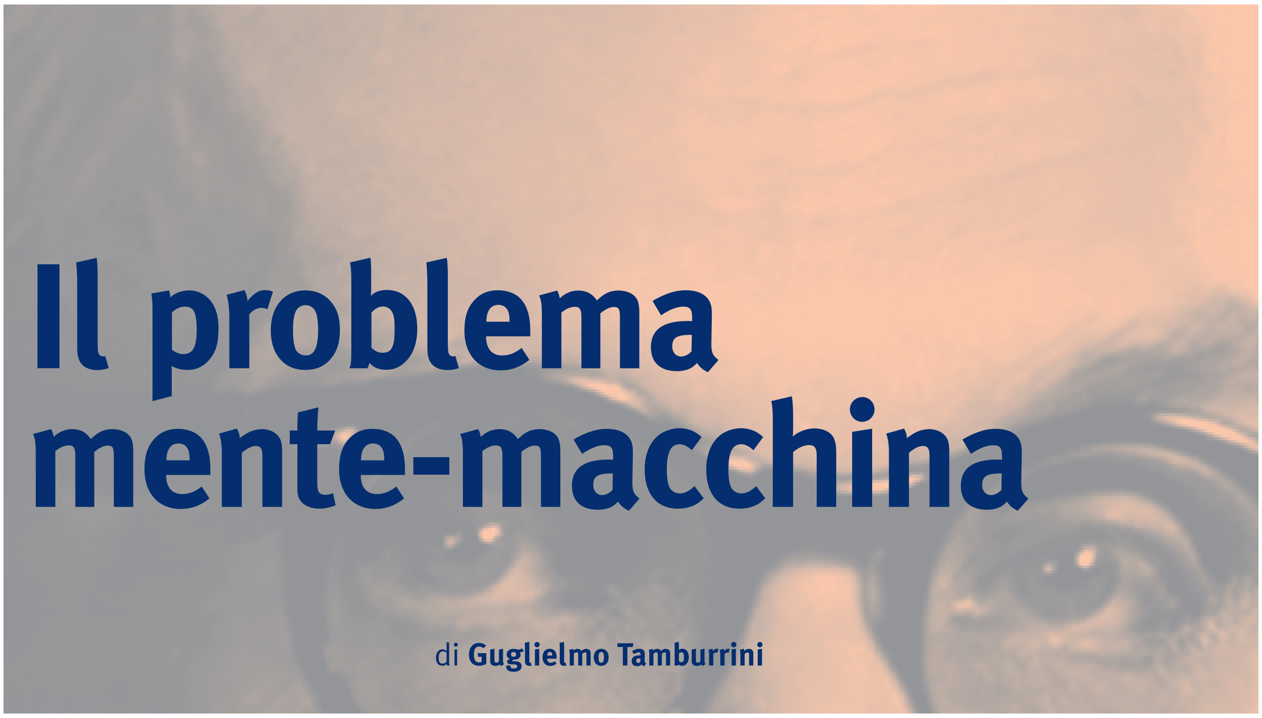

Il problema mente-macchina

di **Guglielmo Tamburrini**

Nel corso degli ultimi venti anni, le nostre conoscenze sulle idee di Kurt Gödel intorno al problema *mente-macchina* si sono notevolmente arricchite. Dapprima, attraverso le testimonianze pubblicate da Hao Wang e da altri logici che con Gödel hanno discusso i problemi filosofici suscitati dai teoremi d'incompletezza; in seguito, attraverso la pubblicazione di vari scritti inediti nel terzo volume dei *Collected Works* (Kurt Gödel, *Collected Works*, voll. I-V, a cura di S. Feferman et al., Oxford UP, d'ora innanzi *CW* I-V) e dell'epistolario nei volumi successivi. La rivista *Philosophia Mathematica* ha dedicato nel 2006 un fascicolo monografico alle riflessioni filosofiche di Gödel sulla Matematica e Logica (vol 14, ottobre 2006), che comprende vari contributi (S. Feferman, P. Koellner, R. Tieszen, W. Sieg, M. van Atten) su *incompletezza* e *problema mente-macchina*, nei quali si tiene conto di queste nuove acquisizioni.

Facendo leva sulla formulazione generale dei teoremi d'incompletezza resa possibile dalla *tesi di Church-Turing*, Gödel ha sostenuto una posizione "antimeccanicistica" intorno alle capacità matematiche della mente umana, la quale non sarebbe soggetta alle medesime limitazioni dei sistemi formali. Questa visione del problema mente-macchina è strettamente legata a una forma di dualismo ontologico tra mente e corpo: nessun sistema fisico può essere utilizzato, secondo Gödel, per superare le limitazioni dei sistemi formali. Le poche osservazioni di Gödel a sostegno di questa tesi di "impossibilità fisica" rivestono un interesse che è indipendente dal problema mente-macchina e dal problema mente-corpo: le sue osservazioni risultano infatti rilevanti per una riflessione critica sull'ipotesi che sistemi di calcolo analogici "superTuring" siano fisicamente possibili. Questa ipotesi è stata formulata e discussa anche in relazione a risultati recenti sulla classificazione di modelli di sistemi dinamici continui o ibridi (che combinano dinamica discreta e continua) in base alle capacità di calcolo dei sistemi descritti da tali modelli.

Ma procediamo con ordine, cominciando dalla versione generale dei teoremi di incompletezza, che costituisce il punto di partenza delle riflessioni gödeliane sul problema mente-macchina.

Teoremi d'incompletezza e Tesi di Church-Turing (1931-1936)

Come suggerisce anche il titolo dell'articolo originale di Gödel del 1931 (*Su proposizioni formalmente indecidibili dei 'Principia Mathematica' e di sistemi affini*), i teoremi d'incompletezza investono, nella loro formulazione iniziale, una particolare famiglia di sistemi formali.

Il problema di pervenire a una versione generale dei teoremi d'incompletezza, valida per *tutti* i sistemi formali che siano coerenti ed adeguati per l'Aritmetica elementare, fu affrontato e risolto negli anni immediatamente successivi, con il contributo dello stesso Gödel e di altri logici (tra i quali spiccano Alonzo Church, Alan Turing, Stephen Kleene ed Emil Post). Per pervenire a tale risultato, questo gruppo di ricercatori attaccò il problema di fornire una definizione generale e uni-

L'autore

Guglielmo Tamburrini insegna Epistemologia presso il Dipartimento di Scienze Fisiche dell'Università Federico II di Napoli. I suoi interessi di ricerca riguardano principalmente l'esplorazione delle neuro-scienze cognitive, la Filosofia e la Storia dell'Intelligenza artificiale e della Robotica.

voca della nozione di *funzione calcolabile* mediante un algoritmo. Una tale definizione avrebbe a sua volta permesso di fornire una definizione univoca e precisa di una proprietà essenziale dei sistemi formali, cioè della richiesta che l'insieme delle formule e l'insieme delle dimostrazioni di un sistema formale siano decidibili mediante un procedimento algoritmico. Ecco come Kleene presenta e commenta una tale richiesta:

> "[In un sistema formale] avremo delle regole che specificano, solo a partire dalla loro forma come oggetti linguistici, quali successioni di simboli siano formule e quali successioni di formule siano dimostrazioni (dell'ultima formula della successione)
> … poiché queste regole devono fornire definizioni complete, pienamente utilizzabili delle classi delle formule e delle dimostrazioni, esse devono rendere disponibili algoritmi da applicare agli oggetti simbolici per rispondere alle classi di domande della forma: "Questa successione di simboli è una formula?" e "Questa successione di formule è una dimostrazione?" Fornire, nel processo di articolazione di un sistema formale, tali algoritmi è una caratteristica dei sistemi formali essenziale agli scopi della formalizzazione."
> (S.C. Kleene, "Turing's analysis of computation, and major applications of it" in *The Universal Turing Machine, A Half-Century Survey*, a cura di R. Herken, Oxford UP, 1986, p. 40.)

Il lavoro di questo gruppo di logici sfociò nella proposta di identificare la classe delle funzioni parziali calcolabili mediante un algoritmo arbitrario con la classe delle funzioni parziali calcolabili da una macchina di Turing (o, equivalentemente, parziali ricorsive). La proposta, avanzata indipendentemente da Alan Turing e Alonzo Church in lavori sviluppati e pubblicati intorno al 1936, è oggi nota come *tesi di Church-Turing*. Poiché la classe delle funzioni parziali calcolabili da una macchina di Turing ha una definizione precisa (al contrario della classe delle funzioni parziali calcolabili mediante un non meglio specificato procedimento algoritmico), la *tesi di Church-Turing* consente di precisare il concetto di sistema formale in maniera tale da pervenire alla dimostrazione di una versione pienamente generale dei teoremi di incompletezza, valida per ogni sistema formale coerente che contenga l'Aritmetica elementare.
Per una discussione più ampia sul ruolo della *tesi di Church-Turing* per le versioni generali dei teoremi d'incompletezza e d'indecidibilità, si veda anche G. Tamburrini, *I matematici e le macchine intelligenti*, Milano, Bruno Mondadori 2002, pp. 51-65.

Mente e macchina

In una breve Nota del 1965 (*CW*, I, pp. 369-70), Gödel osserva:

> "Grazie al lavoro di A. M. Turing, si può ora dare una definizione precisa e indubitabilmente adeguata del concetto generale di sistema formale e si può dimostrare l'esistenza di proposizioni aritmetiche indecidibili e la non dimostrabilità della coerenza di un sistema nello stesso sistema per ogni sistema formale coerente che contiene una certa quantità di Aritmetica finitista."

Gödel aveva reso un analogo riconoscimento a Turing all'inizio di una conferenza che tenne il 26 dicembre 1951 presso la *Brown University*, nell'ambito di una serie di conferenze organizzate dall'*American Mathematical Society* in onore del matematico Josiah Willard Gibbs. Il testo della *Gibbs Lecture*, rielaborato a più riprese e tuttavia mai pubblicato da Gödel, è apparso in *CW* III con il titolo *Some basic theorems on the foundations of mathematics and their implications*:

> "[il lavoro di vari matematici ha dato [a questi teoremi] una forma molto più soddisfacente di quella originaria. Al miglioramento più importante si è pervenuti attraverso la definizione precisa del concetto di procedura finita, che gioca un ruolo decisivo in questi risultati. Vi sono vari modi di arrivare a una tale definizione che, peraltro, portano tutti esattamente allo stesso concetto. Il modo più soddisfacente, secondo la mia opinione, è quello di ridurre il concetto di procedimento finito a quello di una macchina con un numero finito di parti, così come ha fatto il matematico britannico Turing."

È opportuno osservare che Gödel usa l'espressione *procedimento finito* come sinonimo di *algoritmo*, e anche di *procedimento di calcolo* e di *procedimento meccanico*. Ed è anche opportuno rilevare un problema critico riguardo alla valutazione gödeliana del lavoro di Turing: Gödel sopravvaluta la portata reale dell'analisi di Turing, ritenendo il problema di fornire argomenti a sostegno della *tesi di Church-Turing* sostanzialmente risolto con il lavoro di Turing. Vediamo perché.
La *tesi di Church-Turing* riguarda le funzioni calcolabili mediante un qualsiasi processo di calcolo, senza porre distinzioni fra possibili classi di esecutori, mentre l'analisi di Turing si propone di mostrare che i processi di calcolo *eseguibili da un particolare tipo di agente* – l'essere umano – possono essere simulati da una macchina di Turing. Il problema che certamen-

te non è stato risolto dall'analisi di Turing è quello di fornire argomenti a sostegno della *tesi di Church-Turing*, i quali risultino convincenti anche in relazione a processi di calcolo eseguibili da agenti diversi dall'essere umano. Questo punto è stato sottolineato da Robin Gandy che, di Turing, è stato allievo diretto ed amico personale:

> "Alcuni degli argomenti di Turing si possono applicare indifferentemente a uomini e macchine; vi sono però passaggi fondamentali della sua analisi nei quali bisogna valersi del fatto che l'esecutore del processo di calcolo sia un essere umano. Si ricorre a questo fatto per giustificare l'assunto che il calcolo si svolge attraverso una successione di passi elementari. Un essere umano può scrivere solo un simbolo alla volta. Se, tuttavia, si astrae da limitazioni di carattere pratico, è possibile concepire una macchina che stampa simultaneamente un numero arbitrariamente grande di simboli."
> (R. Gandy, "Church's thesis and principles for mechanisms" in *The Kleene Symposium*, a cura di J. Barwise, North-Holland, 1980, pp. 124-125)

Alan Turing

Gandy ha affrontato questo problema attraverso un'analisi dei processi di calcolo eseguibili da agenti computazionali arbitrari. Per un'analisi critica del punto di vista di Gödel sulla *tesi di Church-Turing* basata sugli sviluppi del lavoro di Gandy, si rimanda il lettore interessato al saggio di Wilfried Sieg *Gödel and computability*, apparso nel fascicolo monografico di *Philosophia Matematica* menzionato all'inizio e ai riferimenti bibliografici ivi contenuti. Per una trattazione in lingua italiana, si rimanda a Tamburrini, *op. cit.*, pp. 59-65.
Comunque si giudichi la solidità degli argomenti a sostegno della *tesi di Church-Turing* (seguendo Gödel oppure Gandy, per esempio), è un fatto che la tesi rende possibile la dimostrazione della versione generale dei teoremi d'incompletezza. Gödel utilizza questa versione generale nella prima parte della *Gibbs Lecture*, per impostare il problema dell'inesauribilità o dell'incompletabilità della Matematica, intesa come totalità delle proposizioni matematiche vere. In particolare, Gödel discute l'impossibilità di abbracciare la Matematica con un singolo sistema formale e pone la questione se la Matematica sia in linea di principio inesauribile anche per la mente umana. In questo contesto problematico, si chiede se la mente umana, per quanto concerne le sue capacità dimostrative in Matematica, sia estensionalmente equivalente a un qualche siste-

ma formale S ovvero alla macchina di Turing che si può specificare a partire da una descrizione di S e che enumera l'insieme dei teoremi di S.
Se la mente umana è equivalente a un tale sistema formale (o alla macchina di Turing ad esso associata), allora – sostiene Gödel – esistono problemi *diofantei* che la mente umana non sarà mai in grado di risolvere. In altre parole, dalla suddetta equivalenza deriva l'esistenza di problemi matematici di un tipo particolare che risultano essere *assolutamente* indecidibili per la mente umana. Questa conclusione è

Alonzo Church

resa particolarmente evidente – secondo Gödel – dal secondo teorema d'incompletezza, che asserisce la non dimostrabilità della proposizione che esprime la coerenza di un sistema formale S all'interno dello stesso sistema S, per ogni S che soddisfa le ipotesi del primo teorema d'incompletezza. Tale proposizione ha la forma $\forall(x)R(x)$, laddove R è un predicato primitivo ricorsivo. Se si assume che la mente umana usi esclusivamente regole d'inferenza corrette e assiomi veri che includono gli assiomi di Peano per l'Aritmetica elementare, allora le condizioni dei teoremi d'incompletezza risultano verificate per il sistema S che si ipotizza equivalente alla mente umana. Pertanto, la proposizione che esprime la coerenza di S non è dimostrabile all'interno di S e, ancora per l'equivalenza tra mente umana ed S, questa proposizione non è nemmeno dimostrabile dalla mente umana. In *CW* III, è stato pubblicato un manoscritto nel quale Gödel dimostra che ogni enunciato della forma $\forall(x)R(x)$, con R primitivo ricorsivo, è equivalente a un enunciato della forma $\forall x_1, \ldots, \forall x_n \exists y_1, \ldots \exists y_m\, [p(x_1, \ldots, x_n, y_1, \ldots, y_m) = 0]$, laddove le variabili sono del tipo "numero naturale" e p è un polinomio a coefficienti interi. Questo risultato, migliorato da Martin Davis, Hilary Putnam, Julia Robinson e Yuri Matiyasevich (ponendo m=0 e sostituendo "=" con "≠") nel corso delle ricerche che hanno portato alla soluzione del decimo problema di *Hilbert* (vedi S. Feferman in *Philosophia Mathematica* 14(2), p. 139), giustifica il riferimento di Gödel al carattere diofanteo dei problemi assolutamente indecidibili in questione.
Se viceversa la mente umana non è equivalente ad alcun sistema formale S, allora per ogni tale sistema vi sarà un enunciato dimostrabile dalla mente umana ma non dimostrabile in S.
E' opportuno notare che, per ovvie proprietà degli insiemi che, come l'insieme dei teoremi di S, sono r*icorsivamente enumerabili* (ovvero enumerabili da una macchina di Turing), sotto la seconda alternativa e cioè l'ipotesi di non equivalenza tra mente e sistema formale, la mente umana dovrebbe necessariamente essere in grado di dimostrare un'infinità di proposizioni. Ma allora è naturale chiedersi che cosa ci sia di genuinamente umano in una mente dotata di tali capacità infinite!
La difficoltà che emerge a tale riguardo consiste sostanzialmente nel fatto che la mente "umana", alla quale si attribuiscono capacità dimostrative diverse da ogni calcolatore, è un'entità fortemente idealizzata, dotata di capacità cognitive superiori a quelle che siamo normalmente disposti ad attribuire a una singola mente umana o anche a una non meglio specificata collettività di menti umane. È necessario che una tale entità abbia a disposizione risorse illimitate di tempo e di spazio per eseguire il proprio lavoro dimostrativo e registrarne i risultati. Occorre dinque introdurre un'idealizzazione, almeno rispetto alle risorse di spazio e di tempo, che è analoga a quella incorporata nella nozione di macchina di Turing, anche solo al fine di comprendere il quesito posto da Gödel: "*la mente umana è soggetta oppure no alle limitazioni matematiche dei sistemi formali e dei calcolatori, che sono state messe in luce dai teoremi d'incompletezza?*" In particolare, una siffatta idealizzazione è richiesta per comprendere che cosa significhi rispondere "*no*" a questa domanda, indipendentemente da quali ragioni si possano poi addurre a sostegno di tale risposta.
Gödel non vacilla di fronte alla necessità di queste idealizzazioni e ne introduce altre, che risultano essere ancora più problematiche da un punto di vista epistemologico. In alcune brevi osservazioni a sostegno della risposta "no" (CW II, pp. 305-307), ipotizza che la mente umana potrebbe essere in grado, usando un qualche procedimento non-algoritmico che richiede la comprensione del significato di stringhe simboliche, di generare e di riconoscere la verità di una successione illimitata di assiomi dell'infinito. Questa ipotesi, largamente speculativa, richiede un'altra idealizzazione che non

ha termini di paragone nella teoria delle macchine di Turing, al contrario delle precedenti idealizzazioni relative alle risorse di tempo e di spazio. Gödel, infatti, attribuisce alla mente umana la capacità di comprendere il significato e di emettere un giudizio corretto sulla verità di assiomi che formano un insieme potenzialmente infinito di enunciati.
Le osservazioni "antimeccanicistiche" di Gödel a sostegno della risposta "*no*" sono state criticate in modo puntuale, attraverso una riflessione sullo stato delle ricerche nell'ambito della Teoria degli insiemi (si veda il citato articolo di Feferman e i riferimenti bibliografici ivi contenuti). Ma queste osservazioni risultano problematiche anche per motivi di carattere più generale. Non vi sono, allo stato attuale, buone teorie o spiegazioni scientifiche per un qualche aspetto del fare Matematica che ci consentano di giudicare l'utilità delle idealizzazioni che è necessario introdurre per raccordare il problema mente-macchina ai teoremi di incompletezza. Sembra inevitabile imbattersi in simili problemi concettuali ed empirici nel tentativo di mostrare che le proposizioni matematiche vere dimostrabili dalla mente umana non sono identificabili con un sottoinsieme dell' insieme dei teoremi di un sistema formale. La consapevolezza delle difficoltà che affliggono i tentativi di utilizzare i teoremi d'incompletezza nel quadro della filosofia della mente non è forse estranea alla decisione di Gödel di non pubblicare il testo della *Gibbs Lecture*.

Mente e materia

Le osservazioni "antimeccanicistiche" di Gödel sono strettamente legate a una forma di dualismo ontologico mente-corpo.
Le testimonianze raccolte da Hao Wang (segnatamente nella monografia *From Mathematics to Philosophy*, Humanities Press 1974, e nell'articolo *On physicalism and algorithmism: can machines think?* in *Philosophia Mathematica*, vol. 2, 1993, ma si veda anche la prima parte della *Gibbs Lecture* in *CW* III) non sembrano lasciare dubbi al riguardo. Gödel riteneva un pregiudizio dei nostri tempi la tesi secondo la quale la mente non è separata dalla materia. Credeva, inoltre, che il progresso delle neuroscienze avrebbe portato a una confutazione di tale tesi, mostrando che "*non vi sono sufficienti cellule nervose per eseguire tutte le operazioni osservabili della mente*". Vi è un'ulteriore tesi gödeliana, secondo la quale "*il cervello funziona sostanzialmente come un calcolatore digitale*", la quale fornisce il collegamento tra la visione dualistica mente-corpo e le riflessioni sul problema mente-macchina discusse nel paragrafo precedente. In sostanza, la mente sarebbe in grado di superare le limitazioni dei calcolatori digitali sancite dai teoremi d'incompletezza, ma non in virtù del funzionamento del "suo" cervello, assimilabile al funzionamento di un calcolatore digitale.
È interessante chiedersi perché, secondo Gödel, il funzionamento del cervello sarebbe assimilabile a quello di un calcolatore digitale. (Problema che affrontò in modo più approfondito John von Neumann fornendo, pochi anni dopo la *Gibbs Lecture* di Gödel, una risposta diversa nel suo saggio *The Computer and the Brain*, trad. it. *Il calcolatore e il cervello*, in *La filosofia degli automi*, a cura di V. Somenzi e R. Cordeschi, Bollati Boringhieri 1994.)
Hao Wang afferma che questo convincimento affonda le sue radici in ulteriori riflessioni di Gödel sulla teoria fisica: le leggi fisiche, nelle loro conseguenze osservabili, hanno un limite finito di precisione.
Non è possibile ricostruire o esaminare criticamente in questa sede il collegamento tra questa tesi e il dualismo gödeliano e rimandiamo il lettore interessato alla stimolante ed articolata discussione del problema nel paragrafo 3 dell'articolo di Hao Wang in *Philosophia Mathematica*. È però opportuno rilevare, nel concludere questa Nota, che il limite finito di precisione con il quale è possibile controllare le leggi fisiche costituisce un ostacolo significativo per i tentativi finora compiuti di individuare un dispositivo fisico in grado di superare le capacità di calcolo delle macchine di Turing. Vi sono sistemi descritti da modelli dinamici continui espressi mediante equazioni differenziali ordinarie che hanno un comportamento ingresso-uscita dimostrabilmente *superTuring*, nel senso che tale comportamento non può essere simulato da nessuna macchina di Turing. Anche alcuni sistemi dinamici neurali (costituiti da reti neurali ricorrenti) hanno un comportamento dimostrabilmente *superTuring*. Sulla scorta di tali risultati, è stato ipotizzato che nel mondo fisico possano darsi sistemi di calcolo analogici *superTuring*. Ma questa congettura si scontra con il fatto che la realizzazione fisica di tali sistemi richiederebbe un funzionamento perfetto del calcolatore analogico o un processo di misurazione perfetto di alcuni valori reali: entrambe le condizioni non risultano essere fisicamente plausibili, come si argomenta in un recente articolo (G. Trautteur e G. Tamburrini, *A note on discreteness and virtuality in analog computing*, in corso di stampa su *Theoretical Computer Science C*) al quale si rimanda il lettore interessato. L'opinione di Gödel sull'impossibilità di un comportamento fisico *superTuring* è in pieno accordo con queste valutazioni epistemologiche di risultati, anche recenti, ottenuti in indagini matematiche su modelli non convenzionali di calcolo. ■