E SIGNIFICATO

Appare molto difficile individuare un criterio oggettivo che ci consenta di distinguere il pensiero da una sua imitazione meccanica. Il punto critico è la definizione di pensiero. In un articolo del 1950 Turing trasformò la domanda se una macchina possa pensare in un problema indiziario basato su una prova comportamentistica di tipo dialogico: se un giudice non riesce a distinguere il comportamento verbale di un computer da quello di un essere umano, che per definizione pensa, allora il computer pensa. Il *test di Turing* ha suscitato negli anni, e ancora suscita, molte discussioni, sia per la vaghezza della formulazione originale sia in relazione ai progressi via via compiuti dall'intelligenza artificiale.

1. INTRODUZIONE

“Il metodo della scienza è lo stesso che serve per tutta la conoscenza delle cose di questo mondo: consiste nel limite, o addirittura nell'ignoranza. Posti di fronte alla grande, illimitata unità della natura, noi non possiamo far altro che manipolare col pensiero, scegliendo certi dettagli e (deliberatamente o inconsciamente) isolandoli da tutto il resto”

Edward Carpenter,
Civilisation: its Cause and Cure

“Soltanto il nostro cervello umano può prestare un significato alla cieca capacità dei calcolatori di produrre verità”

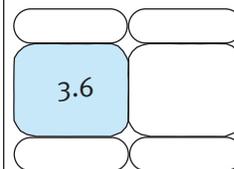
Karl Popper

Nell'articolo *Computing Machinery and Intelligence*, pubblicato sulla rivista *Mind* nel 1950, il grande matematico e logico inglese Alan Mathison Turing (Londra, 1912 - Manchester 1954) descrisse un *Gedankenexperiment*, cioè un esperimento concettuale, per stabilire se una macchina (un computer) sia o no in grado

di pensare. L'esperimento, oggi noto come *test di Turing*, prende lo spunto dal “gioco dell'imitazione”, nel quale un esaminatore cerca di distinguere un uomo da una donna basandosi soltanto sulle risposte a una serie di domande. Turing immagina di sostituire alla donna una macchina, assegnando all'esaminatore il compito di distinguere l'uomo dal computer. La versione modificata del gioco consente a Turing di affrontare il problema: *le macchine possono pensare?* evitando le difficoltà definitorie legate ai termini “macchina” e “pensare”, il cui significato dovrebbe essere precisato in via preliminare. L'utilità di questo esperimento concettuale sta non tanto nella risposta che esso consente di fornire al problema di partenza (non è scontato che dia una risposta), quanto nella possibilità che esso offre di analizzare concetti come *mente, pensiero, intelligenza*. Soprattutto il concetto di intelligenza, che in quegli anni stava diventando il terreno d'incontro di discipline quali la logica, la matematica, la psicologia e la filosofia, che a vario titolo si occupavano dell'uomo, o meglio delle sue capacità cognitive, ritenute le facoltà umane



Giuseppe O. Longo



più caratteristiche. Era importante possedere un criterio se non per *definire* l'intelligenza, almeno per riconoscerne la *presenza* sia negli organismi biologici sia negli artefatti. Non dimentichiamo che erano gli anni della *cibernetica*, una disciplina che si proponeva, secondo il programma del suo fondatore Norbert Wiener, di studiare in modo unificato "la comunicazione e il controllo nell'animale e nella macchina". Di lì a poco, nel 1956, per iniziativa del ma-



FIGURA 1
Alan Turing

tematico John McCarthy si tenne a Dartmouth, nel New Hampshire, un seminario in cui fu scelta la locuzione "intelligenza artificiale" (IA) per indicare un insieme di ricerche piuttosto eterogenee (anche per le discipline di afferenza dei vari partecipanti) che avevano tuttavia l'obiettivo comune di produrre simulazioni o emulazioni di comportamenti intelligenti da parte dei calcolatori elettronici.

Nei paragrafi che seguono esporremo il criterio di Turing e lo sottoporremo a un'analisi critica, per metterne in luce i pregi e i limiti. Pur essendo la trattazione autonoma, faremo costante riferimento all'articolo originale di Turing, la cui lettura è ancora di grande interesse e al quale rimandiamo il lettore (Figura 1).

2. IL GIOCO DELL'IMITAZIONE

Il gioco dell'imitazione consiste in uno scambio linguistico cui partecipano tre persone: un uomo A, una donna B e un terzo individuo C, il giudice o esaminatore, il cui sesso è irrilevante. C, che è isolato dagli altri due, può porre ad A e B domande di qualunque tipo e natura e, basandosi soltanto sulle risposte ricevute, deve stabilire se A è l'uomo e B la donna o viceversa. A e B si sforzano entrambi di essere identificati nella donna, quindi B cerca di aiutare C e A cerca di ingannarlo, fornendo le risposte che a suo parere darebbe la donna. C può avere a disposizione soltanto le risposte e non

Turing: cenni biografici

Nato a Londra nel 1912 da una famiglia della classe media, Alan Mathison Turing manifestò fin da piccolo un carattere solitario e indipendente. Con l'aiuto di un manuale, imparò a leggere da sé in tre settimane. A scuola non brillava nelle materie umanistiche e preferiva dedicarsi a letture scientifiche. Ammesso nel 1931 al King's College dell'Università di Cambridge, vi studiò meccanica quantistica, logica e calcolo delle probabilità, laureandosi nel 1934 col massimo dei voti. Nel 1936, trasferitosi alla Princeton University per specializzarsi, pubblicò il fondamentale articolo *On Computable Numbers, with an Application to the Entscheidungsproblem*, in cui descrisse la cosiddetta "macchina di Turing". Durante la seconda guerra mondiale, ebbe una parte di primo piano nella forzatura del sistema *Enigma*, impiegato dai tedeschi per cifrare i messaggi, e fu vicino al gruppo di crittanalisti che progettò Colossus, il primo computer elettronico digitale programmabile, pur senza prender parte direttamente alla costruzione. Verso la fine della guerra, dopo aver lavorato a dispositivi per la crittografia vocale, Turing ricevette un'alta onorificenza, l'OBE (Ordine dell'Impero Britannico), per i suoi servizi a favore della Gran Bretagna, ma i suoi contributi rimasero segreti per molti anni. Dal 1945 al 1947 si dedicò al progetto dell'ACE (*Automatic Computing Engine*), un computer davvero pionieristico, dotato di un'architettura molto diversa da quella, quasi contemporanea, proposta da von Neumann. Nel 1951 fu eletto membro della Royal Society. Oltre all'articolo di cui si parla in queste pagine, Turing scrisse, sullo stesso argomento, altre due importanti relazioni: *Intelligent Machinery* (1948) e *Digital Computers Applied to Games* (1953), divenendo così un antesignano dell'intelligenza artificiale, anche grazie allo studio delle relazioni tra le macchine e la neurologia e fisiologia degli esseri viventi. Dal 1952 alla morte, nel 1954, si dedicò infatti alla formalizzazione matematica della biologia, in particolare alla morfogenesi, studiando, tra l'altro, i collegamenti tra le forme vegetali e i numeri di Fibonacci. Arrestato nel 1952 per omosessualità, fu condannato a una pena severissima, la castrazione chimica, che lo rese impotente. Fu probabilmente questo il motivo che nel 1954 indusse Turing al suicidio, che mise in atto mangiando una mela atossicata col cianuro di potassio. Benché contestata dalla madre, l'ipotesi del suicidio appare quasi certa. Nel libro *Zeros and Ones*, la Plant suggerisce che la Apple Inc. abbia scelto a proprio simbolo la mela morsicata in omaggio ad Alan Turing, ma l'azienda non ha mai né confermato né smentito questa congettura.

altri indizi (l'aspetto, la voce, la grafia) che renderebbero facile l'identificazione dei soggetti, quindi può comunicare con A e B solo per tele-scrittore o con un mezzo altrettanto neutro. Turing immaginò di modificare il gioco dell'imitazione, sostituendo ad A una macchina, e si chiese se in queste condizioni l'esaminatore C avrebbe sbagliato con la stessa frequenza con cui sbagliava nel gioco dell'imitazione classico. C sbaglia quando ritiene che l'essere umano sia A e non B; in tal caso la macchina vince, e viene dichiarata capace di pensare. Vogliamo ora osservare da vicino il gioco del-

l'imitazione e la variante che ne propose Turing. Chiediamoci intanto: quando esce dai salotti per diventare *Gedankenexperiment*, che cosa dimostra o vorrebbe dimostrare il gioco dell'imitazione? (riquadro 1).

Già proporre il gioco significa ritenere che le risposte di un uomo siano distinguibili da quelle di una donna, cioè che il pensiero maschile differisca in misura apprezzabile da quello femminile. Ma significa anche che, pur essendo forse l'intelligenza della donna diversa da quella dell'uomo, questi può essere tuttavia capace di simularla.

RIQUADRO 1 - Alcune varianti possibili del gioco dell'imitazione

Chiamiamo A, B e C i partecipanti al gioco dell'imitazione. Ecco alcune delle varianti possibili:

1. A uomo, B donna, C giudice umano (gioco originale)
2. A macchina, B umano, C giudice umano (variante di Turing)
3. A macchina, C giudice umano (prima variante a due)
4. A umano, C giudice umano (seconda variante a due)
5. A uomo, B donna, C giudice macchina
6. A umano, C giudice macchina (terza variante a due)
7. A macchina, C giudice macchina (quarta variante a due)

In tutte le varianti il giudice deve stabilire la natura di A. Nella prima variante il giudice deve sapere che ha di fronte, senza vederli, un uomo e una donna. In tutte le altre varianti, il giudice può essere al corrente oppure no delle identità possibili di A (e di B, se c'è). Turing stesso non precisa se nella variante 2 il giudice sappia che uno dei due contendenti è una macchina, anche se è abbastanza ovvio ritenerlo. Se per esempio nella variante 3 il giudice non sa che A può essere una macchina, è possibile che concluda sempre che si tratta di un umano, magari stupido, pazzo o burlone. A maggior ragione ciò accade nella variante 4, in cui gli errori commessi dall'umano per confondere il giudice avrebbero la qualità umana della pertinenza (riquadro 2). Quindi, per rendere più interessante il gioco, il giudice deve essere al corrente delle alternative possibili. Nei casi 5, 6 e 7 fornire al giudice macchina l'informazione sulla possibile natura degli interlocutori è problematico, com'è problematico immaginare quali domande possa fare C per condurre l'esame (e poi c'è da chiedersi se C abbia o no superato a sua volta il test di Turing, vedi Figura). In definitiva sono tre varianti poco significative, ma confermano forse che il test riguarda il giudice quanto e più di A e B: almeno dal nostro punto di vista, solo un'intelligenza umana può giudicare un'intelligenza umana (o d'altro tipo). Ciò sembra confermare che l'intelligenza umana è, da noi, ritenuta superiore a quella della macchina, così come negli esami l'intelligenza (o la sapienza) dell'esaminatore supera (o dovrebbe superare) quella dell'esaminato. Non so se questo antropocentrismo sia inevitabile come mi appare in questo momento. La variante 7 fa tuttavia intravedere una prospettiva nuova: gli umani giocheranno con gli umani e con le macchine fino a quando le macchine non sapranno giocare come gli umani e anche meglio; a quel punto è plausibile che gli umani giochino solo con gli umani e le macchine solo con le macchine (questo comincia ad accadere con i programmi scacchisti e già da un pezzo accade nella corsa: cavalli con cavalli, automobili con automobili e - aggiungiamo - uomini con uomini, donne con donne!). Ma quando le macchine avranno superato gli umani, forse i giudici più qualificati in tutte le varianti saranno macchine, cioè dovremo adottare un punto di vista macchina-centrico.

L'idea di fondo è che, almeno finora, non è possibile formalizzare il concetto generale di intelligenza. Si può convenire di porre il confronto tra la macchina e l'essere umano all'interno di un perimetro ben preciso, una sorta di gioco, appunto, regolato da parametri quantificabili, per esempio - nel caso della conversazione - la ricchezza del vocabolario, le varianti costruttive delle frasi, la concatenazione logica, la frequenza delle metafore e ovviamente la durata della partita. Ma sorge subito un problema: quale dovrebbe essere l'ambito della conversazione? Oppure la conversazione dovrebbe spaziare su qualsiasi tema? S'intuisce che nel secondo caso nessun programma di conversazione può oggi superare il test se il confronto dura anche poco (riquadro 5), mentre nel caso di ambiti molto limitati il superamento del test non dimostra nulla: la conversazione è un processo interattivo, costruttivo, autopoietico, arborescente, in parte aleatorio e in parte determinato, che riflette da vicino gli eventi della vita, con la loro imprevedibilità e la loro capricciosità, ma anche con la loro logica complessa aperta sul mondo. Come dice Licata, "il processo della cognizione umana è ancorato al confine complesso e imprevedibile tra un sistema e l'ambiente che lo circonda".

TURING TEST EXTRA CREDIT:
CONVINCE THE EXAMINER
THAT HE'S A COMPUTER.

YOU KNOW, YOU MAKE
SOME REALLY GOOD POINTS.
I'M ... NOT EVEN SURE
WHO I AM ANYMORE.



Una bizzarra variante del gioco dell'imitazione. Il giudice dice al computer: "hai certe idee davvero interessanti, sai" e soggiunge: "non so nemmeno più chi sono..." la dicitura in alto recita: "avrete un punto in più nel test di Turing se convincerete il giudice che lui è un computer"

Non si dimentichi tuttavia che il gioco è a tre: l'identificazione del sesso di A e B è demandata a un giudice. Se il giudice è in gamba può prendere, aiutato da B, una decisione corretta nonostante gli sforzi di A per sviarlo. Quindi sorge la domanda: si tratta di stabilire se A è più in gamba di B oppure se C è abbastanza in gamba? Forse il gioco è un test per C: cambiando giudice, ma non A e B, l'esito del gioco potrebbe cambiare. Si può anche immaginare di ripetere il gioco con lo stesso giudice, facendo molte "mani" e fornendo una valutazione statistica dei risultati. Oppure si può immaginare che vi siano molti giudici in parallelo che, decidendo tutti in base alle risposte fornite alla stessa serie di domande, possono arrivare a decisioni diverse.

Nella versione originale si tenta di distinguere un uomo da una donna. Questa è la dicotomia più classica, ma potremmo immaginare altre coppie di concorrenti: un bianco e un nero, un italiano e un cinese, un poeta e un matematico (tutti di sesso arbitrario). E la cosa si potrebbe raffinare, mettendo a confronto un matematico cinese e un poeta italiano. Con questi esempi si capisce forse meglio che C, il quale conosce le categorie cui appartengono A e B, è portato a decidere sì in base alle risposte, ma anche in base alle sue credenze, ai suoi pregiudizi, alla sua conoscenza del mondo e così via.

In altre parole, anche se ci si serve di una telescrivente, nelle risposte resta un contenuto che non è puramente sintattico, interno alla lingua, ma anche e fortemente semantico, nel senso che è legato alla natura, non solo cognitiva, di A e di B. Naturalmente la presenza e l'entità affiorante di questo contenuto dipende dalle risposte e quindi dipende dalle domande e dalle aspettative del giudice quanto alle risposte.

Se fosse possibile scegliere domande di natura squisitamente sintattica (di tipo logico, matematico) che contemplassero solo risposte di natura sintattica, forse distinguere un uomo da una donna sarebbe impossibile (ma le recenti scoperte dei neurofisiologi sembrano indicare differenze cerebrali notevoli tra i due sessi, quindi una possibile differenza di pensiero). Naturalmente qui si parla di uomo e di donna generici, cioè scelti a caso, e sconosciuti all'esaminatore. Ne seguirebbe che la distinzione tra sesso maschile e femminile si potreb-

be operare solo sulla base di fatti *non* riconducibili a successioni di simboli. O meglio, a fatti extralinguistici che tuttavia possono lasciare la loro impronta sulle successioni di simboli.

Esiste cioè la possibilità che le differenze (di esperienza, di ruolo sociale, di caratteristiche biologiche e via dicendo) tra uomo e donna introducano differenze apprezzabili, e non colmabili mediante imitazione, nelle risposte di A e B a *certe domande*. Ciò indicherebbe che il pensiero non si può mai isolare o astrarre del tutto dal resto dell'essere umano, per esempio dal suo corpo e dai suoi condizionamenti, quindi dal mondo. Insomma, se le differenze nelle risposte di A e B derivassero dalle differenze *tout court* tra A e B, un giudice abbastanza bravo darebbe sempre la risposta corretta perché B non potrebbe mai imitare A: per imitare A dovrebbe *essere* A.

3. LA MENTE E IL CORPO

Turing non la pensa in questo modo e ritiene che il pensiero si possa separare dal suo supporto. Infatti, perché il gioco abbia senso, C deve essere isolato da A e B, altrimenti distinguerebbe d'acchito la donna dall'uomo oppure, nella variante, la macchina dall'essere umano in base all'aspetto esteriore. Turing estende questa condizione al divieto di considerare caratteristiche o capacità corporee (la bellezza, l'abilità nel ballo e così via), che sarebbero discriminanti nei confronti della macchina e che in realtà, per lui, non aggiungono nulla all'intelligenza di chi le possiede. Traspare da questa clausola il pregiudizio di Turing, ma non solo suo, secondo cui l'intelligenza è un fatto esprimibile e riconoscibile soltanto mediante sequenze ben costruite di simboli. Il corpo non avrebbe nulla a che fare con l'intelligenza. Questo riduzionismo mentale o informazionale di Turing è confermato dal fatto che le attività umane che egli riteneva adatte all'automazione erano quelle che non comportano contatti con il mondo esterno: il gioco degli scacchi, la matematica, la crittografia, forse anche la traduzione da una lingua all'altra. Ma gli aspetti corporei erano esclusi. In una trasmissione radiofonica dichiarò:

"Credo di sicuro, e spero, che non si faranno molti sforzi per costruire macchine dotate di

caratteristiche (a parte quelle intellettuali) specificamente umane, come per esempio la forma del corpo. Mi sembra futile fare tentativi del genere, i cui risultati avrebbero all'incirca la stessa sgradevole qualità dei fiori artificiali. Fare una macchina pensante mi sembra un lavoro che rientra in una categoria diversa”.

Tuttavia nell'articolo del 1950 Turing contraddice in parte questa sua posizione: estendendo l'attività delle macchine intelligenti alla conversazione, anzi facendo della capacità linguistica addirittura il segno distintivo del pensiero, forse senza rendersene conto apre un grande varco e stabilisce un contatto tra la macchina e il mondo esterno. Infatti, l'esaminatore può porre domande arbitrarie, quindi anche domande sui fatti del mondo (cioè del mondo abitato da noi umani). Il legame tra parola e mondo, che si manifesta quando la parola cessa di essere autoreferenziale, cioè puro simbolo sintattico all'interno di un formalismo alla Hilbert, per caricarsi di semantica e alludere alle cose esterne alla lingua, questo legame, per cui la parola crea, modifica e struttura il mondo e ne è condizionata, ora coinvolge anche la macchina (Figura 2).

C'è dunque un certo grado di confusione. Da una parte Turing riconosce che l'intelligenza dell'uomo deriva dalla sua interazione con il mondo e quindi dal suo equipaggiamento corporeo, che è il tramite che consente quell'interazione. Dall'altra emerge la convinzione che la macchina a stati discreti, la cosiddetta “macchina di Turing” (riquadro 2 a p. 16, 17), possa usare le parole in modo umano, al punto da ingannare l'esaminatore, senza essere dotata di organi di senso e di organi effettori che le consentano di interagire con la realtà esterna. Il gioco dell'imitazione si basa sull'interazione linguistica, dunque su un segno esteriore del pensiero, che è inaccessibile. In questo senso è un criterio comportamentistico. Le sequenze verbali sono lì a manifestare la propria esistenza, mentre il pensiero, da cui (forse) scaturiscono, ha uno statuto esistenziale molto più dubitoso e, a quei tempi, nella prima metà del Novecento, era bandito dalle riflessioni di psicologi e filosofi. Turing accetta dunque in pieno la posizione comportamentista allora imperante. Ma nell'accettarla egli non nega l'esistenza del pensiero, anzi con-



FIGURA 2
David Hilbert
(1862-1943)

tribuisce alla formulazione di un problema tuttora vivo nelle scienze cognitive, il rapporto tra i vari livelli di funzionamento e di descrizione. Infatti, tra le obiezioni che egli considera¹, merita attenzione la seguente:

“Non possono forse le macchine comportarsi in una maniera che deve essere descritta come pensiero, ma che è molto differente da ciò che fa un uomo?”

¹ Nell'articolo Turing anticipa le obiezioni dei critici, elencando e discutendo nove opinioni contrarie: 1. *Obiezione teologica*: pensare è una funzione dell'anima immortale, che gli uomini posseggono e le macchine no; 2. *Obiezione dello struzzo*: se le macchine pensassero, le conseguenze sarebbero terribili; speriamo e crediamo che esse non pensino; 3. *Obiezione matematica*: ci sono domande alle quali la macchina non può rispondere in base ai teoremi di Gödel; 4. *Argomento della coscienza*: si veda nel testo a proposito delle dichiarazioni di Jefferson; 5. *Argomento fondato su incapacità varie*: le capacità di tipo fisico elencate impedirebbero alla macchina di pensare (si vedano nel testo le considerazioni a proposito delle fragole con la panna); 6. *Obiezione di Lady Lovelace*: la macchina non è in grado di creare alcunché, può solo eseguire i nostri ordini; 7. *Argomento della continuità del sistema nervoso*: siccome il sistema nervoso non è una macchina a stati finiti, non si può sperare di imitarne il funzionamento con una macchina di quel genere; 8. *Argomento del comportamento senza regole rigide*: siccome l'uomo non si comporta in base a un insieme di regole rigide, ma adotta una condotta flessibile, l'uomo non è una macchina; 9. *Argomento fondato sulla percezione extrasensoriale o ESP* (argomento che Turing definisce molto forte): poiché ci sono prove schiaccianti a favore dell'ESP nell'uomo, uomo e macchina si troverebbero in posizioni diverse quanto alla risposta e l'identificazione sarebbe facilitata.

RIQUADRO 2 - La macchina, Hilbert e Gödel

Come si è accennato, per Turing la “macchina” è una macchina digitale a stati finiti, cioè un calcolatore particolare, prototipo astratto dei calcolatori concreti, chiamato appunto “macchina di Turing” (MT). Si tratta di un dispositivo ideale, munito di un nastro di carta illimitato diviso in caselle, capace di eseguire queste operazioni: far scorrere il nastro avanti e indietro, scrivere un simbolo su una casella vuota e cancellare il contenuto di una casella scritta. Questa macchina può eseguire qualsiasi operazione descrivibile mediante un algoritmo (macchina universale). Turing dimostrò tuttavia che la MT non era in grado di fornire risposta a tutti i problemi matematici e collegò questa scoperta ai teoremi di Gödel (Foto).

Secondo Hodges la MT, che comunica con il mondo solo per il tramite esiguo del programma, rappresenta anche - potenza delle esperienze personali nel foggiare le concezioni scientifiche! - l'ideale di vita di Turing: “una vita in cui lo si fosse lasciato da solo in una stanza tutta sua, a trattare col mondo esterno esclusivamente per mezzo di argomenti razionali”. Contro questa interpretazione si può tuttavia citare il seguente passo dalla tesi di dottorato di Turing a Princeton.

“Così facendo non teniamo conto di quella importantissima facoltà grazie alla quale gli argomenti di qualche interesse vengono distinti da tutti gli altri; al contrario consideriamo sola funzione del matematico quella di determinare la verità o falsità delle proposizioni”.

La facoltà di cui parla Turing dipende dal contatto con la realtà e non può quindi essere posseduta dalla macchina. Anche il passo seguente indica l'importanza che Turing attribuiva al contatto con il mondo.

“È opinione corrente che gli scienziati procedano inesorabilmente da un fatto ben stabilito a un altro fatto ben stabilito, senza lasciarsi mai influenzare da una congettura non dimostrata. Questa opinione è del tutto sbagliata. Purché venga messo chiaramente in evidenza quali siano i fatti provati e quali siano le congetture, non può aversi alcun danno. Le congetture sono di grande importanza, giacché suggeriscono utili linee di ricerca”.

Egli pare dunque oscillare, senza decidersi, tra una vagheggiata autoreferenzialità ideale del mondo linguistico, in particolare logico-matematico, e un'apertura comunicativa e attiva verso il mondo reale.

Un esempio cospicuo di autoreferenzialità in ambito matematico ci è offerto dal cosiddetto “programma di Hilbert”. All'inizio del Novecento il matematico David Hilbert aveva sostenuto la possibilità di costruire tutta la matematica con il metodo assiomatico, usando un numero finito di proposizioni di partenza (gli assiomi) e un numero finito di regole di inferenza. Gli enti matematici sarebbero stati definiti in modo implicito e astratto, senza legami semantici con il mondo esterno. In altre parole Hilbert evitava deliberatamente e programmaticamente di “definire” gli enti matematici, che sarebbero stati “nominati” e poi collegati tra loro e caratterizzati dalle reciproche relazioni. Poi, usando le regole, dagli assiomi si sarebbero potuti ricavare i teoremi con un procedimento meccanico.

Ma nel 1931 il logico Kurt Gödel fissò limiti molto precisi al programma di Hilbert dimostrando due teoremi. Il primo dimostra che in un sistema formale abbastanza potente da consentire almeno una formalizzazione dell'aritmetica (vale a dire una definizione della struttura dei numeri naturali con le operazioni di somma e prodotto) si può ricavare dagli assiomi una proposizione indecidibile, cioè di cui non si può dimostrare né la verità né la falsità. Il secondo teorema dimostra che all'interno del sistema formale è impossibile dimostrare la sua coerenza. Questi teoremi non implicano tanto uno scacco alla matematica o alla ragione umana, quanto una sconfitta della speranza di Hilbert di separare la matematica da ogni riferimento esterno: un matematico non si limita a manipolare simboli, ma fa riferimento al contesto metamatematico, all'importanza pratica dei problemi, compie scelte e individua legami tra problemi e risultati in apparenza lontani. Di fronte a una proposizione indecidibile, per esempio, ricorre alla semantica per ampliare il sistema formale, creandone un altro, il quale conterrà a sua volta una proposizione indecidibile e quindi richiederà un ulteriore ampliamento, e così via. Solo i sistemi formali molto poveri, che non riescono a formalizzare neppure l'aritmetica, sono coerenti e completi, quelli appena più potenti presentano un'apertura semantica indefinita sul mondo. Ciò contraddice la visione della matematica come tautologia finita, cioè come esplicitazione meccanica di verità teorematice implicite negli assiomi.

Nel 1936 Alan Turing pubblicò *On Computable Numbers*, un articolo in cui illustrava la natura atomica dei passi computazionali mediante l'introduzione della macchina di Turing (MT), protagonista di una proposizione, la tesi di Church-Turing, secondo la quale (ne dò qui una versione semplicemente semplificata; per una discussione e altre versioni si veda Hofstadter) *tutti i processi cerebrali derivano da un substrato calcolabile*. Quindi i processi di pensiero dell'uomo sarebbero riproducibili con un automa finito come la MT. Questa proposizione sta alla base della cosiddetta versione forte dell'IA e trova la sua giustificazione nel fatto che la memoria umana è necessariamente limitata.

Nel 1969, tuttavia, Gödel trovò nel ragionamento di Turing una falla che comprometteva la sua visione finitistica della computazione umana. La mente umana, argomentava Gödel, non è statica, ma si sviluppa di continuo durante il suo funzionamento. Benché ad ogni stadio del suo sviluppo il numero dei suoi possibili stati distinti sia finito, non c'è motivo per cui questo numero non possa divergere nel corso di questo sviluppo. La mente è, all'inizio, basata su un certo numero finito di assiomi (che potrebbero corrispondere alle facoltà iniziali determinate dall'informazione genetica), ma poi grazie all'interazione perturbativa con l'ambiente, cioè all'informazione fornita dall'esperienza, il “sistema mente” si complessifica e sfugge alla nostra analisi formale, così come un sistema formale abbastanza potente, nel senso precisato sopra, sfugge ad ogni riduzionismo di tipo assiomatico.

Ciò che caratterizza la mente umana è la capacità di produrre novità, di rompere le simmetrie, di apprendere, di compiere insomma tutte quelle operazioni creative ben descritte da Gregory Bateson nella sua trattazione delle categorie logiche dell'apprendimento e della

segue



Kurt Gödel (1906-1978)

comunicazione. L'interazione mente-ambiente è generativa e dinamica e non si può racchiudere in una descrizione *definitiva* della mente mediante una o più MT. Mentre le strutture dell'IA presentano una logica chiusa e hanno successo nei micromondi specializzati o per i problemi giocattolo, la mente ha una struttura logica aperta, cambia di continuo le regole del gioco e si adegua alla mutevolezza dell'interazione con il (macro)mondo.

Del resto anche la fisica classica, basata, almeno in apparenza, su una chiusura che si esprime nell'inesorabile procedere dei fenomeni secondo ferree leggi eterne, è incapace di spiegare la novità, la nascita dell'informazione e lo sviluppo della storia. Soltanto con le conquiste rappresentate dalla complessità, dall'emergenza e dagli aspetti informativi, si è potuto fornire un quadro evolutivo, storico e aperto della realtà fisica filtrata dalla nostra cognizione. In ciò si può ravvisare quell'intreccio tra la fisica e la mente che in forma filosofica e qualitativa era stato indicato da Bateson con la sua "ecologia della mente".

Non bisogna tuttavia credere che Turing fosse così ingenuo da ritenere che una MT o una famiglia di MT potesse essere un modello adeguato della mente umana. L'organizzazione e il funzionamento della macchina, sostiene Turing sia nell'articolo del 1936 sia in un rapporto del 1948, si devono poter modificare. La prima modifica funzionale è costituita dagli *oracoli*, dispositivi capaci di condizionare dall'esterno la computazione di una MT. C'è un collegamento tra oracoli e intelligenza naturale: il programma inscritto nel genoma è via via modificato dall'esperienza. La disciplina, dice Turing con riferimento all'organizzazione di partenza e alla capacità di obbedire ai programmi, da sola non basta: per l'intelligenza ci vuole anche l'*iniziativa*. Quindi Turing si rende conto che la MT è un concetto troppo povero per descrivere i processi cognitivi: paradossalmente, la macchina deve essere capace di sbagliare, se è infallibile non può essere intelligente. Del resto, un matematico è sottoposto a un lungo apprendistato, durante il quale in genere commette parecchi errori. Perché non dovrebbe essere la stessa cosa per una macchina?

Questa apertura all'errore consente di evitare ciò che di ripetitivo e, appunto, macchinico si ravvisa in certi programmi che vorrebbero imitare le capacità dialogiche dell'uomo. Mentre gli errori delle macchine nel rispondere alle domande del test di Turing possono essere arbitrariamente madornali, negli errori commessi dagli umani si ravvisa sempre una certa dose di *pertinenza*, nozione più ampia e sfumata della correttezza. La pertinenza insomma comprende anche certi tipi di errore: l'importante è che le risposte sbagliate non diano l'impressione di arbitrarietà o di follia. La nozione di pertinenza può aiutarci a cogliere l'idea (vaga) di comprensione più e meglio della nozione (assoluta) di esattezza (riquadro 5).

Insomma, anche se l'attività superficiale, affiorante, cioè l'espressione linguistica, è la stessa, può darsi che i meccanismi sottostanti siano molto diversi nella macchina e nell'uomo. Ma, conclude Turing, riadeguandosi al comportamentismo, se una macchina gioca bene al gioco dell'imitazione, la differenza nei meccanismi sottostanti può essere trascurata.

4. LA VARIANTE DI TURING

Consideriamo ora la variante proposta da Turing, in cui A è una macchina e B un essere umano. Anche in questo caso non è molto chiaro quale sia il significato del gioco. Si può immaginare che, ripetendo la prova più volte, cioè facendo molte mani, certe volte C sbaglia l'attribuzione e altre volte no, cioè che la macchina a volte superi e a volte fallisca la prova. È intuitivo che, se le domande possono essere arbitrarie, basta porne un numero sufficiente per far cadere la macchina. È sufficiente uscire da quello che è il suo orizzonte di competenza, ponendo domande sul mondo in grande. Ma ciò vale anche per gli umani. Forse, come per tutti i giochi, bisogna porre dei limiti e delle condizioni. Ammesso allora di stabilire durata, modalità e così via, del test, il suo superamento è una condizione sufficiente per l'intelligenza? È una condizione necessaria? O è,

semplicemente, un *gioco d'intelligenza*, un passatempo, al massimo uno spunto di riflessione? Per avere un'indicazione in questo senso, poniamoci domande analoghe a proposito di un essere umano, al quale in genere concediamo senza troppe esitazioni la qualifica di intelligente.

5. L'INTELLIGENZA UMANA

Sottoponiamo un soggetto umano a una prova molto particolare: facciamogli giocare un certo numero di partite a scacchi. Possiamo dire che se vince più della metà delle partite è intelligente? E che se ne vince meno della metà non lo è o lo è meno? Credo che pochi sottoscriverebbero queste conclusioni, al massimo potrebbero concludere che il soggetto gioca bene o male a scacchi, *relativamente al suo avversario*. Ora sottoponiamo il soggetto a un'altra prova, più elusiva: chiediamogli di scrivere un racconto. Anche in questo caso, come nel precedente, la riuscita non è condizione sufficiente per dichiararlo intelligente, ma neppure condizione necessaria. Qui giudicare della riuscita sarebbe peraltro molto più difficile.

Molte persone che sono considerate intelligenti (in base a prove indiziarie, come il successo nella vita professionale o la capacità di intessere rapporti umani soddisfacenti o la

bravura nella ricerca scientifica) non sanno giocare a scacchi e non sanno neppure scrivere un racconto. Forse potrebbero cimentarsi nella stesura di un racconto, ma la riuscita, o l'insuccesso, non aggiungerebbe, o toglierebbe, nulla alla loro intelligenza, sarebbe soltanto una prova indiziaria in più, e solo per chi ritiene importanti i racconti: di nuovo il punto di vista del giudice. Per converso molte persone che giocano bene a scacchi o sanno scrivere ottimi racconti spesso danno prova di scarso acume in molte circostanze della vita.

Il fatto è che l'intelligenza degli umani è per così dire a largo spettro (si veda la citazione di Cartesio nel riquadro 3), è multiforme, elusiva e variegata, si esplica e si accresce e si modifica con la comunicazione e più in generale attraverso quell'attività altrettanto multiforme e inafferrabile che è la vita. Da ciò segue che l'intelligenza umana è molto legata al corpo e alla sua immersione nel contesto

ambientale ed è un prodotto dell'evoluzione prima biologica e poi culturale. Inoltre il numero enorme di esseri umani viventi, operanti e comunicanti in ogni istante rende problematico l'uso di soggetti generici come "l'uomo", "la donna", o di caratteristiche sfumate come "l'intelligenza umana". L'intelligenza delle macchine, qualunque ne sia il significato e la portata è, almeno per il momento, a spettro molto stretto, è finalizzata a compiti particolari e in genere comunica con l'ambiente circostante soltanto attraverso l'esile cordone del programma. Non si estende nell'ambiente mondo né possiede la dimensione diacronica dell'evoluzione.

A rafforzare queste perplessità, si ha l'impressione che nell'articolo Turing consideri la sua proposta più un gioco (dell'imitazione, appunto) che un criterio incontrovertibile per l'intelligenza: solo in seguito, dopo la morte dell'autore, si è cominciato a parlare diffusa-

RIQUADRO 3 - Due precedenti illustri: Cartesio e Leibniz

Debbo a Bianchini due interessanti citazioni, di Cartesio e di Leibniz, che costituiscono anticipazioni profonde e sorprendenti degli argomenti di cui si è occupata l'IA. Il passo di Cartesio espone alcuni criteri differenziali tra intelligenza umana e IA, e in particolare pone l'accento sul comportamento linguistico. Non si può non notarne l'affinità con le argomentazioni di Turing. Leibniz sembra accennare alla possibilità che esistano due tipi di spiegazioni, una per i fenomeni meccanici e una per i fenomeni dell'intelligenza (o dello spirito o della sensazione). Questa doppia spiegazione, che non sembra avere natura ontologica, ma solo epistemologica, ricorda la doppia spiegazione di Bateson in ordine ai fenomeni fisici e ai fenomeni mentali. Presi insieme, i due passi costituiscono la formulazione in nuce di uno dei problemi centrali dell'IA, cioè la valutazione dei modelli di spiegazione dell'azione intelligente.

“Se vi fossero macchine simili ai nostri corpi, che ne imitassero le azioni quanto è praticamente possibile, avremmo sempre due mezzi certissimi per riconoscere che non per questo sarebbero dei veri uomini. Il primo è che mai potrebbero usare delle parole o di altri segni impiegandoli, come noi facciamo, per comunicare ad altri i nostri pensieri. Infatti si può senz'altro concepire una macchina costruita in modo da proferire delle parole e addirittura da proferirne qualcuna a proposito di azioni fisiche che determinino qualche mutamento nei suoi organi: per esempio, toccandola in un punto può domandare che cosa le si vuol dire; toccandola in un altro può gridare che le si fa male, e via di seguito; ma non si può concepire che essa coordini le parole diversamente per rispondere al senso di tutto ciò che si dirà in sua presenza, come possono fare anche gli uomini più ottusi. In secondo luogo, le macchine, anche se facessero parecchie cose bene quanto noi, e forse meglio di tutti noi, immancabilmente in qualche altra cosa fallirebbero, dando modo di scoprire che non agiscono in base a conoscenza, ma solo in base alla disposizione dei loro organi. Perché, mentre la ragione è uno strumento universale, che può servire in ogni sorta di occasioni, questi organi hanno bisogno, in ogni azione particolare, di una disposizione particolare; ne consegue la pratica impossibilità che una macchina ne consenta una sufficiente varietà che le consenta, in tutte le occorrenze della vita, di agire come ci fa agire la nostra ragione”.

Cartesio, *Discorso sul metodo*, Quinta parte

“D'altronde bisogna ammettere che la percezione e quello che ne dipende sono inesplicabili mediante ragioni meccaniche, cioè mediante figure e movimenti. Se immaginiamo una macchina costruita in modo che pensi, senta e percepisca, si potrà concepire che venga ingrandita conservando le medesime proporzioni, in modo che vi si possa entrare come in un mulino. Ciò fatto, nel visitarla internamente, non si troverà altro che pezzi, i quali si spingono scambievolmente, e mai alcuna cosa che possa spiegare una percezione. Cosicché questa bisogna cercarla nella sostanza semplice e non nel composto, o nella macchina. Reciprocamente, non altro che le percezioni e i loro mutamenti si possono rinvenire nella sostanza semplice. In essi soltanto possono consistere tutte le azioni interne delle sostanze semplici”.

Leibniz, *Monadologia*, 17

mente di “test di Turing”. L’articolo presenta un’altra caratteristica: nonostante l’importanza dei temi trattati e l’acume con cui sono affrontati, non ci si può sottrarre del tutto all’impressione che il piglio dell’autore sia, appunto, un tantino giocoso e che ogni tanto egli si lasci andare a qualche divagazione bizzarra: si veda nella nota 1 la risposta data all’obiezione del paranormale.

C’è anche, curiosamente, un errore di calcolo: quando il giudice chiede al soggetto (uomo o macchina?) di sommare 34957 e 70764, la risposta è 105621 invece di 105721. Che significato ha questo errore? O è una svista?

6. LA PROIEZIONE COGNITIVA E LA COSCIENZA

È importante riconsiderare la funzione del giudice. Se, come ho suggerito, è C che viene messo alla prova, forse il test riguarda non tanto l’intelligenza delle macchine quanto l’atteggiamento dell’uomo nei confronti delle manifestazioni (in questo caso linguistiche) dell’intelligenza, sia essa simulata o “naturale”.

Questo rovesciamento di prospettiva è confortato dalle vicende di alcuni programmi redatti in seguito, per esempio *Eliza* di Joseph Weizenbaum (Figura 3) (riquadro 4 a p. 20). In ogni caso non si deve dimenticare che la macchina non simula il pensiero dell’uomo, ma un modello molto semplificato del pensiero dell’uomo.

Nei confronti dei programmi come *Eliza* gli umani attuano una sorta di “proiezione cognitiva” analoga alla proiezione affettiva di cui parlano gli psicologi: la somiglianza tra i segni esteriori di una certa entità E e i segni esteriori di un essere umano U ci spinge a compiere un “prolungamento”, cioè ad attribuire a E tutte le caratteristiche di U, in particolare le caratteristiche psicologiche, affettive e cognitive. Ciò accade a maggior ragione quando anche E è un essere umano, cioè quando possiede *tutti* i segni esteriori dell’umanità: l’operazione proiettiva che si compie in questo caso è tanto frequente che passa inosservata e consiste nell’attribuire all’altro sentimenti, capacità logiche e argomentative e così via, analoghe alle nostre. È solo grazie a questa sorta di immedesimazione che si supera il “problema delle

menti altrui” e si persiste nell’ostinato sforzo di comunicare senza sapere che cosa accada veramente dei nostri messaggi quando entrano nell’interlocutore.

Il problema delle menti altrui è analogo al problema della coscienza: non potendo entrare nell’altra persona, non ho modo di sapere se essa possieda la coscienza e altre caratteristiche, quindi le faccio credito operando a suo favore una fiduciosa concessione delle facoltà che possiedo io stesso. Turing conclude affermando che non si vede il motivo per cui un calcolatore dovrebbe essere trattato diversamente.

Di parere contrario era invece il chimico, filosofo ed economista di origine ungherese Michael Polanyi (1891-1976), il quale osservò che una macchina è una macchina e una mente umana è una mente umana e non c’è dato dell’esperienza che possa cambiare questo *a priori*. Polanyi tuttavia non chiarì la distinzione tra macchina e mente umana, che è il nocciolo di tutta la questione: se restiamo alla superficie, tutto sembra chiaro, ma se spingiamo le argomentazioni a fondo cominciano le difficoltà. Ciò ricorda quanto sant’Agostino diceva del tempo: “Se non me lo chiedi so che cos’è, se me lo chiedi non lo so più”. Questo fenomeno di presbiopia cognitiva riguarda anche la lingua: se ci si accontenta di una certa vaghezza legata al buon senso e all’uso ordinario va tutto bene, ma se si vuol tener conto di tutte le accezioni ed eccezioni linguistiche e definitorie si finisce nei



FIGURA 3
Joseph Weizenbaum (1923-2008)

RIQUADRO 4 - La "psicoterapeuta" Eliza

Il programma *Eliza*, allestito nel 1964 da Joseph Weizenbaum (allora ricercatore del MIT), trae il nome dalla protagonista femminile del dramma *Pygmalion* di George Bernard Shaw, la quale diviene oggetto della scommessa del professor Higgins: si può trasformare una sciatta popolana in una raffinata dama di corte agendo su fonetica e linguaggio? Infatti *Eliza*, che consente alle persone di conversare in linguaggio naturale con un computer, può essere via via migliorato da un "insegnante" che fa la parte di Higgins. Il dialogo tra computer e utente avviene tramite telescrivente: l'utente batte una frase e passa la parola a *Eliza*, che analizza la frase e risponde battendo una frase. Ecco un frammento di dialogo tra *Eliza* e una paziente (in corsivo le frasi della paziente):

Gli uomini sono tutti uguali.
In che senso?
In un modo o nell'altro ci tormentano sempre.
Può farmi un esempio?
Be', è stato il mio ragazzo a farmi venire qui.
È stato il suo ragazzo a farla venire qui.
Dice che sono sempre depressa.
Mi dispiace sentire che è depressa.
È vero. Sono infelice.
Pensa che venire qui l'aiuterà a non essere infelice?
Ho bisogno di aiuto, questo è sicuro.
Che cosa cambierebbe per lei se ricevesse aiuto?
Forse imparerei a sopportare mia madre.
Mi dica qualcosa della sua famiglia.
Mia madre si occupa di me.
Chi altri della sua famiglia si occupa di lei?
Mio padre.
Suo padre.
Per certi versi lei somiglia a mio padre.
Che somiglianze vede?
Lei non è molto aggressivo, ma penso che non voglia farmelo capire.
Perché pensa che non sia molto aggressivo?
Lei non litiga con me.
Perché pensa che io non litighi con lei?
Perché ha paura di me.
Le fa piacere pensare che ho paura dei lei?
Mio padre ha paura di tutti.
Che cos'altro le viene in mente quando pensa a suo padre?
I prepotenti.
Questo ha a che fare con il fatto che il suo ragazzo l'ha fatta venire qui?

In sintesi, il procedimento seguito dal programma è questo: *Eliza* cerca nella frase della paziente qualche parola chiave; se la trova, trasforma la frase secondo una regola associata a quella parola chiave, altrimenti fa un'osservazione acotestuale o, in certe condizioni, ripete una trasformazione precedente (in realtà il programma è molto più complesso, ma non è il caso di entrare nei particolari). *Eliza* dà all'interlocutore l'impressione di essere uno psicoterapeuta non direttivo (rogeriano), il quale è comprensivo ma piuttosto inerte: egli pone domande vaghe e parla pochissimo, come si vede dal colloquio riportato, e per un po' risulta piuttosto convincente. In *Eliza* non c'è nessun barlume d'intelligenza: il programma va semplicemente in cerca di parole come "madre" o "depresso" e poi pesca una domanda appropriata da un repertorio; se questa tattica non funziona, produce una frase generica tentando di riavviare la conversazione. Quasi tutti i programmi dialogici si basano su principi analoghi (riquadro 5).

In realtà quello che sembra un dialogo è un monologo dell'interlocutore umano, il quale dà senso a domande e risposte solo sue, comportandosi da "animista". *Eliza* ebbe un enorme successo: chi conversava con "lei" provava spesso sollievo dopo le sedute. Uno psichiatra, Kenneth Colby, dichiarò che di lì a pochi anni sarebbero stati allestiti programmi da usare nella pratica terapeutica, e a sua volta allestì un programma, *Parry*, che simulava il comportamento linguistico di un paranoico e riuscì a ingannare parecchi psichiatri sulla propria identità. È da notare che è molto più facile simulare uno psicotico che un savio. Fu anche organizzato un dialogo tra *Eliza* e *Parry*! Weizenbaum, tuttavia, fu molto turbato dal successo del suo programma. Credo che *Eliza* sia stato determinante nella sua decisione di schierarsi contro l'intelligenza artificiale.

guai. Forse bisognerebbe appagarsi di una certa approssimazione, senza cercare una rischiosa precisione. Chi si contenta gode e chi troppo vuole nulla stringe.

Di parere contrario era anche il grande neurofisiologo e neurochirurgo britannico sir Geof-

frey Jefferson (1886-1961), di cui Turing riporta la seguente obiezione:

"Fino a quando una macchina non saprà scrivere un sonetto o comporre un concerto sulla base di pensieri e di emozioni sentite

come tali e non per una casuale combinazione di simboli - cioè finché non solo possa farlo, ma sappia di averlo fatto - noi non accetteremo mai l'idea che la macchina possa uguagliare il cervello. Non c'è meccanismo capace di provare (e non semplicemente segnalare artificialmente, il che sarebbe facile da ottenere) piacere per i successi conseguiti, dolore quando le sue valvole bruciano, gioia quando riceve complimenti, tristezza per gli errori commessi, l'incanto del sesso, la collera o la delusione perché non riesce ad ottenere ciò che desidera”.

A ciò Turing risponde:

“Questo ragionamento sembra confutare la validità del nostro test. Secondo la versione più radicale di questa posizione, l'unico modo per essere sicuri che una macchina pensa è di essere quella macchina e sentirsi pensare. Si potrebbe allora descrivere questa sensazione al mondo, ma naturalmente nessuno sarebbe tenuto a prenderne atto. Allo stesso modo, secondo questa concezione, l'unico modo per sapere che un dato uomo pensa è di essere quell'uomo. Si tratta in effetti della posizione solipsistica. Può essere la posizione più logica da sostenere, ma rende difficile la comunicazione delle idee. Si crea questa situazione: A è convinto che “A pensa ma B no”, mentre B crede che “B pensa ma A no”. Invece di continuare ad accapigliarsi su questo punto, di solito per educazione si adotta la convenzione che tutti pensino”.

E aggiunge che il gioco dell'imitazione con un solo interrogato è spesso usato nella pratica degli esami orali per accertarsi se lo studente abbia imparato qualcosa a pappagallo o l'abbia davvero capita. Se poi una macchina che scrive sonetti superasse un esame orale sulla composizione poetica con la stessa disinvoltura e la stessa capacità di giustificare le risposte che manifesta uno studente, sarebbe difficile sostenere che la macchina esegue soltanto delle segnalazioni artificiali o che si tratta di un facile trucco. Insomma una prestazione verbale che non si distinguesse da quella di un essere umano spingerebbe ad accettare la va-

lidità del test. Quanto alla coscienza, Turing aggiunge tuttavia:

“Non vorrei dare l'impressione che riguardo alla coscienza non ci sia nessun mistero. Per esempio qualunque tentativo di localizzarla porta a qualche forma di paradosso. Tuttavia non penso che sia necessario risolvere questi misteri prima di poter dare una risposta ai problemi di cui ci occupiamo in questo articolo”.

7. IL RUOLO DEL LINGUAGGIO VERBALE

“E invero tutte le qualità che un accorto novellatore di razza umana, esperto quanto si voglia di caratteri, può rilevare in un animale o attribuirgli, non sono al postutto che mere supposizioni, cui solo il nostro smodato antropomorfismo presta verosimiglianza. Fra noi: in che modo penetrare d'un brutto i pensieri, il vero significato dei suoi gesti, anche ad adottare l'accezione umana di tali termini? Un uomo di fronte a un altro uomo ha almeno una convenzione, se non altro di linguaggio, alla cui stregua commisurare gli attributi; ma riportare questa convenzione sugli animali sarebbe a dir poco arbitrario”

Tommaso Landolfi, *Le due zittelle*

Secondo alcuni il criterio verbale di Turing non è sufficiente per l'attribuzione del pensiero e neppure necessario: un ente potrebbe essere intelligente senza essere in grado di parlare, e in effetti ciò si può dire degli animali superiori. La comunicazione, anche negli umani, non è certo solo linguistica, ma noi occidentali abbiamo privilegiato la comunicazione verbale e consideriamo le altre forme di comunicazione vaghe e ambigue, ma forse è solo questione di tradizione culturale. Ciò che conta è poter comunicare in qualche modo: un'entità intelligente che non potesse comunicare affatto sarebbe come l'albero di Berkeley, che cade nella foresta senza che nessuno ne oda lo schianto. Il criterio verbale è molto antropocentrico, tuttavia è abbastanza evidente che certi uomini non supererebbero il test, mentre certe macchine non intelligenti potrebbero superarlo (inducendo

in errore almeno un giudice per cinque minuti, come nel caso di *Eliza*). Ciò sembra indebolire di molto la portata del criterio. Il tentativo di separare la mente dal corpo, attribuendo l'intelligenza solo alla prima, solleva dunque un certo numero di problemi, che sono tipici dell'impostazione funzionalistica dell'intelligenza artificiale e che hanno portato al fallimento della sua versione forte.

Questi problemi si possono riassumere mediante alcune questioni cruciali:

Può esistere un'intelligenza senza vita e senza altri correlati della vita, per esempio le emozioni e la coscienza? Può esistere una mente senza interazione comunicativa con l'altro? Può esistere un linguaggio autoreferenziale, senza qualche forma di esistenza concreta e quindi di interazione con il mondo? E, forse la più cruciale: può esistere il pensiero senza l'esperienza (che è radicata nel corpo)?

Queste domande denunciano una chiarissima impostazione antropocentrica: noi scegliamo come modello di riferimento la nostra attività cognitiva e confrontiamo con essa ogni alternativa proposta, con il preconcetto che sia inferiore o difettosa. Allo stesso tempo, nelle domande compaiono termini indefiniti, vaghi, imprecisi, come vita, coscienza, esperienza, ai quali tuttavia noi *esseri umani* sappiamo dare significato operativo e di fronte ai quali adottiamo un atteggiamento di sospensione definitoria e di operosità intuitiva. Il trasferimento di questi concetti alla macchina, che non può beneficiare dell'impostazione operativa o dell'intuizione, richiederebbe invece definizioni univoche e inattaccabili. È evidente che le macchine si trovano in condizione d'inferiorità.

Non dimentichiamo che anche Ludwig Wittgenstein si poneva problemi analoghi: la lingua può essere considerata solo un gioco? Oppure ha una relazione con la vita reale? È abbastanza curioso che ancora ci siano pensatori orientati a rispondere in modo positivo alla prima domanda, dimenticando la base concreta della lingua e l'inestricabile groviglio che la lega al mondo attraverso l'attività dell'uomo. Anche le conseguenze dei teoremi di Gödel, che indicano un'apertura indefinita della matematica e una sua connessione

con il mondo dovrebbero rendere evidente l'analogia apertura sul mondo della lingua e della mente (vedi riquadro 2).

Turing oscilla tra due posizioni: una è a favore della smaterializzazione dell'intelligenza, bene espressa dal protagonista del libro *The Small Back Room* di Nigel Balchin: *“È un gran peccato che non si possano abolire la marina, l'esercito e l'aviazione e continuare semplicemente a vincere la guerra senza di loro”*. L'altra invece, illustrata da Turing stesso nell'articolo, riconosce l'importanza del legame con la realtà per la completezza della macchina. Dopo aver elencato vari difetti che, manifestati dalla macchina, la renderebbero per alcuni incapace di pensare: essere gentile, piena di risorse, bella, cordiale, far sì che qualcuno s'innamori di lei, gustare le fragole con la panna, Turing scrive:

“L'incapacità di gustare le fragole con la panna può sembrare frivola al lettore. Forse sarebbe possibile trovare il modo di far gustare a una macchina questo piatto delizioso, ma uno sforzo del genere sarebbe insensato. La cosa importante riguardo a questa incapacità è che essa compromette per esempio la possibilità che si stabilisca tra uomo e macchina lo stesso genere di cameratismo che può esistere tra due bianchi oppure tra due neri”.

Per concludere, si può forse dire che il cosiddetto test di Turing è un esperimento concettuale il cui valore pratico, quanto alla presenza dell'intelligenza nelle macchine, è quasi nullo. Non rappresenta una condizione necessaria e neanche una condizione sufficiente. La sua importanza risiede piuttosto nel tipo di riflessioni che ha provocato sul concetto di intelligenza, ma anche di coscienza, di computazione e via dicendo. Le stesse esitazioni di Turing tra la concezione sintattica dell'intelligenza e il suo radicamento nel corpo e nel mondo sono utili per apprezzare i limiti dell'impostazione forte dell'IA. L'ottimismo di Turing sul futuro dell'IA si è dimostrato ingiustificato, o meglio mal diretto: oggi, nel 2009, le macchine fanno cose straordinarie, ma non fanno ciò che Turing aveva immaginato (riquadro 5 a p. 23). Esse si presentano sempre più come componenti essen-

RIQUADRO 5 - Il premio Loebner

Il test di Turing, almeno a prima vista, è di una semplicità disarmante, ma per alcuni decenni nessuno pensò di attuarlo: come aveva fatto Turing, tutti lo consideravano solo un esperimento concettuale. È vero che nel suo articolo del 1950 Turing aveva pronosticato per le macchine un avvenire luminoso, sia pure in termini piuttosto vaghi:

“io credo che tra una cinquantina d’anni sarà possibile programmare calcolatori aventi una capacità di memoria di circa un miliardo, in modo da farli giocare così bene al gioco dell’imitazione che un esaminatore medio avrà una probabilità non superiore al 70% di compiere l’identificazione giusta dopo cinque minuti di interrogatorio. Credo che la domanda iniziale - le macchine possono pensare? - sia troppo priva di senso per meritare una discussione. Ciò nonostante credo che alla fine del secolo l’uso delle parole e l’uso corrente saranno talmente mutati che si potrà parlare di macchine pensanti senza aspettarsi di essere contraddetti”.

Ma evidentemente i ricercatori successivi erano stati meno ottimisti e non si erano cimentati con il test. Nel 1991, tuttavia, Hugh Loebner, programmatore e uomo d'affari del New Jersey, istituì un premio di 100.000 dollari da assegnare al primo computer che superasse il test. Ritenendo piuttosto remoto questo evento, Loebner decise di assegnare un premio annuo d’incoraggiamento di 2000 dollari al concorrente migliore. Da allora la competizione è seguita con grande interesse, per quanto nessun computer si sia mai avvicinato neppure lontanamente al traguardo, sicché il concorso si è più o meno trasformato in un evento pubblicitario. Nel 1996 la gara riservò una sorpresa: tra tutte le macchine pensanti, o candidate a pensare, si era intrufolato un impostore, il programma HeX, scritto da Jason Hutchens, un dottorando dell’University of Western Australia. E, tra l’indignazione generale, HeX vinse. Dopo la proclamazione, Hutchens dichiarò che il suo programma non era più intelligente di un distributore automatico di bevande: HeX si basava su un dizionarietto di domande, risposte e frasette disinvolte, che venivano emesse quando il programma riconosceva certe parole (vedi riquadro 4). Hutchens non aveva fatto ricorso a nessuna tecnica IA e anzi le aveva evitate per dimostrare che, in rapporto all’IA, il concorso Loebner era una pura perdita di tempo. Benché in seguito avesse dichiarato di aver partecipato solo per ottenere “gloria e denaro”, non si può neppure sostenere che lo studente avesse imbrogliato, perché il premio Loebner non ha nessuna regola precisa sul contenuto o sugli intenti dei partecipanti (questa vaghezza, in fondo, riflette la vaghezza del test di Turing). Nonostante il ricorso alle tecniche IA più raffinate, non è che i concorrenti battuti avessero scritto programmi molto brillanti. Ecco uno scampolo di conversazione tra l’esaminatore e il programma Fred 16.43:

Fred: Salve esaminatore 4

Esam: Ciao. Com’è il tempo fuori?

Fred: Non sono stato molto fuori oggi. A me piacciono gli asparagi.

Esam: Quanti anni hai?

Fred: Circa trenta.

Esam: Circa?

Fred: Potrei parlare di Star Trek.

Esam: No, parliamo di te. Quanti anni hai, di preciso?

Fred: Ah, tu sei Batman.

HeX non era molto meglio. Ecco un suo scampolo:

HeX: Vado all’University of Western Australia, che è una delle più belle università d’Australia. È situata sul Fiume dei Cigni, a Perth.

Esam: Che significa “situata”?

HeX. Non ho idea di che cosa “situata” significhi essere.

Esam: Ehm... sei stanco?

HeX: Tutti sanno che sono stanco. E non faccio spesso lo spaccone. Che cosa pensa di quest’affare del concorso Loebner?

Esam: Sta diventando una cosa spassosa, no?

HeX: Eh eh eh... questa sarebbe una “domanda”!? Ha visto qualche bel film di recente?

Si consideri la conversazione che segue: il concorrente è un computer o un essere umano (test di Turing per il lettore)

Esam: Siamo in aprile o in maggio?

Conc: Ehm... credo in aprile.

Esam: Che cosa te lo fa credere?

Conc: Questa è un’ottima domanda. Perché i mesi e i giorni e così via sono arbitrari... credo.

Esam: Mi sembra di notare in te una sfumatura di agitazione. Una sfumatura umana.

Conc: Davvero? Anche lei però è un po’ reticente.

Esam: Non ho mai detto che tu sei reticente.

Conc: Lo so.

Esam: Sei compiaciuto.

Conc: Moi?

Esam: Penso che tu sia un essere umano.

Dato il livello di questi programmi, non ha molta importanza che Hutchens abbia imbrogliato. Dai brani riportati si vede quanto i computer siano ancora lontani dall’intelligenza umana: si ha l’impressione che tutte le conversazioni siano sconnesse, assurde, non *pertinenti*, appunto artificiali. Non c’è dubbio che col tempo le prestazioni miglioreranno, ma finché i programmi resteranno chiusi nel loro mondo virtuale non potranno rispondere a domande sul caldo o sul Fiume dei Cigni. Bisogna aver immerso la mano nell’acqua, aver boccheggionato di sete, essere stati lì per annegare. Turing ha basato il suo gioco solo sulla lingua, cercando di isolare la “purezza” astratta del pensiero, e ciò comporta che si possa smascherare un’intelligenza che non abbia l’esperienza del mondo: basta porre un numero sufficiente di domande ed ecco emergere la differenza *qualitativa* tra l’uomo e il computer, che deriva essenzialmente dalla formazione evolutiva e dall’immersione sistemica dell’uomo, cioè dalla presenza del corpo. In queste condizioni si rileva tutta l’inadeguatezza della locuzione “intelligenza artificiale”: l’accento andrebbe posto molto più sull’aggettivo che sul sostantivo.

