

Sul problema dell'autocoscienza

R. Nobili - Università degli Studi di Padova

Materiale preparato nell'occasione di alcuni dibattiti scientifico-filosofici tenuti a Padova all'inizio del 2009 presso l'Aloisianum (Istituto di Studi Filosofici della Compagnia di Gesù) volti a chiarire le tesi sostenute dall'autore nell'articolo "La Macchina della Mente I".

Indice dei paragrafi

1. *Introduzione*
2. *Incertezza e informazione*
3. *Informazione e coscienza*
4. *I "possibili" nella fisica*
5. *Il correlato neurologico dell'intuizione del possibile*
6. *La macchina della mente*
7. *La complessità secondo Kolmogorov*
8. *La complessità secondo von Neumann*
9. *Un esempio di complessità indescrivibile: l'insieme di Mandelbrot*
10. *Macchine ricorsive e processi ricorsivi*
11. *Processi sequenziali e processi paralleli*
12. *Le macchine logico-aritmetiche e la nozione di calcolatore universale*
13. *Le macchine costruttrici e la nozione di universalità costruttiva*
14. *Capacità autoreferenziale del formalismo logico-aritmetico*
15. *Capacità autoreferenziale del linguaggio logico*
16. *Capacità autoriproduttiva dei costruttori universali*
17. *Capacità autoreferenziale del pensiero umano*
18. *Le prospettive dell'intelligenza artificiale*

1. Introduzione

Circa il problema dell'autocoscienza, di come questa possa definirsi e spiegarsi, io distinguerei due aspetti, uno sostanziale, l'altro formale.

Per aspetto sostanziale intendo il fenomeno dell'autocoscienza come manifestazione della realtà soggettiva interiore; quel sapere di sé, in rapporto ad un mondo esterno, che si presenta in ogni istante come una determinazione specifica entro l'insieme delle possibilità esistenziali e comportamentali intuite dal soggetto pensante in concomitanza con gli eventi della percezione.

Per aspetto formale intendo la rappresentazione dell'attività autoriflessiva del pensiero intesa come processo di elaborazione ricorsiva dell'informazione cerebrale. Questo riguarda, in ultima analisi, il problema di come possa funzionare, in termini informatici, la capacità di pensare il pensato e il processo stesso del pensiero. Dal punto di vista della logica formale, si tratta di spiegare come una macchina capace di elaborare e interpretare i messaggi di un linguaggio, inteso come un sistema di comunicazione basato su rappresentazioni simboliche, possa elaborare ed interpretare (in termini comunicazionali e comportamentali) anche i messaggi generati all'interno dalla macchina stessa, ad esempio quelli che descrivono i procedimenti generali usati dalla macchina nello svolgimento delle sue funzioni.

2. Incertezza e informazione

Per cercare di porre bene il problema farò una digressione su una materia che potrà sembrare piuttosto tecnica, ma che in realtà è di grande interesse scientifico e filosofico. Nel 1949, Claude Shannon pubblicò la *Teoria della Comunicazione*¹ basando la nozione d'informazione su quella di *entropia informazionale*: una grandezza puramente matematica dotata di significato qualitativo e quantitativo del tutto simile a quella che il fisico austriaco Ludwig Boltzmann introdusse nella meccanica statistica alla fine dell'800 per calcolare l'entropia di un sistema termodinamico. La misura dell'entropia informazionale è espressa dal logaritmo medio delle probabilità di occorrenza di un insieme completo di messaggi reciprocamente esclusivi che una sorgente d'informazione² può inviare ad un destinatario reale o potenziale. La caratteristica più importante di questa formula logaritmica è di essere completamente individuata dalle seguenti proprietà: *i*) essere una funzione delle sole probabilità; *ii*) essere additiva per sorgenti di messaggi indipendenti; *iii*) essere massima quando i messaggi sono equiprobabili; *iv*) annullarsi quando un messaggio è certo mentre gli altri sono impossibili; *v*) assumere il valore di 1 *bit* (*binary digit*; it. cifra binaria) per due messaggi alternativi equiprobabili.

Grazie a queste proprietà l'entropia informazionale riesce a rappresentare nel modo più soddisfacente l'*incertezza* con cui i messaggi sono mediamente attesi dal destinatario. Sulla base di questa nozione Shannon poté quindi definire la *quantità d'informazione* ricevuta dal destinatario come la *diminuzione di entropia informazionale* (neg-entropia informazionale) che consegue all'acquisizione del messaggio, vale a dire come la differenza tra l'entropia prima del messaggio e quella dopo il messaggio. Chiaramente, se il messaggio ricevuto elimina tutta l'incertezza, dopo la ricezione l'entropia è zero e pertanto la misura dell'informazione acquisita uguaglia quella dell'incertezza di attesa del messaggio.

La portata di tale nozione va oltre il dominio della teoria della comunicazione. Ciò che vale per i messaggi può essere riferito ad un qualsiasi altro insieme completo di eventi reciprocamente esclusivi. Così, ad esempio, il lancio di una moneta produce 1 bit d'informazione in quanto riduce a zero l'incertezza associata all'alternativa TESTA/CROCE. Appare dunque chiaro che, sia che si tratti di messaggi attesi da un destinatario sia di eventi naturali attesi da un osservatore, l'*informazione*, anche in senso qualitativo, può essere definita come *riduzione di uno stato iniziale d'incertezza*.

3. Informazione e coscienza

Tornando alla distinzione posta nell'introduzione, si può notare che un "momento di autocoscienza", nel suo aspetto sostanziale, si presenta come un *evento informazionale* nel senso shannoniano del termine. In tale "momento", infatti, si può riconoscere sia un'incertezza aprioristica sia una riduzione d'incertezza. L'incertezza aprioristica è relativa all'insieme delle possibilità esistenziali attese dal soggetto pensante; in altri termini, lo spazio degli eventi immaginabili dall'individuo cosciente nel quale si pone ad ogni istante la cognizione di ciò che realmente accade. La riduzione d'incertezza è invece la restrizione di tale insieme alla pura e semplice percezione di quella, tra le infinite possibilità, che di fatto si dà nell'istante presente; in altri termini, la determinazione di un significato particolare all'interno dell'insieme di possibilità che definisce e delimita l'incertezza iniziale. Incidentalmente, dato che in ogni momento di autocoscienza la certezza di esistere ha un contenuto d'informazione solo in rapporto al dubbio di esistere, la celebre proposizione cartesiana *cogito ergo sum* si presenta formalmente equivalente alla proposizione *dubito ergo sum*.

Ora, è proprio l'analisi di questo aspetto sostanziale che presenta le maggiori difficoltà. Come nasce l'intuizione del possibile? Cos'è, nella sua essenza, il sentimento d'incertezza associato a questa intuizione? Qual è il correlato neurologico di questa incertezza? Quale quello fisico? Queste sono, a

¹ Shannon, C.E. and Weaver, W. (1949) *Mathematical Theory of Communication*, Illinois University Press, Urbana.

² Si definisce *entropia informazionale* di un insieme completo di eventi reciprocamente esclusivi, attesi con probabilità p_1, p_2, \dots, p_n , la quantità $H = p_1 \log_2(1/p_1) + p_2 \log_2(1/p_2) + \dots + p_n \log_2(1/p_n)$. C.Shannon, (1948) A Mathematical Theory of Communication, *Bell System Technical Journal*, 27:379-423, 623-656. Questo articolo è facilmente reperibile in rete.

mio avviso, le domande cruciali che incidono maggiormente sulla possibilità di soluzione del problema.

4. I “possibili” nella fisica

La difficoltà maggiore nasce proprio in seno alla fisica, dato che la nozione di “possibilità”, e di conseguenza anche quella d’incertezza, si pone in modi profondamente diversi nella fisica classica e in quella quantistica.

Nella fisica classica, secondo la quale il mondo è governato da leggi deterministiche, il termine “incertezza” riveste un significato *soggettivo assoluto*, del tutto equivalente a “stato d’ignoranza dell’osservatore”. Nella quantistica invece, secondo la quale il mondo è governato da leggi probabilistiche dipendenti dai modi di osservazione, lo stesso termine riveste primariamente un significato *oggettivo relativo*, del tutto equivalente a “stato d’indeterminazione intrinseca dei fenomeni naturali rispetto agli apparati di osservazione”.

E’ tuttavia importante osservare che, mentre nella fisica classica si ritiene che gli stati fisici possano essere osservati senza subire perturbazioni - cosicché si può affermare che i fatti che si osservano sono esattamente quelli che accadrebbero anche in assenza di osservazioni - nella quantistica, invece, le perturbazioni degli stati conseguenti all’interazione osservato-osservatore incidono sulla catena causale degli eventi naturali rendendo più o meno imprevedibili i risultati delle osservazioni. In questo caso, la conoscenza precisa di uno stato fisico ad un dato istante non consente, in generale, di prevedere con esattezza cosa sarà effettivamente osservato subito dopo tale istante.

La conoscenza precisa dello stato ci consente solo di prevedere quale ventaglio di possibili fenomeni potranno essere osservati con un determinato apparato di osservazione e quali sono le probabilità di tali fenomeni. In altri termini, in ogni previsione della fisica quantistica c’è un margine più o meno grande d’incertezza, talvolta anche nullo, non tanto su quello che potrà essere osservato, ma su quello che sarà effettivamente osservato. Ovviamente, questo margine d’incertezza svanisce nel momento stesso in cui l’apparato di osservazione rivela quale, tra i possibili fenomeni attesi, si è realmente manifestato. Così, in generale, in assenza di osservazioni la corrispondenza tra stato e fenomeno verrà meno. Ma essa si ristabilirà prontamente nell’istante stesso in cui avverrà l’osservazione. Si potrà perciò ancora affermare che subito dopo il processo di osservazione lo stato è esattamente quello che corrisponde al fenomeno osservato. Questo può essere confermato eseguendo un secondo processo di osservazione immediatamente dopo il primo. Ma poco dopo, in seguito all’evoluzione del sistema, la corrispondenza verrà perduta e la conoscenza del nuovo stato prodotto dal processo di osservazione non ci permetterà di prevedere con precisione cosa verrà realmente osservato dopo un tempo non immediatamente successivo.

Più in generale, nel corso dell’evoluzione naturale, in corrispondenza del manifestarsi degli eventi fisici, si aprono continuamente nuovi ventagli di possibilità osservazionali che di volta in volta introducono nuovi margini d’incertezza. Questi ventagli si restringono, o si riducono alla singola possibilità che si è realmente manifestata, nell’istante stesso in cui sono effettuate nuove osservazioni. Ciò determina, di volta in volta, una diminuzione di entropia e di conseguenza la produzione di ulteriori elementi d’informazione.

La meccanica quantistica insegna, infatti, proprio questo: nel momento stesso in cui l’apparato di osservazione cattura l’informazione, lo stato fisico immediatamente precedente all’osservazione subisce una trasformazione improvvisa che lo “mette in forma” nel modo esattamente corrispondente al fenomeno osservato. Ciò significa che i processi di osservazione quantistici sono eventi informativi in un duplice senso: sia soggettivamente, come produzioni d’informazione per gli osservatori, sia oggettivamente, come produzioni di nuovi stati fisici. Si stabilisce così una corrispondenza biunivoca tra l’incertezza di aspettazione degli eventi naturali, che possiamo identificare con l’*entropia informazionale*, e l’indeterminazione oggettiva delle grandezze osservabili,

che in ultima analisi può essere identificata con l'*entropia termodinamica*. Su questa base, l'intera storia dell'universo può essere interpretata come la realizzazione di un duplice processo: da un lato una produzione d'informazione per soggetti osservatori reali o potenziali - i quali, si badi, possono esistere solo come parti macroscopiche di un universo termodinamico fuori equilibrio - dall'altro un processo termodinamico che trasforma irreversibilmente gli stati macroscopici della materia, che funge così come la memoria fisica infinita dell'informazione cosmologicamente realizzata.

In questa visione del mondo fisico l'accadimento dei fatti naturali si presenta come un'immensa produzione di fenomeni, di cui è chiaramente presupposta l'osservabilità. Possiamo così dare un senso al *principio del parallelismo psicofisico* postulato da von Neumann (1932)³, secondo il quale tutti gli eventi informativi che sono consapevolmente esperiti dagli ego astratti dei soggetti pensanti sono concomitanti ad eventi oggettivi che modificano gli stati fisici dei soggetti stessi. Ne segue che tutti i processi mentali che contribuiscono a generare l'autocoscienza, e prima ancora l'intuizione soggettiva del possibile, possono essere interpretati come processi naturali.

5. Il correlato neurologico dell'intuizione del possibile

Sulla scorta di queste considerazioni, la soluzione scientifica al problema che abbiamo posto nell'introduzione potrebbe sembrare meno ardua di quanto si poteva sospettare. Assumendo che l'intuizione soggettiva del possibile abbia un preciso correlato fisico-neurologico, saremmo portati in ultima analisi ad ipotizzare che l'aspetto sostanziale e quello formale siano, per così dire, due facce di una stessa medaglia. Potremmo allora sperare che la comprensione del secondo possa giovare a quella del primo. Devo però confessare che alla domanda di quale siano i correlati neurologici dell'intuizione o del sentimento del possibile non so dare una risposta precisa. Anzi mi chiedo se le conoscenze scientifiche disponibili allo stato attuale permettano di porre bene questo problema. Per queste ragioni, eviterò di speculare più oltre sull'aspetto sostanziale e sulla sua possibile relazione con quello formale e limiterò le mie considerazioni soltanto al secondo.

Tuttavia, anche nel merito dell'aspetto formale è necessario restringere di campo. Basterebbe la sola lettura dei libri di Antonio Damasio⁴ per disporsi allo studio di questo aspetto con grande prudenza. Riguardo il problema di come funziona la mente in relazione ai correlati neurologici normali o patologici del cervello, Damasio spiega in modo circostanziato e convincente come sia necessario tenere conto del fatto che la cosiddetta "coscienza" si presenta all'analisi clinica come un complesso stratificato di livelli. Si procede da un livello più profondo e primitivo, la cui sede neurologica sta nelle regioni ventro-mediali della corteccia cerebrale e in vari nuclei del tronco encefalico - che i neurologi identificano come i centri generatori di *emozioni e sentimenti*⁵ - via via fino alla coscienza cognitiva, nei suoi diversi modi e gradi di sviluppo, e infine alla coscienza autobiografica, che è certamente assai più sviluppata negli esseri umani che negli altri animali. Tutti questi livelli interagiscono attraverso una complessa gerarchia di circuiti neuronali riverberanti, che fanno supporre l'esistenza di un complesso sistema di processi ricorsivi a tutti questi livelli corrispondono altrettante forme di memoria cognitiva.

Purtroppo, a tutti i livelli indicati dal celebre neurologo portoghese si pongono problemi di caratterizzazione e funzionamento del cervello che l'indagine neurologica non ha ancora saputo comprendere bene e risolvere; senza dire che in realtà non sappiamo ancora abbastanza bene quali siano i principi generali e i processi neuronali su cui si basa la straordinaria potenza di elaborazione d'informazione del cervello.

³ Von Neumann, (1932) *Mathematical Foundations of Quantum Mechanics*, Princeton University Press.

⁴ A. Damasio - *L'errore di Cartesio. Emozioni, sentimenti e cervello*. (1995); *Emozione e coscienza* (2000); *Alla ricerca di Spinoza. Emozioni, sentimenti e cervello* (2003). Adelphi Ed.

⁵ Queste regioni ricevono e trasmettono l'informazione che riguarda lo stato fisiologico di tutti gli organi del corpo (*informazione somatica*).

6. *La macchina della mente*

Data l'impossibilità di chiarire l'aspetto sostanziale nell'ambito delle scienze neurologiche, limiterò le mie considerazioni ad un problema meno complesso, ma non per questo meno impegnativo: quello di come una macchina elaboratrice d'informazione possa generare un processo *autoriflessivo*. Alcuni studiosi d'intelligenza artificiale, in prima fila Douglas Richard Hofstadter (1984, 2009)⁶, ritengono che questo problema sia formalmente simile a quello di come un sistema logico-matematico sufficientemente complesso riesca ad ammettere una relazione interna di tipo autoreferenziale. L'individuazione di questo problema e la sua soluzione si possono ricondurre alle indagini di Kurt Gödel, Alonzo Church, Alan Turing, Alfred Tarski, John von Neumann e altri sommi logici e matematici.

Purtroppo, una trattazione abbastanza completa di questo problema ed un'analisi delle sue implicazioni e conseguenze non sembra essersi ancora affermata e diffusa nella letteratura scientifica e filosofica contemporanea. In questo scritto cercherò di dare un contributo alla divulgazione di questo affascinante argomento cercando di mettere in evidenza e collegare tra loro, secondo la mia visione personale, alcuni capisaldi concettuali della materia. Elencherò qui di seguito, e poi descriverò succintamente, alcune definizioni e nozioni matematiche che mi sembrano necessarie alla comprensione del problema e della sua soluzione. Le descriverò così come si sono presentate con sufficiente chiarezza alla mia attenzione nel corso degli anni, seguendo un ordine logico anziché cronologico.

7. *La complessità secondo Kolmogorov*

La complessità di un oggetto matematico, finito o infinito, fu definita da Kolmogorov come la lunghezza del più breve programma capace d'istruire una macchina calcolatrice a produrre tale oggetto mediante le operazioni di un algoritmo⁷. Per questa ragione essa è chiamata *complessità algoritmica*. Ad esempio, la complessità algoritmica di una sequenza del tipo 000..., oppure 111..., è piccolissima dato che un programma atto a istruirne la produzione ha all'incirca la forma: "scrivi 0, oppure 1, e ripeti questa operazione infinite volte". La complessità algoritmica di $\pi = 3.1456\dots$, il numero reale che rappresenta il rapporto tra la circonferenza e il diametro di un cerchio, è un po' più grande, dato che richiede, ad esempio, il programma che istruisce il calcolo di tale numero mediante una procedura iterativa, ad esempio quella descritta dalla formula di Leibniz: $\pi = 4/1 - 4/3 + 4/5 - 4/7 + \dots$. Per contrasto, una sequenza casuale infinita di 0 e 1, quale potrebbe ottenersi lanciando ripetutamente una moneta con 0 e 1 al posto di testa e croce, possiede una complessità algoritmica infinita, poiché non esiste alcun modo di produrre una sequenza di questo genere impartendo ad un calcolatore un'istruzione finita.

Chiaramente, questa nozione di complessità appare alquanto ambigua dato che è relativa al tipo di macchina che produce gli oggetti e dal linguaggio di programmazione prescelto. La sua sensatezza dipende dal fatto che la differenza tra le lunghezze dei programmi usati da due macchine diverse, o istruiti in due linguaggi di programmazione diversi - per generare lo stesso oggetto - è circa proporzionale al logaritmo medio di tali lunghezze. Poiché il logaritmo di un numero n cresce con n molto più lentamente di n , tale differenza diventa relativamente trascurabile quando le lunghezze minime dei programmi diventano molto grandi. Ciò significa che, quando il programma è abbastanza lungo, vale a dire quando la complessità algoritmica è abbastanza grande, la dipendenza della sua

⁶ Hofstadter, D. R. (1984) *Gödel, Escher, Bach. Un'eterna ghirlanda brillante*. Adelphi Ed; (2009) *Anelli nell'Io. Che cosa c'è al cuore della coscienza?* Mondadori Ed.

⁷ Kolmogorov, A. N. (1965) Three approaches to the definition of the concept of "amount of information. *Problemy Peredachi Informatsii*, 1:3-11. MR32 #2273.

lunghezza dalle particolarità della macchina e del linguaggio di programmazione diventa trascurabile. Possiamo pertanto affermare che la definizione di Kolmogorov riveste un significato *quasi oggettivo*.

8. La complessità secondo von Neumann

Von Neumann non ha mai fornito una definizione precisa del termine “complessità”, che egli usa spesso come sinonimo di “complicazione”. Nei suoi scritti la nozione sembra significare talvolta il numero di parti di cui si compone un oggetto⁸, altre volte la lunghezza della descrizione letteraria di un oggetto. Questi due significati possono considerarsi equivalenti se i simboli alfanumerici che compongono la descrizione letteraria di un oggetto sono considerati come “parti” elementari della descrizione. Per questa ragione essa può definirsi *complessità descrittiva*.

Poiché ogni descrizione può essere codificata in una sequenza di 0 e 1, la complessità descrittiva può essere matematicamente precisata come il numero di 0 e 1 che bisogna usare per codificare la descrizione letteraria di un oggetto. Secondo questa definizione una sequenza infinita di 0 o di 1 ha una complessità descrittiva infinita, sebbene possa essere codificata in modo semplicissimo nella forma “scrivi 0, oppure 1, e ripeti questa operazione infinite volte”. Chiaramente, la definizione non è priva di ambiguità dato che dipende dalla ridondanza della lingua⁹, da quella della descrizione¹⁰, dalle procedure di codifica usate per indicare le parti ripetitive del testo ecc. Ad esempio una sequenza di dodici zeri può essere codificata anche nella forma QQQ , dove $Q=0000$, e così via.

Ad essere precisi, la nozione di complessità descrittiva fu usata da von Neumann nell’ambito della teoria degli automi in un senso più qualitativo che quantitativo con lo scopo di dare risalto ad una differenza fondamentale tra due nozioni di diverse complessità, che possiamo rispettivamente chiamare *complessità strutturale* e *complessità comportamentale*. Ecco come il grande matematico introdusse questa distinzione in una conferenza tenuta all’Università dell’Illinois nel 1948:

*Normalmente, una descrizione letteraria di ciò che si suppone un automa faccia è più semplice dello schema completo dell'automa. Non è a priori vero che debba essere sempre così. Ce n'è abbastanza nella logica formale per indicare che la descrizione delle funzioni svolte da un automa è più semplice dell'automa stesso fintanto che l'automa non è molto complicato, ma che quando si raggiungono livelli di complicazione superiore, l'oggetto reale è più semplice della descrizione letteraria di ciò che fa. (Von Neumann, *The Theory of Self-Reproducing Automata*).*

Come è chiarito in altre parti del libro ora citato, ciò che von Neumann intende dire è che la complessità descrittiva di una macchina può essere più piccola, anzi molto più piccola, di quella dei processi, anzi dei singoli processi, che la macchina può eseguire. Così, dato che in generale la complessità degli oggetti prodotti riflette quella dei processi che li producono, si riesce a spiegare come la complessità descrittiva degli oggetti prodotti possa essere maggiore, e persino infinitamente maggiore, di quella della macchina che li produce.

Un esempio di come questo stato di cose possa capitare è offerto dal processo di calcolo che genera quella figura piana delimitata da un contorno infinitamente frastagliato che va sotto il nome di *insieme di Mandelbrot* (il *frattale di Mandelbrot* è il contorno). Chiunque abbia potuto analizzare la forma di questo impressionante oggetto matematico, ingrandendo i dettagli dell’immagine fornita da un calcolatore, non può avere alcun dubbio che si tratta di un oggetto di complessità indescrivibile, anzi di forma infinitamente più complicata di ogni altra che si possa osservare in natura. A dispetto di ciò, il programma che lo genera consiste di poche righe. In altri termini, la complessità algoritmica del

⁸ Von Neumann, J. (1966) *The Theory of Self-Reproducing Automata*, A.W.Burks Editor, University of Illinois Press,.

⁹ La ridondanza di un testo linguistico abbastanza lungo è pari alla percentuale di lettere alfabetiche che possono essere cancellate a caso senza che venga compromessa la possibilità di ricostruire il testo. Si è stimato che la ridondanza della lingua inglese è circa il 65%. Pertanto la complessità descrittiva di una descrizione letteraria in lingua inglese può essere stimata circa il 35% della lunghezza del testo

¹⁰ Una descrizione ripetuta n volte conta come una sola descrizione.

frattale è infinitamente più piccola della sua complessità descrittiva. Identificando la complessità strutturale del calcolatore come la somma della complessità descrittiva del microprocessore che fa “girare” il programma e di quella del programma generatore, si conclude che anche questa è infinitesima rispetto alla complessità descrittiva del frattale.

A questo punto è evidente che, per caratterizzare la complessità di un oggetto matematico, ma più in generale di un oggetto di natura qualsiasi, è opportuno avvalersi sia dell’approccio di Kolmogorov sia di quello di von Neumann. L’approccio di Kolmogorov si presta bene per caratterizzare la complessità delle procedure di calcolo e degli oggetti matematici, mentre quello di von Neumann si presta bene per caratterizzare la differenza tipologica tra la complessità descrittiva di un sistema fisico e la complessità descrittiva della sua dinamica. Ciò è particolarmente rilevante se si considera che nell’insegnamento della fisica si è fatto sempre credere che le dinamiche dei sistemi siano sempre descrivibili.

9. Un esempio di complessità indescrivibile: l’insieme di Mandelbrot

A questo punto è opportuno fare alcune considerazioni su certi problemi posti da queste diverse nozioni di complessità. La complessità algoritmica di una semplice sequenza infinita di 0 è grosso modo una breve riga di programma, mentre quella dell’insieme di Mandelbrot è circa una ventina di righe di programma. D’altronde, sia la complessità descrittiva di una sequenza infinita di 0 e quella dell’insieme di Mandelbrot sono infinite. Non c’è tuttavia alcun dubbio che si tratta di complessità descrittive di tipo assai diverso. Per comprendere dove sta la differenza dobbiamo cercare di capire dove nasce la difficoltà di descrizione dei due tipi di oggetti. Il punto cruciale è che, mentre ci basta un’occhiata per capire quale è il programma che genera una sequenza di zeri, non è per niente facile capire quale sia il programma che genera il frattale di Mandelbrot, per quanto accurate e dettagliate siano le descrizioni degli insiemi che delimitano questa immagine per gradi di approssimazione crescente.

Il fatto è che nessuna immagine dell’insieme di Mandelbrot limitata ad un grado di precisione finita, come d’altronde è ogni immagine realmente esistente, permette di risalire in modo univoco al programma generatore del frattale. Si possono, infatti, immaginare infiniti programmi, anche di complessità algoritmica illimitata, che sono capaci di riprodurla in modi identici ad ogni predefinito grado di precisione, ma che produrrebbero immagini diverse ai gradi di precisione superiore. In altri termini, solo una descrizione infinitamente precisa del frattale permetterebbe di risalire in modo univoco alla formula che lo genera. Ma questo è impossibile, poiché una descrizione infinitamente precisa del frattale non può esistere in alcun luogo del nostro universo. Emerge qui in modo drammatico un aspetto paradossale della nozione di complessità algoritmica: se conosciamo il programma generatore possiamo produrre l’oggetto in modo preciso o con precisione arbitrariamente grande, ma se la complessità descrittiva dell’oggetto è infinita non possiamo risalire al programma generatore.

E’ chiaro che la differenza basilare tra la complessità descrittiva della sequenza di zeri e quella dell’insieme di Mandelbrot sta nel fatto che per ogni porzione finita della prima possiamo facilmente trovare il programma che la genera, mentre è impossibile fare la stessa cosa per il secondo. Nel primo caso la ridondanza della sequenza è palese - e ciò ci consente di ricodificare facilmente la descrizione in modo economico; nel secondo caso non è affatto evidente quali relazioni sussistono tra le diverse parti del magico insieme, tanto più che le girandole di braccia e filamenti esibite dalle parti del frattale coinvolgono in modo assai complicato tutti i numeri primi, che sono infiniti! Converterà pertanto distinguere la nozione di *complessità descrittiva infinita* da quella di *complessità indescrivibile*, riservando questa seconda denominazione al caso in cui sia impossibile risalire al programma generatore a partire da ogni porzione finita dell’oggetto.

Tutte queste diverse nozioni di complessità possono disorientare il lettore, ma è opportuno rilevare il fatto che in realtà esse sono soltanto nozioni accessorie preliminari del grande tema della complessità.

Se dovessimo tentare di elencare tutte le diverse forme e i diversi modi teoricamente concepibili della complessità ci troveremmo di fronte ad un'impresa disperata, giacché la nozione di complessità è essa stessa indescrivibilmente complicata.

10. Macchine ricorsive e processi ricorsivi

Una proprietà caratteristica delle macchine calcolatrici è di poter transitare tra un insieme discreto di stati, ed eventualmente ritornare allo stato iniziale, sotto il controllo di un orologio interno che regola come un direttore di orchestra i passi di calcolo della macchina (cicli). Una macchina capace di utilizzare ad ogni ciclo i risultati da essa stessa prodotti nei cicli precedenti è detta *ricorsiva*. Anche i processi di calcolo eseguiti da una macchina che opera in modo ricorsivo, e i programmi che istruiscono la macchina ad un comportamento ricorsivo, sono detti ricorsivi. Di questo tipo è ad esempio il processo che genera il frattale di Mandelbrot, come pure il programma che lo istruisce.

Una macchina ricorsiva di complessità finita ma dotata di memoria infinita e capace di funzionare per un tempo infinito è detta *macchina di Turing*¹¹. Si tratta ovviamente di una macchina ideale. I processi ricorsivi eseguibili da una macchina di Turing sono in ogni caso istruiti da programmi finiti, pertanto la complessità algoritmica degli oggetti che essa può produrre è sempre finita. Tuttavia, essa è capace di funzionare per un tempo infinito generando oggetti di complessità descrittiva infinita. Chiaramente, i processi di una macchina di che si arrestano dopo un numero finito di cicli producono solo oggetti di complessità descrittiva finita.

Esistono processi ricorsivi di vario genere. Possiamo immaginarne di molto semplici, ad esempio quelli che effettuano ad ogni ciclo un numero fisso di operazioni, incrementando di volta in volta un certo numero di parametri e producendo in questo modo un numero fisso di risultati. Di questo tipo sono quelli che generano le sequenze di cifre descritte nel paragrafo sulla complessità di Kolmogorov. Possiamo immaginarne altri in cui i risultati prodotti ad ogni ciclo aumentano progressivamente, vuoi con legge di potenza vuoi con legge esponenziale o vuoi ancora con legge di esponenziale di esponenziale ecc., e persino processi in cui i dati temporaneamente immagazzinati nella memoria della macchina per essere usati nel ciclo successivo aumentano indefinitamente con le più svariate leggi nel corso del tempo. L'analisi della complessità di calcolo ricorsivo è un capitolo importante della teoria delle macchine¹². A dire il vero, sarebbe un capitolo inesauribile dato che la varietà dei comportamenti ricorsivi possibili è infinita e indescrivibilmente complicata.

La rilevanza epistemologica della teoria di Turing è espressa nel modo più chiaro dalla tesi (tesi di Church (1936¹³) secondo la quale tutti i processi di calcolo eseguibili dalla mente umana sono eseguibili, in linea di principio, anche da una macchina di Turing opportunamente programmata.

Un altro notevole risultato di questa teoria è che i processi di elaborazione ricorsivi basati sulla copiatura, la cancellazione e l'incollatura di stringhe alfanumeriche e di parti di queste stringhe (Post, 1936¹⁴) sono equivalenti a processi eseguibili dalle macchine di Turing. Il nostro interesse per questo argomento sta nel fatto che i sistemi biologici sono, in senso generalizzato, macchine ricorsive naturali che al posto di simboli alfanumerici usano molecole semplici reperibili in un ambiente fisico. In un certo senso i sistemi di Post valgono come modelli matematici di processi biologici. Tuttavia, la gran differenza tra i processi ricorsivi delle macchine artificiali e quelli delle macchine biologiche (autoriproduzione, embriogenesi, attività mentale ecc.) è che i primi sono "sequenziali" mentre i secondi sono "paralleli". Questa differenza diversifica in modo sostanziale le proprietà temporali dei processi: velocità, aspetti riguardanti la contemporaneità, la messa in concerto di più processi ecc.

¹¹ Turing, A. M. (1936) On computable Number, with an application to the Entscheidungsproblem, *Proceedings of the London Mathematical Society*, (Ser. 2, Vol. 42);

¹² G. Ausiello (1975) *Complessità di calcolo delle funzioni*. Bollati Boringhieri.

¹³ Church, A. (1936) An unsolvable problem in elementary *number theory*, *American Journal of Mathematics*, 58:345-363.

¹⁴ Post, E. L. (1936), Finite Combinatory Processes - Formulation 1, *Journal of Symbolic Logic I*: 103-105.

11. Processi sequenziali e processi paralleli

Per ragioni inerenti allo sviluppo dell'informatica contemporanea, precisamente le difficoltà ad implementare una fitta rete di connessioni multiple tra gli elementi che compongono un circuito elettronico, le macchine calcolatrici attualmente disponibili sul mercato hanno un funzionamento di tipo *sequenziale*. Esse eseguono, una di seguito all'altra, operazioni logico-aritmetiche su insiemi di dati alfanumerici eseguendo di volta in volta un numero esiguo di operazioni. Il deficit di connettività di queste macchine è compensato dalla grandissima velocità di trasmissione dei segnali, in pratica la velocità della luce, che attualmente consente agli ordinari calcolatori da tavolo di eseguire alcuni miliardi di operazioni logico-aritmetiche per secondo. I limiti alla loro prestazioni derivano principalmente dalla natura meccanica dei dispositivi di memoria (dischi fissi).

Per contrasto, un sistema che processa informazione come il cervello possiede una grandissima connettività interna ma una velocità di trasmissione dei segnali approssimativamente equiparabile a quella del suono, vale a dire circa un milione di volte inferiore a quella della luce e, in generale, dei segnali elettrici. Dato che i neuroni non riescono a produrre potenziali di azione con frequenza superiore ad un paio di millisecondi, possiamo fissare come limite superiore teorico della capacità di calcolo di un singolo neurone il valore di circa 500 operazioni per secondo. In realtà questa cifra è più piccola perché l'attività di sparo dei neuroni è pilotata e sincronizzata da oscillazioni di potenziale elettrico intrinseche delle membrane neuronali alle frequenze tipiche degli elettroencefalogrammi.

E' importante rilevare che la capacità di generare processi di complessità illimitata dipende in ogni caso dal funzionamento ricorsivo. Nel caso delle macchine parallele, questo richiede che i flussi di dati trasportati da una regione all'altra del cervello da fasci di fibre nervose parallele siano sincronizzati. Si sa, del resto, che i processi di sincronizzazione e desincronizzazione dei potenziali di azione dei neuroni hanno un ruolo fondamentale nel funzionamento del cervello. Considerando che la frequenza di sincronizzazione tipica delle funzioni corticali è circa 80 cicli al secondo (frequenza gamma), dobbiamo portare la stima della capacità di calcolo di un singolo neurone a circa 80 cicli al secondo.

Per rendere un'idea di quello che può essere un processo ricorsivo parallelo, Hofstadter ha fatto un esperimento (da me riprodotto con successo) piazzando, proprio di fronte allo schermo di un televisore, una telecamera che riprende l'immagine riprodotta dallo schermo stesso. L'informazione visiva che esce dallo schermo si comporta come un flusso di dati parallelo che è continuamente catturato dalla telecamera (nel mio caso una videocamera digitale) e rinviato, attraverso lo schermo, di nuovo alla telecamera. Questi dati sono leggermente deformati dall'apparato ottico della telecamera e dall'apparecchiatura elettronica che rinvia i segnali allo schermo del televisore (nel mio caso lo schermo del mio portatile).

L'effetto di questo anello ottico è la produzione di immagini astratte variamente colorate, totalmente diverse da quelle che possono essere catturate dalla telecamera quando non è rivolta verso lo schermo del televisore. Esse mutano di forma imprevedibilmente generando figure cangianti in modi sempre diversi. L'inserimento temporaneo di un oggetto tra la telecamera e lo schermo innesca un processo ricorsivo che evolve rapidamente verso figure sempre più complesse, totalmente diverse dall'immagine iniziale; immagini che tuttavia "ricordano" in qualche modo l'informazione visiva inizialmente catturata.

Naturalmente, il processo ricorsivo parallelo televisivo è molto rozzo rispetto a quelli che presumibilmente avvengono nel cervello. E' interessante rilevare che l'anatomia del cervello evidenzia l'esistenza di un massiccio circuito capace di generare un processo ricorsivo parallelo. Esso è formato dall'insieme dei circuiti neuronali che percorrono la sequenza *talamo->corteccia cerebrale->gangli basali->talamo*. Questo anello circuitale ricorda la disposizione degli apparati elettronici su descritta, col talamo, più precisamente l'insieme dei nuclei talamici, che fa la parte della telecamera, o di un sistema di telecamere, e la corteccia, più precisamente l'insieme di aree corticali, quella dello schermo, o di un sistema di schermi.

C'è tuttavia una gran differenza tra la semplicità delle trasformazioni subite dal flusso d'informazione veicolata del circuito televisivo e quelle subite dal flusso d'informazione veicolata dai circuiti talamo-corticali. Il numero di circuiti che compongono l'anello talamo-corticale, attraversando complessivamente alcune decine di stadi formati da popolazioni neuronali omogenee, può essere approssimativamente stimato dell'ordine di grandezza di alcune decine di miliardi (si stima che l'intero cervello ne abbia 100 miliardi). Assumendo per ogni neurone una velocità di calcolo media di circa 80 operazioni al secondo possiamo attribuire alla macchina ricorsiva cerebrale un potenza di calcolo di circa mille miliardi di operazioni al secondo. Questa cifra è circa duecento volte maggiore della potenza di calcolo dei microprocessori attuali più avanzati (5 miliardi di operazioni al secondo). Si può tuttavia prevedere che tra una decina di anni il progresso tecnologico avrà colmato questa differenza.

Ma nel confronto tra le prestazioni del cervello e quelle del calcolatore, le differenze di prestazione più serie sono quelle che riguardano la capacità di memoria e il rapporto tra il tempo di accesso alla memoria e i tempi dei cicli di calcolo. Queste differenze sono enormemente più sfavorevoli per il calcolatore. La capacità di memoria di una sinapsi neuronale si può stimare di circa 1 byte (8 bit). Dato che ogni neurone possiede mediamente diecimila sinapsi, possiamo stimare che la capacità di memoria di un neurone sia di circa 10 kilobyte e quella del circuito talamo-corticale sia di alcune decine di terabyte (migliaia di miliardi di byte). Probabilmente molto di più se anche altre cellule del cervello (in particolare gli astrociti) possiedono meccanismi di memoria. Circa la velocità di accesso alla memoria, quella di una sinapsi si può stimare 1-2 millisecondi (pari al tempo di attivazione di una sinapsi), che è una frazione trascurabile del tempo medio di un ciclo di un circuito riverberante neuronale.

Quanto alla memoria esterna dei calcolatori, bisogna considerare che oggi sono in commercio dischi fissi che hanno una capacità di memoria di alcuni terabyte, una quantità confrontabile con quella del cervello. Purtroppo il tempo di accesso ai dati memorizzati in un disco fisso è di alcuni millisecondi. Ciò ritarda in modo enorme la velocità di calcolo effettiva dei microprocessori. Infatti, per eseguire un'operazione di calcolo utilizzando dati memorizzati nel disco fisso bisogna prima leggere i dati dal disco e poi scrivere i risultati sul disco (*swapping*). Credo che tutti coloro che hanno un calcolatore con poca memoria RAM (*random access memory*) sappiano quanto sono noiose queste operazioni di *swapping*.

Per contrasto, la memoria RAM, che funziona in modo totalmente elettronico, ha un tempo di accesso di circa 20 nanosecondi (circa due cicli macchina), che dunque non è trascurabile rispetto al ciclo di calcolo del microprocessore. Questo in pratica diminuisce il tempo effettivo del ciclo di calcolo di cinque volte ($2+2+1$). In vista dei futuri progressi della tecnologia, possiamo dire che se la memoria RAM fosse tanta quanta quella di un disco fisso, la potenza di calcolo effettiva di un calcolatore si avvicinerebbe molto a quella di un cervello umano. Purtroppo però le memorie RAM ora disponibili sono al massimo dell'ordine di grandezza di alcuni gigabytes (miliardi di byte).

Queste cifre ci danno la misura di quanto sia distante la possibilità di simulare con un calcolatore le funzioni del cervello umano. Esse spiegano inoltre perché il sistema visivo è capace di riconoscere un oggetto posto in un ambiente complesso in un decimo di secondo, mentre un calcolatore, per svolgere questa funzione con un grado di efficienza accettabile, impiegherebbe un tempo migliaia di volte maggiore. Bisogna tuttavia considerare che il progresso tecnologico sembra non conoscere limiti, cosicché non è azzardato pensare che in un futuro non tanto lontano sarà possibile disporre di sistemi di processamento d'informazione capaci di simulare in tempo reale i processi paralleli del cervello umano. E' pertanto legittimo chiedersi se si riuscirà a produrre un automa "autocosciente".

12. Le macchine logico-aritmetiche e la nozione di calcolatore universale

Come già detto, i comuni calcolatori sono essenzialmente macchine programmabili capaci di eseguire operazioni logiche e aritmetiche. Esse sono inoltre dotate di memoria e di dispositivi per la

presentazione dei loro risultati. La dimostrazione che una macchina calcolatrice programmabile dotata dei requisiti minimi per poter eseguire le operazioni logiche e aritmetiche è per ciò stesso in grado di eseguire qualunque processo informazionale, fu fornita da Alan Turing nel 1936. Il possesso di tale capacità algoritmica basta a rendere un calcolatore *universale*, vale a dire capace di simulare il comportamento di qualsiasi altro calcolatore. In ultima analisi, le differenze di prestazione tra calcolatori universali riguardano solo la velocità e la quantità di memoria.

Questa scoperta portò i matematici ad identificare il dominio del “matematicamente calcolabile” con quello del calcolabile mediante una macchina di Turing universale. Uno dei grandi meriti della teoria di Turing fu quello di rendere possibile la traduzione delle scoperte di Gödel in termini informatici. In particolare, il teorema d’indecidibilità di Gödel fu tradotto da Turing nel cosiddetto problema dell’*halting*: nessun calcolatore è generalmente in grado di predire in un tempo finito se un processo ricorsivo eseguito da un altro calcolatore avrà termine o no.

Una macchina universale ricorsiva è in grado di generare tutti i processi possibili, anche quelli di complessità arbitrariamente elevata. I tempi impiegati dalle macchine sequenziali per eseguire processi molto complessi può essere proibitivamente grande, ma questo limite può non valere per le macchine parallele.

13. Le macchine costruttrici e la nozione di universalità costruttiva

La teoria degli automi di von Neumann (1948-1956) costituisce un’estensione concettuale della teoria di Turing. Un *automa costruttore* è una macchina capace di utilizzare oggetti reperibili nell’ambiente per produrre altri oggetti da riporre nel medesimo ambiente. In questo senso è una macchina ricorsiva che opera sugli oggetti di un ambiente invece che su insiemi di simboli. Per svolgere questa funzione l’automa deve possedere un repertorio sufficientemente ricco di sensori e strumenti di lavoro che gli permettono di eseguire operazioni di cernita e assemblaggio dei componenti necessari per la produzione di oggetti fisici più complessi. Un automa costruttore è in generale dotato di programmi di lavoro capaci d’istruire i processi di costruzione; può essere inoltre dotato di un calcolatore interno ed essere pertanto capace di elaborare informazione. Se l’automa è organizzato in modo troppo semplice non riuscirà a produrre altro che oggetti di complessità limitata, ma se è abbastanza complesso e opportunamente organizzato potrà funzionare come una *macchina universale*. Vale a dire come una macchina costruttrice capace di fare tutto quello che può fare qualsiasi altra macchina.

Come per le macchine di Turing, l’universalità costruttiva è resa possibile soltanto se il funzionamento della macchina è istruito da un programma. Se è dotata di opportuni programmi, essa è in grado di produrre oggetti di complessità descrittiva uguale o superiore alla sua; in particolare, anche una copia di sé stessa e del suo stesso programma. La riproduzione cellulare e l’embriogenesi sono esempi lampanti di questi processi di generazione. Del resto von Neumann, con la sua teoria degli automi, mirava proprio a fornire le basi teoriche dei fenomeni biologici.

Come ha osservato von Neumann, un costruttore universale che non sia internamente dotato di un calcolatore universale può produrre oggetti di grande complessità descrittiva solo se è istruito da programmi di altrettanto grande complessità descrittiva. Ma se è dotato di un calcolatore universale, esso è in grado di produrre anche oggetti di grandissima complessità descrittiva utilizzando programmi piccoli che riflettono la complessità algoritmica dell’oggetto (come nel caso della macchina che genera il frattale di Mandelbrot).

L’istanza fondante della profonda visione di von Neumann è la distinzione tra la complessità strutturale e la complessità comportamentale di un sistema formale o di un sistema materiale fisico o biologico. Sotto un certo livello di complessità strutturale, i comportamenti di un algoritmo, di una macchina calcolatrice o di un automa costruttore sono perfettamente prevedibili e descrivibili in termini finiti sulla base della conoscenza della sua struttura. In queste stesse circostanze, la potenza interpretativa dell’algoritmo, quella di calcolo del calcolatore o la capacità produttiva dell’automa non

può che essere limitata. Corrispondentemente, l'universalità algoritmica, quella computazionale e quella costruttiva, di conseguenza anche l'autoriproduzione, sono impossibili.

Tuttavia, se tali sistemi possiedono strutture sufficientemente ordinate e articolate, si crea un fatto sconvolgente: la maggioranza dei comportamenti esibiti dalla macchina *non sono logicamente prevedibili e descrivibili sulla base dalla conoscenza della sua struttura e dei programmi da essa utilizzati*, per quanto perfetta possa essere questa conoscenza. L'unico modo di conoscerli è quello di osservarli nel loro effettivo svolgimento. Proprio in queste stesse circostanze, la potenza interpretativa dell'algoritmo, la capacità di simulazione del calcolatore o quella produttiva dell'automa diventano universali.

14. Capacità autoreferenziale del formalismo logico-aritmetico

Una delle più importanti conquiste del pensiero umano avvenute nel '900 è certamente l'indagine condotta da Kurt Gödel (1930-31)¹⁵ sulla capacità della logica di descrivere e spiegare la potenza espressiva dell'aritmetica e della matematica in generale. E' un'opinione piuttosto diffusa che l'importanza dell'indagine gödeliana consista principalmente nel mettere in evidenza i limiti del pensiero matematico. In questo senso, essa avrebbe una portata sostanzialmente negativa per le aspettative di onnicomprensione del pensiero scientifico.

Questa credenza è fondamentalmente sbagliata. La via attraverso la quale Gödel giunge ai suoi risultati dimostra, in realtà, che gli algoritmi matematici, in particolare l'aritmetica, sono infinitamente più potenti delle teorie logiche che pretendono di descriverli e governarli. L'esistenza di proposizioni logiche vere, ma *indimostrabili*, non è affatto una dichiarazione d'impotenza dell'aritmetica, ma del fatto che esistono procedure algoritmiche *deterministiche*, e tuttavia di comportamento imprevedibile, che potrebbero produrre in un tempo infinito oggetti che nessuna deduzione logica è in grado di prevedere, descrivere o classificare. Questo è tutt'altro che l'evidenza di una limitatezza del pensiero matematico; essa è, al contrario, l'evidenza del fatto che gli algoritmi, e più in generale i processi analitico-costruttivi della matematica, possiedono una potenza creativa infinita!

Merita forse dilungarci un po' di più su questo argomento. L'indagine di Gödel si basa sulla scoperta che le dimostrazioni logiche possono essere tradotte in un linguaggio aritmetico ed interpretate da calcoli aritmetici. Questa possibilità origina dal fatto che le elaborazioni logiche concrete, ad esempio le dimostrazioni dei teoremi di un sistema assiomatico (geometria euclidea, topologia, meccanica razionale, teoria dei numeri ecc.), non sono altro, in ultima analisi, che produzioni algoritmiche di dati a partire da altri dati, esattamente come lo sono i calcoli aritmetici. Infatti, le operazioni della logica booleana, applicate a collezioni infinite di possibili oggetti matematici (variabili), di cui consiste una dimostrazione logica, non sono altro che una sorta di algoritmo applicato a proposizioni logiche ben formate per la produzione di altre proposizioni logiche ben formate. Ora, l'aritmetica, giacché è un algoritmo universale, può simulare tutte le procedure dimostrative di una qualsiasi teoria logica associando, in primo luogo, ogni formula ben formata ad un numero intero particolare¹⁶ e codificando gli assiomi della teoria nella forma di operazioni aritmetiche (gödelizzazione). In tal modo, la dimostrazione di un teorema si traduce in una sequenza di operazioni aritmetiche condotte su un particolare insieme di numeri interi. La sequenza di complessificazioni e semplificazioni del discorso che intervengono in una dimostrazione logica si tradurrà, in corrispondenza, in una sequenza di crescite e diminuzioni dei numeri. Procedendo in questo modo, l'intero sistema dei ragionamenti logici ammissibili dalla teoria sarà rappresentato da un sistema di corrispondenti operazioni aritmetiche che si concluderanno con la produzione del numero che codifica il teorema.

¹⁵ Gödel, K. (1931) Über formal unentscheidbare Sätze der Principia Mathematica und verwant der Systeme, I, *Monatshefte Math. Phys.* 38:173-198.

¹⁶ Enumerando con degli interi i simboli alfabetici si può persino determinare in modo univoco il numero di Gödel della Divina Commedia! Si trova un numero gigantesco.

Il passaggio chiave dell'analisi gödeliana consiste nella "gödelizzazione" della stessa teoria assiomatica dell'aritmetica: un processo autoreferenziale ricco di straordinarie implicazioni e conseguenze. Grazie a questa procedura di *auto-gödelizzazione*, l'algoritmo aritmetico è messo nelle condizioni di poter interpretare se stesso e di dimostrare, nella forma di calcoli aritmetici particolari, le proprietà logicamente dimostrabili dell'aritmetica stessa. Così ad esempio il teorema "esistono infiniti numeri primi", proposizione che non è ovviamente verificabile analizzando caso per caso quali numeri siano primi, sarà codificata da un numero particolare.

In tale procedimento autoreferenziale, certe dimostrazioni riguardanti le proprietà generali dell'aritmetica si traducono in particolari procedure aritmetiche ricorsive. Quelle che producono numeri di Gödel finiti sono finite. Esse corrispondono ai teoremi *veri e dimostrabili*. Quelle che producono numeri di Gödel infiniti sono indescrivibilmente complicate. Esse corrispondono ai teoremi *veri ma non dimostrabili*. E' di straordinario interesse rilevare come la procedura autoreferenziale gödeliana permetta di interpretare un universo di significati logici *generali* mediante una collezione di casi *particolari*.

15. Capacità autoreferenziale del linguaggio logico

Gödel, interpellato da Burks dopo la morte di von Neumann circa il possibile significato del riferimento fatto dal grande ungherese ai suoi teoremi, si limitò a rilevare che la distinzione tra tipi logici citata da von Neumann è analoga a quella tra linguaggio e metalinguaggio, che fu introdotta da Tarski nel fatidico 1936. In realtà, come si evince da altri scritti, von Neumann non intendeva affatto riferirsi alla teoria del linguaggio di Tarski ma a quella delle macchine di Turing.

La sorprendente affermazione di Gödel potrebbe farci credere che il sommo logico slovacco avesse frainteso il senso della domanda. Ma, ad un'analisi più attenta l'idea che esista una profonda relazione tra teoria delle macchine e quella dei linguaggi appare molto suggestiva.

Secondo la definizione di Tarski, gli enunciati di un linguaggio sono messaggi che veicolano informazione su cose che accadono in un dominio del mondo esterno. Quelle del *metalinguaggio* sono messaggi che veicolano informazione sulle relazioni tra i messaggi del linguaggio che descrive il mondo esterno, che può essere pertanto definito *linguaggio oggetto*. La ragione per cui il metalinguaggio appartiene a un tipo logico superiore rispetto al linguaggio oggetto è che le deduzioni logiche del metalinguaggio si comportano come procedure di calcolo ricorsivo applicate alle proposizioni del linguaggio oggetto assunto come universo del discorso. Per svolgere questa funzione, il metalinguaggio deve contenere al suo interno, oltre a tutte le proposizioni del linguaggio oggetto, anche quelle necessarie per definire nuovi significati, per descrivere le operazioni logiche eseguibili su queste proposizioni e i criteri che permettono di stabilirne le verità o le falsità ecc.

Partendo da questa considerazione, Tarski dimostrò che esistono enunciati metalinguistici che sono intraducibili nel linguaggio oggetto. Ad esempio il concetto stesso di verità non è definibile nel linguaggio oggetto (Tarski, 1936, 1972)¹⁷. Abbiamo qui un'impressionante analogia tra la differenza di tipo logico del linguaggio in rapporto al metalinguaggio e la differenza di tipo logico della complessità strutturale di in rapporto a quella comportamentale indicata da von Neumann. Si tratta in definitiva di un'ulteriore versione del teorema di indecidibilità di Gödel, come pure del problema dell'*halting* di Turing. Queste similitudini suggeriscono che la macchina generatrice del linguaggio (cioè la mente umana) elabori significati metalinguistici attraverso procedure ricorsive simili a quelle dei calcolatori, con la differenza tutt'altro che trascurabile che si tratta di una macchina parallela invece che sequenziale. Abbiamo qui un'indicazione assai significativa dei meccanismi che stanno alla base della produzione del pensiero.

¹⁷ Tarski, A. (1956) The Concept of Truth in Formalized Languages, in *Logic, Semantic, Metamathematics*, 152-278, Clarendon Press, Oxford; (1972) Verità e dimostrazione - *Lecture da "Le Scienze"*, 50:70-79.

L'analogia tra la teoria delle macchine e quella del linguaggio ci permette di cogliere un altro importante aspetto della teoria della complessità. Quello che riguarda il rapporto tra il tipo dei linguaggi atti a descrivere le strutture delle macchine e quelli atti a descriverne i comportamenti e i prodotti di tali comportamenti. Per descrivere gli oggetti generati da processi ricorsivi illimitati non basta il repertorio di termini di un normale linguaggio descrittivo, ma diventa inevitabile usare un linguaggio metaforico, che riesce a rendere tanto meglio le idee quanto più estende il suo campo di significati oltre il dominio delle scienze e delle tecniche dove tutto è logicamente descrivibile.

Per esemplificare queste affermazioni, torniamo di nuovo all'insieme di Mandelbrot. Il programma che lo genera per approssimazioni successive consiste di poche righe che istruiscono un procedura ricorsiva condizionata da un criterio selettivo. Ciononostante, l'insieme di Mandelbrot è stato definito come l'oggetto geometrico più complesso mai visto in matematica. Così, in un certo senso, la sua immagine fornisce una rappresentazione sinottica del comportamento del calcolatore sotto l'azione del minuscolo programma.

Nel tentativo di descrivere la sua forma, Roger Penrose¹⁸ (1989) usò non solo termini appartenenti al normale linguaggio geometrico, ed esempio segmenti, cerchi, spirali, cardioidi ecc., ma anche altri assai stravaganti e inconsueti come "code di cavallucci marini", "valli di cavallucci marini", "anemoni marine", "dragoni a n teste", "apparenze floreali", "molteplicità filamentose", "mondi bambini" e così via. Chiaramente, non c'è alcun modo di tradurre gli esatti significati di queste espressioni in puri termini geometrici o algebrici. Ingrandendo i dettagli dell'insieme di Mandelbrot - ciò che può farsi con un calcolatore di media potenza - si può facilmente mettere in evidenza un repertorio infinito di figurazioni sempre diverse. Queste possono essere vagamente descritte come complesse arborizzazioni generate per infinita combinazione di cicli ed epicicli che hanno come periodi i numeri primi. Osservando questo gioco infinitamente complesso avvertiamo un senso di armonia, anche se ogni tentativo di descriverlo appare disperato e se alla fine ci risulta noioso.

L'insieme di Mandelbrot è un esempio di come il passaggio dal linguaggio che descrive la struttura del calcolatore e del programma che lo istruisce al linguaggio che ne descrive in qualche modo il comportamento richiede l'uso di metafore ispirate alle forme di organismi biologici, di parole i cui significati oltrepassano il mondo delle strutture artificiali, di un linguaggio che può definirsi metastrutturale. Possiamo farci un'idea di come le elaborazioni che un metalinguaggio è in grado di operare sul suo linguaggio oggetto possano far emergere valori e significati che trascendono il livello del linguaggio oggetto e che, per essere espresse, richiedono un più ampio bagaglio di nozioni e concetti primitivi.

16. Capacità autoriproduttiva dei costruttori universali

Von Neumann - richiamandosi al teorema di indecidibilità di Gödel - afferma che l'automa non potrà "auto-copiarsi" semplicemente esaminando direttamente la propria struttura interna mediante sensori o altri ipotetici apparati d'introspezione, poiché nessuna auto-misurazione o auto-osservazione può essere completa. Di fatto, nessun apparato di misura può compiere una misura su sé stesso. Affinché l'autoriproduzione abbia luogo è essenziale che l'automa costruttore disponga di una descrizione della sua struttura interna e di un apparato di decodificazione che gli permetta di dedurre le procedure costruttive o, equivalentemente, di un elenco completo delle istruzioni riguardanti le modalità di costruzione e assemblaggio delle sue parti. Tale è appunto il codice genetico degli organismi viventi.

Si ha in ciò l'analogo della codificazione della teoria assiomatica dell'aritmetica in un insieme di operazioni aritmetiche, senza la quale l'auto-gödelizzazione dell'aritmetica non potrebbe effettuarsi. È notevole il fatto che von Neumann giunse ad affermare tutto questo tre o quattro anni prima della

¹⁸ Penrose, R. (2000) *La Mente Nuova dell'Imperatore*, Rizzoli Ed., Milano; (2000) *Il grande, il piccolo e la mente*, Raffaello Cortina Ed., Milano.

scoperta del DNA (Watson e Crick, 1953), suggerendo, tra l'altro, sulla base di argomenti relativi a condizioni di funzionalità ottimale, che l'informazione genetica sia codificata in sequenze lineari di dati (come i simboli sui nastri delle macchine di Turing). Il sommo matematico ha fornito l'esempio di un semplice automa planare capace di autoriprodursi. Esso è costituito da un insieme di cellette quadrate uguali, arrangiate a scacchiera su un piano, ciascuna delle quali può transitare in tempi discreti tra 29 possibili stati interni. Lo stato di ogni celletta all'istante successivo è determinato dagli stati della celletta stessa e di quelli delle cellette contigue.

Come è stato già detto nel paragrafo 13, una condizione necessaria perché una macchina possa generare oggetti di arbitraria complessità utilizzando programmi di complessità descrittiva inferiore a quelli della macchina stessa è che sia *un costruttore universale internamente dotato di una macchina calcolatrice universale*. In questo caso, la lunghezza del programma fornisce una misura approssimativa della complessità algoritmica dell'oggetto prodotto dalla macchina.

Lo stesso stato di cose vale se l'oggetto da produrre è una copia della macchina stessa, come è il caso degli organismi viventi. Nel caso di organismi semplici come i batteri la complessità descrittiva del genoma ha lo stesso ordine di grandezza di quella del resto del batterio. Ciò fa supporre che i batteri non possiedano qualcosa che funzioni come un calcolatore universale. Ma nel caso degli organismi multicellulari la complessità descrittiva del genoma dell'ovulo fecondato è certamente inferiore a quella dell'organismo sviluppato. E' dunque presumibile che lo sviluppo dell'organismo sia governato da calcolatore universale (di collocazione e natura finora sconosciute).

Importa qui rilevare che l'impossibilità delle macchine, artificiali o naturali, a riprodursi per auto-ispezione è rimediata e risolta con l'introduzione di parti che contengono le istruzioni per la costruzione di un oggetto uguale alla macchina stessa: i programmi genetici e, per le macchine molto complesse, i calcolatori universali.

17. Capacità autoreferenziale del pensiero umano

Un'importante estensione della teoria delle macchine universali riguarda la capacità di rappresentazione universale e di autorappresentazione della mente umana, avanzata negli anni '80 da Hofstadter, che investe direttamente il problema dell'autocoscienza.

Molti autori hanno osservato che l'autocoscienza non può generarsi da un processo d'*introspezione*. Ciò è perfettamente comprensibile poiché la tesi dell'introspezione darebbe luogo a regressioni all'infinito del genere: *io* sono consapevole di guardare il tavolo; → *io* sono consapevole di essere consapevole di guardare il tavolo; → *io* sono consapevole di essere consapevole di essere consapevole di guardare un tavolo; ecc.

La questione è per alcuni aspetti analoga a quella della macchina costruttrice universale che non può copiare sé stessa mediante un semplice procedimento di auto-ispezione. Ora, questa analogia con la macchina che copia se stessa grazie all'ausilio di un programma suggerisce che anche la "macchina della mente" sia dotata di un programma capace di produrre la rappresentazione della macchina stessa. In tal modo, quella entità che nell'esempio su riportato si presenta come "*Io che guardo il tavolo*" sarebbe in realtà l'immagine costruita da un programma di un personaggio chiamato *Io* che guarda l'immagine di un tavolo. Incidentalmente, questo si accorda abbastanza bene ciò che è testimoniato da persone soggette a stati di dissociazione mentale, le quali affermano di avere talvolta la sensazione di osservare sé stessi dall'esterno.

Con ciò si vuole asserire che ogni volta che noi diciamo *io*, in realtà non ci riferiamo, come siamo portati a credere, a noi stessi in carne ossa e nervi, né all'attività pensante in atto del nostro cervello, ma ad *una rappresentazione di noi stessi posti in relazione con la rappresentazione di altre cose*. Rappresentazione che, in omaggio ad Hofstadter, possiamo chiamare il *simbolo del sé*.

Il merito di Hofstadter va riconosciuto perché questo brillante autore introdusse per primo, nel suo libro *Gödel, Escher e Bach*, l'idea che il processo mentale generatore dell'autocoscienza assomigli al

procedimento di auto-gödelizzazione dell'aritmetica. Nella visione di Hofstadter, il simbolo del sé svolge il ruolo dell'aritmetica assiomatica, che in certo senso rappresenta il programma generatore dell'algoritmo aritmetico, o quello del programma genetico dei viventi. In *Anelli dell'Io*, Hofstadter precisa che il simbolo del sé è generato da un processo cerebrale ricorsivo dotato di programma residente nel cervello. In questo stesso libro, Hofstadter arriva anche a dichiarare che i processi generatori del simbolo del sé possono persino generare, sia pure in forma larvale, i simboli del sé di altri esseri umani coi quali il soggetto pensante principale abbia sufficientemente interagito e comunicato nel corso della sua vita.

Rispetto alla visione di Hofstadter, quella ispirata dalla teoria delle macchine di von Neumann esprime gli stessi concetti in altra forma. Io la preferisco perché ha il merito portare l'attenzione sul funzionamento della macchina cerebrale. Se la macchina della mente è, come tutto porta a credere, un calcolatore parallelo universale, allora si spiega come il processo ricorsivo che genera il simbolo del sé sia in grado di produrre rappresentazioni mentali di complessità enormemente maggiore di quella del programma che lo istruisce. Del resto, basta pensare alla straordinaria complessità dei sogni per rendersi conto che quei mondi onirici popolati da una miriade di persone e cose, che si comportano e si muovono in modo da dare l'illusione della realtà, non possono essere sfornati belli e pronti dalla memoria del cervello, semplicemente perché non ce n'è abbastanza, ma possono essere generati solo da una macchina ricorsiva universale.

Che la produzione ricorsiva di oggetti mentali complessi non richieda dopo tutto programmi proibitivamente complicati lo dimostra il fatto che gli ambienti architettonici e i personaggi animati dei comuni videogames sono gestiti da programmi di complessità descrittiva molto più piccola di quella dei mondi da essi generati. Il problema è semmai di quale strutture ausiliarie la macchina parallela che produce le immagini mentali possa avvalersi per assicurare la capacità autoriflessiva del pensiero.

Purtroppo Hofstadter non ha fatto alcuno sforzo per scoprire quale possa essere il correlato neurologico di questa attività di produzione del simbolo del sé, né quali processi ausiliari debbano entrare in gioco per generare la capacità autoriflessiva del pensiero. Quello che ho creduto di capire, leggendo il suo primo libro, è che il modello di auto-gödelizzazione dell'aritmetica sembra fornire implicitamente una soluzione del problema.

La capacità dell'algoritmo aritmetico di rappresentare al proprio interno, nella forma di particolari calcoli aritmetici, le stesse proprietà universali generali del calcolo aritmetico, si basa sull'esistenza di una teoria assiomatica dell'aritmetica (teoria di Dedekind e Peano) che si avvale di dimostrazioni logiche per studiare le proprietà generali del calcolo aritmetico. La teoria assiomatica media in qualche modo la relazione dell'algoritmo aritmetico, in quanto sistema di procedure di calcolo particolari, con se stesso, in quanto algoritmo dotato di proprietà generali. Pertanto, se il modello gödeliano deve valere fino in fondo, bisogna rispondere alle seguenti domande. Quale è il corrispettivo mentale dell'algoritmo aritmetico? Quale il corrispettivo della teoria logica dell'aritmetica? L'ipotesi che ho avanzato nell'articolo *La Macchina della Mente I*, disponibile in rete nel sito <http://pd.infn.it/~rnobili>, è che il primo sia identificabile nell'attività produttiva delle immagini mentali, e il secondo nell'attività generatrice del linguaggio.

18. Le prospettive dell'intelligenza artificiale

Gli argomenti presentati in questo articolo non mancheranno di suscitare opinioni divergenti. Mi pare che la divergenza principale sia quella che porta a rispondere in modo opposto alla questione se sia o no appropriato considerare gli organismi viventi e pensanti alla stregua di macchine. Come fisico devo dire che è nella natura stessa della fisica interpretare i fenomeni naturali con gli stessi mezzi concettuali e formali che si usano per progettare e costruire le macchine. L'idea che la natura sia governata da leggi matematiche è un altro modo di dire la stessa cosa. Le macchine sono rappresentabili matematicamente

perché le operazioni che ne permettono il controllo sono rappresentabili come operazioni matematiche. D'altronde, l'enorme potere predittivo dimostrato dalla fisica moderna non trova tanto spiegazione in qualche miracolosa capacità divinatoria del linguaggio fisico, ma nella capacità della fisica-matematica di rappresentare molto efficacemente, sebbene in modo idealizzato, le operazioni di controllo e di trasformazione della materia, vale a dire di funzionare essa stessa come una macchina dotata di capacità operazionali infinite e infinitamente precise. La macchina della fisica teorica, operando in modo ricorsivo sugli enti di un universo puramente simbolico, riesce a produrre anticipatamente come oggetti formali ciò che la macchina naturale è capace di produrre come oggetti materiali.

Può questa visione dell'universo macchina essere estesa anche agli esseri viventi? La mia risposta è sì, dal momento che tutte le parti di un organismo vivente possiedono una struttura fisica. Resta tuttavia da chiarire alcuni punti in corrispondenza dei quali la fisica incontra il suo limite naturale. Si tratta principalmente del fatto che la comprensione del funzionamento dei sistemi biologici richiede un ampliamento concettuale che va oltre i confini della fisica e che si proietta nel dominio delle scienze dell'informazione. Non si può, infatti, descrivere il funzionamento degli esseri viventi senza usare concetti come memoria, messaggio, codificazione ecc.

Tuttavia, oltrepassando i confini della fisica, uscendo dal mondo delle macchine quantistiche e termodinamiche, la scienza non abbandona il mondo delle macchine naturali ma entra in un mondo di macchine naturali più ampio. Un mondo nel quale, a dispetto del fatto che a scala macroscopica le macchine funzionano in modo praticamente deterministico, l'ideale fisico della predittività teorica incontra i suoi limiti computazionali. Dal punto di vista delle possibilità di rappresentazione matematica, questo è il mondo delle procedure indecidibili, dei processi ricorsivi dagli esiti logicamente imprevedibili, delle verità indimostrabili, della generazione di oggetti di complessità indescrivibile. L'analisi teorica dei fenomeni che trovano solo in questo più ampio campo interpretativo la spiegazione della loro complessità si presenta principalmente come teoria delle macchine calcolatrici universali di Turing e degli automi costruttori universali di von Neumann. Questo più ampio dominio della scienza moderna ci permette non solo di spiegare la complessità strutturale degli esseri viventi, ma anche d'interpretare i loro processi e i loro comportamenti come processi e comportamenti di macchine materiali.

Si pone allora in modo naturale la domanda se questo sapere teorico così esteso possa arrivare al punto di rendere possibile la costruzione di organismi viventi artificiali diversi da quelli prodotti dalla natura e persino di macchine artificiali pensanti.

A questo proposito penso che, sebbene la scienza e la tecnologia siano ancora molto distanti da questa possibilità, non si possa escludere a priori che i futuri progressi non permetteranno mai di raggiungere tali obiettivi. L'attendibilità di una prospettiva di questo genere si basa sul fatto che il progresso scientifico-tecnologico è esso stesso un processo ricorsivo capace di raggiungere livelli di complessità indescrivibile in modo esponenzialmente crescente nel corso del tempo, per cui non già gli esseri umani direttamente con le loro mani e i loro cervelli, ma gli esseri umani con le loro macchine sempre più complesse e i loro calcolatori sempre più potenti potranno creare generazioni di macchine ancor più potenti, fino a raggiungere la possibilità di dirigere la costruzione di sistemi materiali di complessità confrontabile a quella degli esseri viventi. Se questa meta, per ora soltanto fantascientifica, venisse realmente raggiunta, l'umanità si troverebbe nelle condizioni di propagare le nuove forme di vita e d'intelligenza in mondi lontani dalla terra.