



UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO
DIPARTIMENTO DI RICERCA E INNOVAZIONE UMANISTICA
CORSO DI LAUREA IN SCIENZE FILOSOFICHE

TESI DI LAUREA IN
SCIENZE NEL MONDO CONTEMPORANEO

PENSIERO E CALCOLO: INDAGINE SU
FORMALISMO E RIDUZIONISMO NELLE SCIENZE
COGNITIVE TRA TURING, GÖDEL E
INTELLIGENZA ARTIFICIALE

Relatore:

Chiar.mo Prof. Francesco Paolo de Ceglia

Correlatore:

Chiar.mo Prof. Lorenzo Leporiere

Laureando:

Alessio Donvito

Anno Accademico 2022-2023

INDICE

INTRODUZIONE.....	1
I.....	5
ORIGINI LOGICO-MATEMATICHE DELL'ANALOGIA COMPUTAZIONALE	5
1.1. FONDAZIONALISMO MATEMATICO. HILBERT E L'ENTSCHEIDUNGSPROBLEM.....	5
1.1.1. <i>Preistoria della computabilità.....</i>	<i>5</i>
1.1.2. <i>Algoritmi, completezza, meccanizzazione.....</i>	<i>6</i>
1.2 RISOLVERE L'ENTSCHEIDUNGSPROBLEM: MACCHINE DI TURING.....	10
1.2.1. <i>Regole di calcolo. Verso una nuova analogia uomo-macchina.....</i>	<i>10</i>
1.2.2. <i>Stati mentali e configurazioni-macchina.....</i>	<i>13</i>
1.2.3. <i>Halting Problem.....</i>	<i>15</i>
1.3 INTELLIGENZA DELLE MACCHINE.....	17
1.3.1. <i>Su vaghezza della nozione di intelligenza e Turing Test.....</i>	<i>18</i>
1.3.2. <i>Obiezioni.....</i>	<i>19</i>
1.3.3. <i>Filosofia ed euristica della ricerca turinghiana.....</i>	<i>27</i>
1.3.4. <i>Quale riduzionismo?.....</i>	<i>29</i>
II.....	34
STRUMENTI DELL'ANALOGIA: IL PARADIGMA RIDUZIONISTICO E I SISTEMI FORMALI.....	34
2.1 NICOLAS RASHEVSKY: LA BIOLOGIA COME SISTEMA FORMALE.....	36
2.1.1. <i>Scienze teoretiche, scienze sperimentali.....</i>	<i>36</i>
2.1.2. <i>Sul trattamento formale dei fenomeni biologici.....</i>	<i>38</i>
2.1.3. <i>Sistemi complessi e riducibilità ai livelli inferiori.....</i>	<i>45</i>
2.2 MCCULLOCH E PITTS: LA COGNIZIONE COME CALCOLO LOGICO NEURONALE.....	47
2.2.1. <i>Atomi psicologici e altri problemi delle teorie riduzionistiche.....</i>	<i>50</i>
2.2.2. <i>Neuroni artificiali, reti logiche.....</i>	<i>54</i>
2.2.3. <i>Sull'equivalenza tra reti di McCulloch-Pitts e funzioni Turing-computabili... ..</i>	<i>56</i>
2.2.4. <i>Qualche problema ulteriore.....</i>	<i>58</i>
2.2.4.1. <i>Si fa presto a dire "calcolo".....</i>	<i>58</i>
2.2.4.2. <i>Ancora sull'equivalenza di reti logiche e macchine di Turing.....</i>	<i>59</i>
2.2.4.3. <i>Semplificare, astrarre, idealizzare.....</i>	<i>60</i>
2.2.4.4. <i>Reti sintattiche, reti semantiche.....</i>	<i>61</i>
2.2.4.5. <i>Uno scherzo formalistico.....</i>	<i>63</i>
2.2.4.6. <i>Sulla non-eccezionalità delle descrizioni formali.....</i>	<i>65</i>
III.....	66
EPISTEMOLOGIA DELL'IA COME MODELLO TEORICO DEI SISTEMI COGNITIVI.....	66
3.1 GENESI CONCETTUALE DEL PARADIGMA COMPUTAZIONALE.....	66
3.2 FORME D'INFERENZA.....	68
3.2.1. <i>Logic Theorist: formalizzabilità in senso proprio.....</i>	<i>69</i>
3.2.2. <i>Induzione e abduzione: formalizzabilità in senso analogico.....</i>	<i>71</i>
3.3 EFFETTI DEL PARADIGMA FORMALISTICO-RIDUZIONISTICO NELLE SCIENZE COGNITIVE.....	78
3.3.1. <i>Pars construens.....</i>	<i>79</i>
3.3.2. <i>Pars destruens.....</i>	<i>82</i>
3.4 RIDUZIONISMO E ANTI-RIDUZIONISMO IN OTTICA OPERATIVA.....	85
CONCLUSIONI.....	89
BIBLIOGRAFIA.....	94

INTRODUZIONE

In matematica e in informatica è detta “teoria della computabilità” (o della calcolabilità, o della computabilità effettiva) la disciplina che si occupa di stabilire l’esistenza di algoritmi per il calcolo automatico di classi specifiche di funzioni.

Essa valuta la loro calcolabilità di principio, posta l’esistenza di problemi e funzioni non calcolabili¹. Cruciali per la teoria sono le definizioni formali (ovvero matematiche) di computabilità e di algoritmo; l’intreccio dei concetti in questione ha rappresentato l’atto di nascita degli studi informatici.

Nelle pagine seguenti sarà nostro obiettivo ripercorrere alcune delle tappe fondamentali dell’impresa informatica novecentesca, mostrandone i protagonisti e – per quanto ci è consentito – i principali sviluppi. Questo lavoro fungerà da base per l’esplorazione teorica delle molte propaggini logico-formali ed epistemologiche che si diramano dal nucleo centrale dell’informatica come scienza.

Si mostrerà, in particolare, la stretta provenienza di certi programmi di ricerca dell’informatica tardo-novecentesca e contemporanea (tra tutti, quelli concernenti lo sviluppo di intelligenze artificiali²) dai primi passi compiuti nel campo della teoria della computabilità effettiva.

Per gli stessi motivi, si potrà osservare la correlazione esistente tra la nascita dell’informatica in generale e il grande fermento di studi logici e matematico-fondazionali a cavallo tra fine Ottocento e inizio Novecento. In questo senso specifico si potrà pensare all’informatica come a un esito non atteso della matematica e della logica formale del nuovo secolo.

Il vasto successo riscosso dall’informatica a partire dalla metà del Novecento ha contribuito alla nascita di un intero campo – di carattere, per così dire, extra-

¹ Un esempio di problema insolubile è il celebre *Entscheidungsproblem* hilbertiano, del quale parleremo ampiamente più avanti.

² Da ora in avanti IA o AI, a seconda che si abbrevi il termine in italiano o in inglese.

tecnico – di studi ad essa correlati. È venuta così delineandosi una vera e propria filosofia dell'informatica, strutturata secondo una vasta congerie di ramificazioni sottostanti, alcune di natura ontologica, altre a forte connotazione etica e altre ancora – come è stato anticipato – a carattere logico ed epistemologico: queste ultime direttamente influenzate dalle questioni aperte dalle filosofie della logica e della matematica di inizio Novecento.

Ricostruiremo, avvalendoci di tale evidenza, i processi sottesi alla nuova, più ampia e decisiva diffusione di concettualità afferenti alla sfera della riduzione e della formalizzazione dei sistemi complessi, con particolare riferimento ai processi cognitivi: dal lavoro di Alan Turing in avanti si è determinata una decisiva crescita di interesse nei confronti di questioni collocabili all'intersezione tra teoria della computabilità (e corollari logico-epistemologici) e filosofia della mente (a sua volta combinata con problemi e metodi di ricerca provenienti da linguistica, neuroscienze, psicologia e scienze cognitive in generale).

Entro l'intersezione appena prospettata si inquadra la ripresa di motivi teorici ampiamente attestati in età moderna (Hobbes, Leibniz, La Mettrie), rinvigoriti e dotati di strumenti nuovi dagli studi di Turing, Rashevsky, McCulloch e Pitts, dai progetti di riduzione e unificazione delle scienze perseguiti dai neopositivisti viennesi e ancora dalla nascita dell'IA come campo disciplinare autonomo: si produce così una nuova analogia mente-macchina, la quale assurge allo status di *standard* teorico almeno implicito, a sua volta informante lo sviluppo del discorso scientifico intorno ai processi della cognizione e alla loro replicabilità tecnica.

Rintraceremo nell'opera di Hilbert e poi di Turing le radici precipue dell'analogia computazionale e ne vedremo l'espansione che la fa crescere sino a determinare un paradigma computazionale vero e proprio. Si mostrerà la possibilità di fare ordine nel groviglio dei concetti ad esso sottesi mediante una distinzione tra approcci riduzionistici differenti, alcuni da intendersi in senso proprio, altri invece di segno operativo o debole.

Nel secondo capitolo si presentano due esperienze scientifiche cruciali per l'affermazione del paradigma computazionale, quelle di Nicholas Rashevsky

– padre della cosiddetta “biofisica matematica” – e di Pitts e McCulloch, autori di una teoria formale dell’attività neurale costruita sulla logica del primo ordine. I risultati che conseguono offrono uno spunto prezioso per comprendere quali siano i meccanismi all’opera nei processi di riduzione della complessità di un fenomeno, di cui l’uso più o meno “largo” dei formalismi matematici è strumento principe.

Si vedrà che le descrizioni riduzionistiche e formalizzanti dei fenomeni complessi operano secondo uno schema di astrazione, riduzione, idealizzazione e – alle volte – introduzione nella teoria di enti e concetti sviluppati *ad hoc* per la sua tenuta descrittiva.

Si otterranno così informazioni fondamentali sullo statuto dei formalismi matematici e sulle possibilità di una loro estensione in ambiente “extra-formale”; in particolare si guadagna un principio di “non-eccezionalità” delle descrizioni formali che si rivela ausilio prezioso nell’operazione di discriminare applicazioni proprie ed applicazioni improprie degli approcci riduzionistici e formalizzanti.

Nel terzo, infine, si sfruttano le evidenze ottenute per sondare le conseguenze dell’uso del paradigma computazionale nel campo delle scienze cognitive. Ne distinguiamo di costruttive e di distorsive, verificando in ultima istanza che il successo delle procedure di riduzione e formalizzazione non può essere valutato in termini assoluti, ma deve interpretarsi in relazione al particolare livello operativo della descrizione teorica che si costruisce.

Quando una descrizione riduzionistica e formalistica opera entro i limiti del principio di non-eccezionalità, il suo uso è tendenzialmente vantaggioso, pur manifestando i limiti che le abbiamo riferito più sopra; quando – invece – tali limitazioni sono oltrepassate e se ne tenta un’applicazione “uni-livellare” ne risulta una compressione inefficace della complessità del fenomeno, la quale esclude molti suoi elementi fondamentali dal contesto della trattabilità scientifica.

Il bilanciamento tra vantaggi e distorsioni dell’uso delle descrizioni formali risulterà riflettere la distinzione tra riduzionismi di diverso segno introdotta più

sopra; si individuerà pertanto un criterio di demarcazione nel concetto di *operatività* della descrizione.

I.

ORIGINI LOGICO-MATEMATICHE DELL'ANALOGIA COMPUTAZIONALE

1.1. Fondazionalismo matematico. Hilbert e l'*Entscheidungsproblem*

1.1.1. Preistoria della computabilità

Come chiarito in sede di introduzione, la nostra indagine sugli incroci di filosofia della scienza e informatica contemporanee comincia da alcuni grandi problemi di natura logico-matematica. Entro questo campo, infatti, hanno preso a svilupparsi metodi e concetti cruciali per la nascita successiva dell'informatica come scienza. Di simili esiti i primi contributori non potevano – come è facile immaginare – avere alcuna cognizione.

Sebbene la storia dei concetti di calcolo e di algoritmo sia assai lunga e articolata³, gli intenti della nostra ricerca ci autorizzano a scegliere come punto di partenza il lavoro di David Hilbert: al matematico tedesco risale, infatti, la prima formulazione di un problema per la decidibilità di proposizioni nei sistemi matematici.

Nel corso della sua presentazione al Congresso internazionale di Parigi del 1900, Hilbert sottopose all'attenzione dei colleghi una celebre lista di grandi problemi matematici per il secolo appena iniziato.

Il secondo dei ventitré totali esige una dimostrazione di coerenza per gli assiomi dell'aritmetica⁴. Come si vedrà, il computazionalismo filosofico (qui

³ Cfr. M. DAVIS, *Il calcolatore universale*, Adelphi, Milano, 2012. Il matematico statunitense ne propone una ricostruzione che parte dal sogno leibniziano della *characteristica universalis* e ripercorre le principali tappe che conducono ai primi tentativi espliciti di costruire una macchina computazionale universale (di cui furono protagonisti, tra gli altri, Turing e von Neumann). L'opera mostra – benché con atteggiamento divulgativo – l'estrema complessità teorica alla quale simili tentativi hanno dato origine: emerge l'immagine di una vicenda all'incrocio tra matematica, logica, filosofia e saperi tecnici.

⁴ Il problema fu risolto da Kurt Gödel, con risposta negativa, nel 1931. Il logico austriaco pubblicò allora l'articolo *Über formal unentscheidbare Sätze der «Principia mathematica» und verwandter Systeme*, con cui dimostrava che la coerenza di PM (l'aritmetica espressa dai *Principia Mathematica* di Russel e Whitehead) non era dimostrabile entro PM stessa, ovvero che in un sistema coerente di assiomi esiste sempre almeno una proposizione indecidibile dentro il sistema.

inteso come diretta conseguenza delle maggiori acquisizioni della teoria matematica della computabilità) fiorì precisamente nell'ambito del dibattito matematico-fondazionale.

Davis scrive:

«Il problema della coerenza dell'aritmetica era ormai una piaga in suppurazione, ma solo negli anni Venti Hilbert decise di aggredirlo direttamente insieme ai suoi allievi e seguaci – con conseguenze che nessuno di loro era in grado di prevedere»⁵.

Occorre precisare che le conseguenze impreviste cui Davis si riferisce non riguardano immediatamente la comparsa della teoria computazionale della mente, ma soltanto la nascita “accidentale” della scienza informatica a partire dal tentativo di risolvere alcuni dei problemi della metamatematica hilbertiana.

Esiste pertanto una linea diretta di contatto tra il fondazionalismo e il richiamo al rigore logico dell'aritmetica, da una parte, e il lavoro di analisi e formalizzazione del calcolo in quanto processo mentale, dall'altra.

Questa particolare relazione gode di ampio riconoscimento nella letteratura scientifica; molto meno conosciuta è la dipendenza delle filosofie computazionaliste – nelle loro diverse formulazioni – dalla tradizione di studi citata poc'anzi. Cercheremo di fare luce su entrambe le connessioni.

Riprendiamo da Hilbert. Lo studioso tedesco aveva collocato il problema della coerenza dell'aritmetica ai primi posti tra le urgenze matematiche del Novecento già all'inizio del secolo. Cominciò, tuttavia, ad occuparsene sistematicamente all'incirca un ventennio più tardi. Con lui lavorarono John von Neumann, Paul Bernays e Wilhelm Ackermann.

1.1.2 Algoritmi, completezza, meccanizzazione

Insieme ad Ackermann pubblicò, per la prima volta nel 1928, un manuale di logica⁶ in cui si enunciavano due problemi cruciali: il primo consistente nella

⁵ M. DAVIS, Op. cit., p. 123.

⁶ D. HILBERT, W. ACKERMANN, *Grundzüge der theoretischen Logik*, Springer, Berlin, 1928.

richiesta di esibizione di una prova della *completezza* della logica del primo ordine⁷; il secondo chiedeva l'ottenimento di un algoritmo (ovvero di un insieme formalizzato e finito di istruzioni di calcolo) per determinare la validità di una qualsiasi formula della logica del primo ordine.

Quest'ultimo, noto come *Entscheidungsproblem* (problema della decisione), può essere espresso in diversi modi equivalenti: la sua formulazione originale chiedeva se esistesse «[...] una procedura per stabilire se un'espressione della logica del primo ordine fosse valida in ogni interpretazione possibile».⁸

Dalle dimostrazioni di Gödel in avanti si è ritenuto inteso che la formulazione hilbertiana fosse da considerarsi equivalente a quella riportata da Davis ed altri: trovare un algoritmo per determinare, date alcune premesse e una ipotetica conclusione scritte nel linguaggio della logica del primo ordine, se le regole del linguaggio medesimo consentano di derivare la seconda dalle prime⁹.

Il progetto di Hilbert e colleghi mostra un'ambizione considerevole: lo si può intendere come il tentativo di costruzione di un algoritmo dall'estensione e dalla potenza inedite il quale «[...] in linea di principio [...] avrebbe dovuto ridurre tutti i ragionamenti deduttivi umani a calcolo bruto, realizzando in buona misura il sogno di Leibniz».¹⁰

Mostrare – in altri termini – che la matematica è un sistema chiuso e governato unicamente da sistemi di regole: così può sintetizzarsi l'ambizione hilbertiana. Hilbert riteneva equivalenti la verità di un enunciato e la sua prova formale:

⁷ La logica del primo ordine, o anche logica predicativa, è – all'incirca – il linguaggio formale sviluppato da Gottlob Frege. I ragionamenti che può esprimere si avvalgono di connettivi logici ($\neg, \wedge, \vee, \rightarrow, \leftrightarrow$), quantificatori (\exists, \forall), funzioni e relazioni. Si dice “del primo ordine” perché i suoi quantificatori riguardano solo gli elementi dell'insieme di riferimento e non anche quelli dei suoi sottoinsiemi (si dicono “del secondo ordine” le logiche che quantificano sui sottoinsiemi). La *completezza* di un linguaggio formale, invece, esprime la possibilità di dimostrare o confutare, dentro la teoria, qualsiasi enunciato formulato mediante il linguaggio della teoria medesima (un sistema formale completo non presenta enunciati indecidibili). Per ogni precisazione ulteriore cfr. F. BERTO, *Logica da zero a Gödel*, Editori Laterza, Bari, 2008 e E. MENDELSON, *Introduzione alla logica matematica*, Bollati Boringhieri, Torino, 1977, da cui sono state tratte le definizioni.

⁸ M. DAVIS, op. cit., p. 295.

⁹ Oltre a M. DAVIS, op. cit., p. 190, si veda anche S. BARRY COOPER, J. VAN LEEUWEN, *On Computable Numbers, with an Application to the Entscheidungsproblem – A Correction*, in *Alan Turing, His Work and Impact*, pp. 13-115, Elsevier, 2013.

¹⁰ M. DAVIS, op. cit., p. 190.

basterà tracciarne i passi, e verificare che nessuna regola sia stata trasgredita, perché si dia conoscenza nuova.

L' *Entscheidungsproblem*, così ci pare di vedere, è filosoficamente rilevante in quanto portatore di una esplicita visione meccanicistica della matematica, nonché di una implicita visione meccanicistica dell'intelligenza: se l'attività del matematico non segue che regole formali, l'intelligenza stessa potrà spiegarsi come processo fondato su sistemi chiusi di regole – come macchina. Tale è, pressappoco, la prospettiva che guida i pionieri dell'idea originaria di Intelligenza Artificiale, evidentemente debitori del lavoro di Hilbert e colleghi. L'ingresso in campo dell'*Entscheidungsproblem* fa compiere al fondazionalismo matematico novecentesco un salto decisivo: da una richiesta di maggior rigore nell'aritmetica, soddisfacibile mediante l'uso del simbolismo logico, si approda al tentativo di meccanizzare le procedure per la dimostrazione di ogni relazione matematica e – più in generale – del ragionamento deduttivo *tout court*.

Non deve sorprendere, per questi motivi, che il programma di ricerca di Hilbert e dei suoi collaboratori possa aver suscitato riluttanza o, talvolta, esplicita avversione.

Andrew Hodges, ad esempio, riporta la reazione infastidita del matematico di Cambridge G. H. Hardy, secondo cui il tentativo sarebbe infine – senza alcun dispiacere – risultato vano:

«Naturalmente un teorema così non esiste; ed è una fortuna, perché se esistesse potremmo disporre di un insieme di regole meccaniche per la soluzione di tutti i problemi matematici, il che porrebbe fine alla nostra attività come matematici».¹¹

La preoccupazione manifestata da Hardy anticipa, per così dire, le impreviste linee di sviluppo del problema presentato da Hilbert e seguaci.

¹¹ A. HODGES, *Alan Turing: The Enigma*, New York, Simon & Schuster, 1983, trad. it. *Storia di un enigma. Vita di Alan Turing 1912-1954*, Bollati Boringhieri, Torino, 1991, p. 129.

Potremmo, in effetti, sintetizzare approssimativamente le attese del programma di Hilbert nei termini seguenti: in primo luogo, la completa assiomatizzazione della matematica, a partire dalla sua teoria fondamentale (l'aritmetica), può essere svolta; compiuto questo sforzo, si potranno individuare procedure di calcolo (o algoritmi) che – data una qualsiasi proposizione della matematica – stabiliscano attraverso un numero preciso e finito di passaggi se essa sia vera oppure falsa. Le proposizioni vere entro una particolare architettura logica sono dette valide: dunque il problema della decisione hilbertiano è un caso particolare del problema della validità, posto – come abbiamo visto – nel contesto della logica del primo ordine.

Nonostante l'evoluzione avversa alle aspettative degli studiosi tedeschi, i quali ritenevano di essere ormai prossimi alla soluzione dell'*Entscheidungsproblem*¹², i successivi sviluppi teorici si sono mossi proprio sulla linea dell'automazione del processo di calcolo (ovviamente nel campo, ancora da definire, della computabilità di principio e nozioni annesse, le quali saranno introdotte più avanti).

I lavori di Church, Kleene e Turing in questo ambito danno ragione ad Hardy: in seguito alla pubblicazione del cruciale articolo gödeliano del 1931 sull'indecidibilità dei sistemi formali si intuisce che il problema della decisione debba avere risposta negativa; naturalmente è necessaria una dimostrazione.

Gli intenti della nostra ricerca rendono essenziale un'analisi, ancorché rapida, delle principali dimostrazioni di insolubilità dell'*Entscheidungsproblem*. Esse costituiscono, infatti, il punto da cui prendono a dispiegarsi i processi che conducono all'elaborazione di un vero e proprio campo disciplinare della computabilità, e dunque, sul versante filosofico, alla nascita delle molte varianti della Teoria Computazionale della Mente.

Uno sguardo veloce al lavoro di Church e Turing ci fornirà, quindi, un quadro accurato del salto dal fondazionalismo logico-matematico, all'informatica, ai modelli di identificazione mente-calcolatore.

¹² Cfr. M. DAVIS, op. cit., p. 191: «[...] la lacuna fra quei casi particolari per i quali erano stati trovati degli algoritmi e i casi ai quali era stato ridotto il problema generale era diminuita a tal punto da far sperare che bastasse ormai un ultimo, piccolo passo a colmarla per intero, fornendo l'algoritmo cercato da Hilbert».

1.2 Risolvere l'*Entscheidungsproblem*: macchine di Turing

Sebbene Alonzo Church sia pervenuto per primo ad una dimostrazione negativa dell'*Entscheidungsproblem*, è conveniente partire dall'elaborazione turinghiana del modello teorico di calcolatore, per due ragioni: i risultati ottenuti dal matematico di Cambridge hanno immediatamente riscosso maggiore successo di quelli dell'omologo di Princeton e, soprattutto, l'analisi del processo di calcolo proposta da Turing ci introduce velocemente nel campo di maggiore rilievo per gli intenti che abbiamo dichiarato.

Si vedrà, infatti, che una embrionale teoria computazionale della mente è già implicitamente contenuta (o, quantomeno, facilmente ottenibile) nell'algoritmo che egli elabora.

1.2.1 Regole di calcolo. Verso una nuova analogia uomo-macchina

I risultati in questione sono esposti in un articolo, ormai celeberrimo, pubblicato nel 1936 e intitolato *On Computable Numbers, with an Application to the Entscheidungsproblem*¹³.

Turing apre con una definizione preliminare di numero computabile: sono computabili¹⁴ i numeri reali la cui espressione in decimali è calcolabile “con mezzi finiti”¹⁵.

Specifica, inoltre, che nonostante l'oggetto della pubblicazione siano i numeri computabili, si possono definire ed investigare mediante gli stessi metodi anche funzioni e predicati computabili¹⁶.

Si giunge così ad una prima definizione di interesse cruciale:

¹³ A. TURING., *On Computable Numbers, with an Application to the Entscheidungsproblem*, in “Proceedings of the London Mathematical Society”, serie seconda, 42, 1936, pp. 230-67; errata corrige in 43, 1937, pp. 544-46.

¹⁴ Intendiamo il termine come equivalente all'espressione “calcolabile mediante procedure meccaniche”.

¹⁵ A. TURING, op. cit., p. 230: «The computable numbers may be described briefly as the real numbers whose expressions as a decimal are calculable by finite means».

¹⁶ *Ibid.*

«According to my definition, a number is computable if its decimal can be written down by a machine»¹⁷.

Se è così, allora bisogna in primo luogo definire una macchina calcolatrice. Turing sceglie di procedere comparando una persona alle prese con l'esecuzione di un qualsiasi calcolo ad una macchina capace di assumere un numero finito di configurazioni q_1, q_2, \dots, q_R , che egli chiama “ m -configurazioni”.

L'identità persona-macchina rispetto alle capacità di calcolo, che nell'articolo del 1936 rimane implicita almeno fino a §9, deriva da un'astrazione compiuta proprio sugli algoritmi che regolano il processo di calcolo: esso, spogliato di ogni caratteristica inessenziale¹⁸ risulta in un insieme ben determinato di azioni di base, che possiamo immaginare – senza problema alcuno – svolte su un nastro unidimensionale suddiviso in celle, ognuna delle quali reca una cifra o il simbolo di un operatore matematico. Con le parole di Turing:

«In elementary arithmetic the two-dimensional character of the paper is sometimes used. But such a use is always avoidable, and I think that it will be agreed that the two-dimensional character of paper is no essential of computation. I assume then that the computation is carried out on one-dimensional paper, *i.e.* on a tape divided into squares»¹⁹.

L'operatore, per effettuare il calcolo, muove lo sguardo lungo il nastro avanti e indietro e scrive i risultati ottenuti di volta in volta prendendo in considerazione le cifre del calcolo da svolgere e gli operatori matematici di riferimento.

La scrittura di ogni nuovo simbolo nelle celle del nastro dipende – questo è un passaggio cruciale – non solo dai simboli su cui opera, ma anche dal suo “stato

¹⁷ *Ibid.*

¹⁸ Cfr. M. DAVIS, op. cit., p. 193. Sono da considerarsi elementi inessenziali per la definizione del processo di calcolo, per esempio: il mezzo di scrittura (penna o matita), il supporto utilizzato per compiere le operazioni (un foglio o un nastro diviso in celle orizzontali) e persino l'incolonnamento delle cifre.

¹⁹ A. TURING, op. cit., p. 249.

mentale”²⁰: deve, cioè, decidere se sommare, sottrarre, moltiplicare, dividere o compiere qualsiasi altra operazione.

Nell’esempio proposto da Davis il calcolo da effettuare è una semplice moltiplicazione: l’operatore scrive prima i due prodotti parziali, ottenuti moltiplicando le cifre a due a due; dopodiché prende a sommarli tra di loro e ottiene il risultato finale.

È necessario notare che le cifre sommate sono le stesse ricavate in seguito alla moltiplicazione: ciò che cambia è lo “stato mentale” dell’operatore, che gli suggerisce, lavorando su simboli, prima di moltiplicare e poi di sommare.

Ogni processo di calcolo – che si operi in aritmetica o su strutture algebriche, che riguardi il calcolo di derivate o di integrali – obbedisce sempre ad un insieme ristretto di regole:

- In ogni stadio del calcolo l’attenzione è rivolta solo a pochi simboli.
- In ogni stadio l’azione intrapresa dipende solo da quei simboli su cui si focalizza la sua attenzione e dal suo stato mentale del momento.²¹

L’elenco può essere esteso tenendo in debita considerazione le caratteristiche dell’esperimento teorico che stiamo definendo. Pertanto il processo di calcolo può essere sintetizzato nei termini seguenti:

- Viene eseguito scrivendo dei simboli nelle caselle di un nastro di carta.
- A ogni passo la persona che esegue il calcolo fa attenzione al simbolo scritto in una sola di queste caselle.
- L’azione successiva dipenderà da questo simbolo e dallo stato mentale della persona.

²⁰ *Ivi*, p. 194.

²¹ *Ivi*, p. 195.

- Tale azione consisterà nello scrivere un simbolo nella casella osservata ed eventualmente nello spostare l'attenzione sulla casella immediatamente a destra o a sinistra.²²

Le regole appena elencate definiscono il calcolo come un'operazione su simboli meccanicamente eseguibile.

1.2.2 *Stati mentali e configurazioni-macchina*

Ne ricaviamo la possibilità di sostituire il calcolatore umano con una macchina calcolatrice; è sufficiente che gli “stati mentali” del calcolatore vengano rimpiazzati da un insieme di m -configurazioni della macchina. La derivazione del formalismo che definisce la macchina dall'analisi dei processi del pensiero umano alle prese con il calcolo è confermata da Church:

«[...] in particular, a human calculator, provided with pencil and paper and explicit instructions, can be regarded as a kind of Turing machine»²³

Potrà essere utile osservare la trasposizione logico-formale di questa analisi, ossia vedere una macchina di Turing in azione, per capire fino a dove possiamo spingere l'identificazione tra pensiero e automatismi di calcolo.

In primo luogo, una considerazione essenziale: essendo una macchina di Turing (o *a-machine* – macchina automatica – nell'articolo del 1936) nient'altro che un insieme di regole di calcolo, esisteranno infinite macchine di Turing, una per ogni procedura o meccanismo che la definisce. Ognuna di esse opera sui simboli scritti su un nastro unidimensionale diviso in celle, muovendosi avanti e indietro a seconda della configurazione che la guida. Ad ogni istante scansiona un solo simbolo del nastro e – sempre sulla base della sua configurazione – ne scrive uno nuovo; prosegue spostandosi verso destra o verso sinistra, oppure continuando a scansionare la stessa cella.

²² *Ivi*, pp. 195-196.

²³ A. CHURCH, *Review: On Computable Numbers, with an Application to the Entscheidungsproblem*, *The Journal of Symbolic Logic*, Vol. 2, No. 1 (Mar., 1937), pp. 42-43.

Turing sviluppa una notazione simbolica utile a descrivere il comportamento delle macchine. Ogni azione da esse svolta è definita da una quintupla (ossia un insieme di cinque simboli) che esprime un enunciato.

Per esempio, la regola:

*Se, nella configurazione B, è scansionato il simbolo c, la macchina
sostituisce c con d, si sposta a sinistra e passa nella configurazione*

A

È sintetizzata dalla quintupla:

$$B c : d \leftarrow A$$

Se vogliamo che la macchina si sposti a destra, ci basterà cambiare direzione alla freccia:

$$B c : d \rightarrow A$$

Un enunciato che richieda alla macchina lo stesso comportamento ma senza alcuna traslazione di cella, invece, è espresso dalla quintupla:

$$B c : d \star A$$

Usando pochi simboli è possibile definire un insieme praticamente sconfinato di regole di calcolo. Basterà indicare il numero di configurazioni o di “stati mentali” necessari e far corrispondere, ad ognuno di essi, il comportamento della macchina quando scansiona ciascuno dei simboli che può incontrare sul nastro. Le tavole di quintuple che definiscono il comportamento della macchina sono – in termini più aggiornati – il programma della macchina stessa.

In §3 (*Examples of computing machines*) Turing prospetta qualche modello di macchina; i suoi esempi sono – tuttavia – leggermente più complessi di quello indicato da Davis, che si limita a definire la configurazione completa, ossia l’insieme: configurazione di partenza-simbolo scansionato-operazione della macchina-configurazione finale, di una macchina di Turing che distingue i numeri pari da quelli dispari.

Senza elencare tutte le quintuple che la costituiscono, sarà sufficiente comprendere il suo funzionamento in linea generale: dato un numero a più cifre scritto sul nastro, la macchina comincerà a scansarlo una cella alla volta, passando da uno stato all'altro e muovendosi verso destra man mano che riconosce un numero pari o un numero dispari.

Arrivato alla prima cella vuota (dunque alla fine della scansione del numero) stamperà uno 0 se si trova nello stato cui giunge in seguito alla scansione di un numero pari o un 1 se si trova nello stato cui giunge in seguito alla scansione di un numero dispari²⁴.

Potremmo domandarci quale sia la connessione tra questi risultati e la gamma di problemi da cui il nostro discorso ha preso le mosse.

In effetti, come suggerisce il titolo assegnato da Turing al suo articolo, l'analisi e formalizzazione del concetto di computabilità mediante l'introduzione dell'idea di calcolo automatico gode di una particolare possibilità di applicazione all'*Entscheidungsproblem*.

Segnatamente, l'utilizzo della nozione di macchina formulata dal matematico di Cambridge consente di dimostrare l'insolubilità del problema della decisione. Hilbert cercava, come abbiamo visto, un algoritmo che decidesse la verità o falsità di qualsiasi enunciato matematico. Di conseguenza, dimostrando l'esistenza di un problema matematico non risolvibile algebricamente si proverebbe l'insolubilità complessiva dell' *Entscheidungsproblem*.

1.2.3 *Halting Problem*

In questo la nozione turinghiana di computabilità torna estremamente utile: posto che ogni procedura algebricamente computabile sia eseguibile da una macchina di Turing (è questo, approssimativamente, il contenuto della tesi di Church-Turing, secondo cui ogni problema calcolabile può essere risolto da una macchina di Turing), basterà mostrare che esiste finanche un solo problema che nessuna macchina risolve per ritenere provata l'inesistenza di un algoritmo come quello cercato da Hilbert.

Ancora con le parole di Turing:

²⁴ Cfr. M. DAVIS, op. cit., pp. 197-200.

«The simplest and most direct proof of this is by showing that, if this general process exists, then there is a machine which computes it»²⁵

Quello che si dimostra algebricamente insolubile nell'articolo del 1936 è il problema detto dell'arresto o della fermata (*Halting Problem*). Turing mostra l'impossibilità di costruire un algoritmo (*i. e.* una macchina di Turing o programma, in termini computazionali) che decida, prendendo come *input* un algoritmo o programma ulteriore, se esso terminerà oppure continuerà la sua esecuzione all'infinito. Se – come si dimostra – il problema dell'arresto è algebricamente insolubile, l'*Entscheidungsproblem* è automaticamente e identicamente dimostrato insolubile in modo complessivo.

Naturalmente la nostra presentazione del problema non rende conto dei molti tecnicismi matematici pur fondamentali per una comprensione efficace della dimostrazione. Ce ne smarchiamo per ovvie ragioni di sintesi e di attinenza all'argomento principale del discorso, ma ne rileviamo la relativa accessibilità, soprattutto nella sintesi proposta da Davis.

Per tutto ciò che è stato detto fino ad ora, è evidente che le macchine definite dal formalismo di Turing non necessitano di alcuna costruzione effettiva:

«[...] Dopotutto sono solo astrazioni matematiche; ciò che importa è che questa analisi della nozione di calcolo abbia consentito di stabilire che tutto ciò che è calcolabile mediante un processo algoritmico può essere calcolato da una macchina di Turing»²⁶.

Ad ogni modo, quello che qui ci interessa evidenziare è il tentativo di catturare i processi mentali sottesi al calcolo. La formalizzazione che approda alla costruzione della macchina ne è il prodotto fondamentale.

Sin dall'inizio, il concetto di “meccanismo” (nel senso turinghiano dell'automatismo computazionale) ha mirato a definire ciò che una mente umana fa quando esegue una procedura. Il risultato è la scomposizione del calcolo –

²⁵ A. TURING, op. cit., p. 246.

²⁶ M. DAVIS, op. cit., p. 196.

esteriormente visibile come processo unico e indiviso – in un insieme di operazioni atomiche elementari:

«Let us imagine the operations performed by the computer to be split up into “simple operations” which are so elementary that it is not easy to imagine them further divided»²⁷

Entriamo così in possesso di un numero sufficiente di elementi utili a proseguire la nostra indagine in direzione dei modelli di equiparazione mente-calcolatore.

1.3 Intelligenza delle macchine

Sino ad ora abbiamo fatto luce sulla connessione tra i problemi del fondazionalismo matematico e lo sviluppo dei primi programmi di ricerca in ambito informatico. Abbiamo potuto confermare – come suggerito all’inizio – la diretta (nonché fortuita) provenienza di tali programmi dal campo complesso e variegato della metamatematica otto-novecentesca; resta da analizzare il problema della derivazione di TCM e modelli affini dalle ricerche di Turing, Church e colleghi.

Al momento, infatti, ci siamo limitati ad abbozzare i pochi elementi che caratterizzano la presenza, in esse, di una tentata similitudine tra processi mentali e “programmabilità” meccanica. In altri termini, quanto stabilito sino ad ora mostra un tentativo ancora timido di spiegare i processi che costituiscono il pensiero umano in generale come del tutto formalizzabili, e pertanto eseguibili anche da una macchina opportunamente programmata.

On Computable Numbers non manifesta dirette ambizioni in questo senso: sebbene sia evidente l’origine della *a-machine* dalla modellizzazione dell’operato di un essere umano che lavora secondo un insieme di regole, alla metà degli anni ’30 Turing non ha ancora sviluppato i celebri argomenti sull’intelligenza delle macchine che caratterizzano il suo lavoro più tardo.

Il fulcro teorico dell’articolo riguarda – infatti – la definizione del concetto di computabilità, condotta mediante la formalizzazione delle procedure di calcolo

²⁷ A. TURING, op. cit., p. 250. Sebbene per noi sia difficile sfuggire all’immaginario del moderno calcolatore, il termine *computer*, nell’articolo del 1936, è da intendersi nel senso di calcolatore umano.

e l'introduzione della macchina, la presentazione di alcuni possibili modelli di *a-machine* – come si è detto più sopra – e si conclude con la dimostrazione di insolubilità dell'*Entscheidungsproblem*.

1.3.1 *Su vaghezza della nozione di intelligenza e Turing Test*

Di tutt'altro tenore è invece l'articolo intitolato *Computing Machinery and Intelligence*²⁸. La diversa estrazione dei due lavori si evince già dagli estremi di pubblicazione: l'articolo del 1936 compare negli atti della *London Mathematical Society*, mentre quello del 1950 è pubblicato in *Mind*, prestigiosa rivista di psicologia e filosofia.

La collocazione editoriale si riflette nell'eterogeneità contenutistica: *On Computable Numbers* è un articolo a forte caratterizzazione tecnico-matematica e per questo distinto da un rigore considerevole; *Computing Machinery and Intelligence* mostra diversi elementi di arbitrarietà e lascia aperti ampi margini di discussione critica.

Qui, le concettualità afferenti ad una possibile interpretazione dei fenomeni mentali come espressione di un sistema computazionale – elemento generale tipico delle diverse sotto-formulazioni della teoria computazionale della mente – sono largamente presenti. Una veloce analisi dell'articolo ce ne darà prova.

La discussione proposta da Turing prende le mosse da una celebre domanda sull'intelligenza delle macchine. È possibile ritenere che esse siano in grado di pensare? Per tentare di rispondere alla domanda, il matematico propone la formulazione di un altrettanto celebre esperimento, passato alla storia come Test di Turing. Siccome il problema richiede una definizione del significato dei termini “pensare” e “macchina”, bisogna sostituire la domanda di partenza con una seconda, di fatto equivalente ad essa ma espressa in termini meno vaghi:

«The new form of the problem can be described in terms of a game which we call the “imitation game”. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the

²⁸ A. TURING, *Computing Machinery and Intelligence*, *Mind*, vol. 59, n. 236, pp. 433-460, 1950.

game for the interrogator is to determine which of the other two is the man and which is the woman.»²⁹

Si può – a questo punto – immaginare di sostituire ad A una macchina. Entro la nuova condizione, l'interrogatore fallirà la scelta con la stessa frequenza? Riuscirà, in altri termini, a distinguere la macchina dall'uomo con lo stesso rapporto di fallibilità che caratterizza la scelta tra donna e uomo? Queste domande rimpiazzano quella di partenza.

Il test è utile – secondo il matematico – a tracciare una distinzione netta tra capacità intellettuali e capacità fisiche di un essere umano, in modo tale da stabilire una sorta di *fair play* per la macchina³⁰.

Nel caso in cui – come l'autore ritiene – una macchina opportunamente programmata riuscisse ad ingannare l'interrogatore, ad essa bisognerebbe riconoscere intelligenza.

Una posizione simile si presta evidentemente a molte obiezioni; Turing le prefigura e, entro una certa misura, tenta di rispondervi. Obiezioni e risposte relative hanno carattere marcatamente filosofico; analizzarle ci servirà a chiarire le posizioni del tardo Turing e – in seguito – a constatare la vicinanza ad esse dei modelli computazionali della mente.

1.3.2 Obiezioni

L'esame delle posizioni avverse è condotto dal matematico a partire da §6. È tuttavia preceduto da una definizione di computer digitale che ripresenta la coincidenza operativa tra intelletto umano e macchina:

«The idea behind digital computers may be explained by saying that these machines are intended to carry out any operations which could be done by a human computer»³¹

²⁹ *Ivi*, p. 433.

³⁰ Cfr. *Ivi*, p. 434: «The new problem has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man. [...] We do not wish to penalize the machine for its inability to shine in beauty competitions, nor to penalize a man for losing in a race against an airplane. The conditions of our game make these disabilities irrelevant».

³¹ *Ivi*, p. 436.

Più avanti è introdotta la nozione di “discretezza”. I calcolatori digitali possono essere considerati macchine a stati discreti. Esse operano saltando da uno stato definito all’altro³².

Si noterà che la trasposizione del concetto di discretezza nell’ambito dell’analogia mente-macchina suggerisce che, come avviene per il calcolatore, anche le operazioni dell’intelletto umano sono svolte secondo salti discreti tra stati mentali nettamente definiti.

Il problema, che qui citiamo *en passant*, meriterebbe una discussione specifica. Non potendola sostenere in queste pagine, rileviamo soltanto la possibilità di confronto tra questa tesi e l’ipotesi bergsoniana della continuità degli stati di coscienza³³.

In questo confronto, discretezza *versus* continuità, pare all’opera un effettivo scontro tra paradigmi contrastanti. Tuttavia, registriamo la complessità della questione anche nella discussione condotta da Turing, il quale introduce consapevolmente la nozione di discretezza solo per convenzione, in virtù della sua semplicità procedurale:

«Strictly speaking there are no such machines. Everything really moves continuously. But there are many kinds of machine which can profitably be *thought of* as being discrete-state machines. [...] There must be intermediate positions, but for most purposes we can forget about them»³⁴

In generale, lo vedremo, il matematico non sottovaluta affatto il peso dell’obiezione.

A partire dalla sesta sezione di *Computing Machinery and Intelligence*, dunque, si esaminano le posizioni contrarie a quella enunciata dall’autore, la quale è esplicitamente comunicata nello spirito – per così dire – di una auspicata chiarezza metodologica: Turing ritiene plausibile la costruzione – entro la fine del secolo – di macchine capaci di ingannare una percentuale consistente di

³² Cfr. *Ivi*, 439: «The digital computer considered in the last section may be classified amongst the “discrete-state machines”. These are the machines which move by sudden jumps or clicks from one quite definite state to another».

³³ Cfr. H. BERGSON, *Saggio sui dati immediati della coscienza*, Raffaello Cortina Editore, Milano, 2002.

³⁴ A. TURING, *Computing Machinery and Intelligence*, *Mind*, vol. 59, n. 236, pp. 433-460, 1950, p. 439.

interrogatori nel gioco dell'imitazione. L'autore stesso rileva il carattere congetturale ed euristico della sua convinzione, tanto da pronunciarsi poco più avanti in una breve prolusione di natura epistemologica:

«The popular view that scientists proceed inexorably from well-established fact to well-established fact, never being influenced by any improved conjecture, is quite mistaken. [...] Conjectures are of great importance since they suggest useful lines of research»³⁵

Serve ora evidenziare un problema fondamentale, sul quale ritorneremo più volte. Gli intenti della presente ricerca impongono infatti – volendo tracciare una genealogia dei modelli di identificazione mente-computer come appaiono nelle filosofie computazionaliste – uno sforzo verso una sorta di filologia del pensiero turinghiano, poiché facciamo risalire ad esso lo sviluppo dei campi disciplinari di cui abbiamo parlato più sopra.

Questo sforzo ci deve far rilevare che la domanda di Turing sull'intelligenza delle macchine trova una risposta che elude consapevolmente il problema profondo dell'intelligenza come comparirebbe in filosofia e in psicologia. Si afferma che:

[...] the question, "Can machines think?" should be replaced by "Are there imaginable digital computers which would do well in the imitation game?"³⁶

E aggiunge poco più avanti, nella sezione successiva:

«The original question, "Can machines think?" I believe to be too meaningless to deserve discussion»³⁷

Questi pochi passi sono indizi ulteriori a sostegno di quanto già detto: segnalano che la posizione turinghiana in merito di intelligenza delle macchine non riguarda un'analisi effettiva del problema dell'intelligenza in quanto tale.

³⁵ *Ivi*, p. 442.

³⁶ *Ibid.*

³⁷ *Ibid.*

Turing procede così all'esposizione delle posizioni contrarie, che possiamo esaminare rapidamente.

Il matematico comincia dall'obiezione teologica, la quale postula che il pensiero sia una funzione dell'anima umana, direttamente concessa da Dio a tutti gli uomini e tutte le donne ma non agli animali e non – certamente – alle macchine. Chiunque si preoccupi della scientificità delle proprie affermazioni non potrebbe accogliere l'argomento neppure in minima parte; in effetti, Turing lo rifiuta completamente, pur tentando una breve replica in termini teologici prima di passare all'argomento successivo, che chiama della “testa sotto la sabbia”³⁸.

Chi sostiene questa posizione teme la costruzione di macchine pensanti, per questo spera (e infine si convince) che esse non possano pensare. L'argomento in questione intrattiene un legame esplicito con quello teologico, poiché, ugualmente, postula una indimostrata superiorità dell'intelletto umano. Anche la seconda obiezione, come la prima, non è neppure sufficientemente solida da richiedere una confutazione³⁹.

La terza è un'obiezione di carattere matematico: certi risultati conseguiti in logica matematica, segnatamente i teoremi gödeliani degli anni Trenta del Novecento, mostrano che in ogni sistema formale potente (lo abbiamo visto, più sopra, per quello dei *Principia Mathematica* di Whitehead e Russell) esistono proposizioni indecidibili, che non possono cioè essere dimostrate o confutate, a meno che il sistema sia incoerente. Le macchine digitali sono costruite come sistemi formali, dunque il problema sollevato dagli studi di Gödel le riguarda direttamente. I limiti dei sistemi computazionali automatici sono mostrati anche Turing, i cui risultati sono compresi nell'obiezione:

«The result in question refers to a type of machine which is essentially a digital computer with an infinite capacity. It states that there are certain things that such a machine cannot do»⁴⁰

³⁸ Cfr. *Ivi*, p. 444, «The ‘Heads in the Sand’ Objection».

³⁹ Cfr. *Ibid.* L'autore aggiunge, ironicamente: «Consolation would be more appropriate».

⁴⁰ *Ibid.*

In altri termini, esistono domande alle quali una simile macchina, preparata ai compiti necessari allo svolgimento del Test di Turing, non risponderà correttamente oppure non risponderà affatto, qualunque sia il tempo ad essa concesso per elaborare una risposta.

In questo caso è formulata una contro-obiezione: per quanto esistano dei limiti a quello che le macchine possono fare, non si potrebbe forse ritenere che queste e più pesanti limitazioni gravino anche sull'intelletto umano? Senza alcuna dimostrazione, infatti, noi asseriamo che tali difetti riguardano solo le macchine e non anche i processi che caratterizzano il pensiero umano.

Si giunge dunque al quarto argomento, che Turing chiama della coscienza. Secondo tale obiezione non sarebbe possibile parlare di intelligenza senza tenere in debita considerazione il problema della coscienza. L'argomento è riportato nella forma in cui lo esprime, nel 1949, il neurologo britannico Geoffrey Jefferson:

«Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain—that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants»⁴¹

Non basta, cioè, che una macchina si limiti ad emulare l'intelligenza umana – per esempio componendo un concerto o un sonetto – per ritenere valida la relazione di uguaglianza tra mente e macchina medesima; è necessario che essa abbia coscienza dei risultati che consegue e delle attività che svolge.

La risposta di Turing propone una sorta di *reductio ad absurdum* dell'argomento: esso, condotto alle sue conseguenze più estreme, implicherebbe che l'unico modo per accertare l'intelligenza di una macchina sia di *essere* la macchina e di *sentirsi pensare*. Bisognerebbe poi, allo stesso modo, applicare il

⁴¹ *Ivi*, p. 446.

ragionamento al pensiero umano: l'unico modo per accertarsi che qualcuno pensi veramente sarà di essere quell'individuo determinato.

L'argomento della coscienza approda ad un solipsismo radicale che – dice Turing – con tutta certezza anche il professor Jefferson rigetterebbe.

Prima di passare all'analisi delle obiezioni successive, Turing ci lascia un'indicazione utile a comprendere la sua posizione con maggiore accuratezza: egli afferma di non ritenere affatto che ogni problema sullo statuto della coscienza sia risolto. Le sue convinzioni sull'intelligenza delle macchine, in altri termini, non dipenderebbero da una sottovalutazione della complessità della coscienza umana.

Cionondimeno, riferisce di non credere che per rispondere alla domanda di partenza sia necessario prima svelarne ogni mistero. Potremo ritornare sulla questione a breve.

La classe di argomenti di cui Turing si occupa successivamente è una sorta di estensione dell'obiezione precedente. Questi argomenti postulano l'esistenza di un insieme di compiti che nessuna macchina potrebbe svolgere, per esempio: essere gentile o amichevole, prendere iniziativa, avere senso dell'umorismo, commettere errori, innamorarsi, imparare dall'esperienza, essere oggetto del proprio pensiero, avere la stessa varietà comportamentale degli umani, fare qualcosa di veramente nuovo e originale⁴² (l'ultima caratteristica, quella dell'originalità, è lo scheletro dell'obiezione che – più avanti – Turing fa risalire ad Ada Lovelace, la matematica britannica nota per il suo contributo allo sviluppo della macchina analitica di Charles Babbage).

Data la parentela con l'argomento precedente, anche qualora si potessero programmare macchine che svolgono qualcuno di questi compiti (e oggi, a distanza di settant'anni dall'articolo turinghiano, i progressi in ambito IA lo consentono in larga misura), l'obiettore continuerebbe a non ritenerli utili a provare l'esistenza di una forma di reale intelligenza, asserendo che la macchina si limita – ancora una volta – ad imitarla mediante l'esecuzione di procedure di calcolo simbolico.

⁴² Cfr. *Ivi*, p. 447.

Quanto al problema dell'originalità del comportamento delle macchine, l'idea secondo cui esse, al contrario della mente umana, non possano fare niente di nuovo, ma solo ciò per cui sono state strettamente programmate, sarebbe ancora il risultato di una fallacia in cui tipicamente si imbattono matematici e filosofi:

«The view that machines cannot give rise to surprises is due, I believe, to a fallacy which philosophers and mathematicians are particularly subject. This is the assumption that as soon as a fact is presented to a mind all consequences of that fact spring into the mind simultaneously with it»⁴³

Di nuovo, i limiti che predichiamo della macchina non sono altrettanto accuratamente esaminati nel caso del pensiero umano. Se la macchina si limita ad elaborare un problema deducendo le proprie risposte secondo un insieme di regole e principi ben definiti, cosa ci assicura che qualcosa di simile non valga anche per noi? Non si potrebbe ritenere che ciò che chiamiamo, nell'ambito dell'intelligenza umana, "creatività", non sia altro che il risultato dell'elaborazione di tutte le conseguenze dipendenti da ciò che abbiamo imparato, dai dati a cui siamo stati sottoposti e dai principi generali che informano il nostro ragionamento?

Oppure si potrebbe presentare il problema dal versante opposto. Possiamo considerare la maggior parte delle menti umane – dice Turing – come "subcritiche" (*subcritical*): presentando ad una mente siffatta una qualsiasi idea, essa ne produrrà – al massimo – un'altra sola in risposta. Una strettissima minoranza di menti, invece, è di tipo "supercritico": un'idea presentata ad una mente simile genera un'intera teoria, fatta di idee secondarie, terziarie, remote, eccetera. Possiamo a questo punto chiederci se non sia possibile programmare macchine supercritiche.⁴⁴

La lista di obiezioni continua – nell'articolo del matematico inglese – ancora per diverse pagine.

⁴³ *Ivi*, p. 451.

⁴⁴ Cfr. *Ivi*, 454.

Turing ritorna, per esempio, sul problema della discretezza degli stati della macchina. Diversamente da quello che abbiamo detto più sopra, quando abbiamo ipotizzato un confronto con le teorie bergsoniane sulla continuità degli stati di coscienza, egli presenta il problema riferendolo alle caratteristiche del sistema nervoso e non alla coscienza in sé. Dunque, pur permanendo - rispetto al tema dell'intelligenza - entro uno schema implicitamente fisicalista (essa sarebbe, in questo caso, il diretto risultato della configurazione biologica del cervello umano, il bilancio tra *input* e *output* elettrici neuronali) Turing riconosce il peso del problema della continuità. Egli afferma:

«The nervous system is certainly not a discrete-state machine»⁴⁵

Perciò, a rigore, neppure la sola metafora che identifica pensiero e calcolo meccanico sarebbe accurata. Non si potrebbe, cioè, imitare il comportamento del sistema nervoso attraverso l'uso di macchine a stati discreti.

Turing risponde, pur riconoscendo che una macchina a stati discreti debba necessariamente essere diversa da una macchina "continua", che una volta riportato il problema al gioco dell'imitazione, la costituzione (discreta o continua) della macchina non produce più alcuna differenza.

Le ultime pagine dell'articolo, quelle che danno corpo a §7 (*Learning Machines*), contengono numerose altre indicazioni che ci autorizzano ad attribuire ad Alan Turing il grosso della paternità sulle più tarde discussioni in merito all'analogia mente-macchina.

Per esempio:

«The "skin-of-an-onion" analogy is also helpful. In considering the functions of the mind or the brain we find certain operations which we can explain in purely mechanical terms. This we say does not correspond to the real mind: it is a sort of skin which we must strip off if we are to find the real mind. But then in what remains we find a further skin to be stripped off, and so on. Proceeding in this way do we ever come to the "real" mind,

⁴⁵ *Ibid.*

or do we eventually come to the skin which has nothing in it? In the latter case the whole mind is mechanical»⁴⁶

Si ripropone qui una lettura meccanicistica, quantomeno parziale, dei fenomeni mentali. Per Turing il problema dell'imitazione meccanica della mente non è teorico o “di principio”. È, piuttosto, un mero problema di capacità di programmazione e di avanzamento tecnico (più il primo che il secondo).

In altri termini, a condizione di trovare gli strumenti per programmarle adeguatamente, le macchine potrebbero ottimamente riprodurre i processi del pensiero umano.

1.3.3 Filosofia ed euristica della ricerca turinghiana

È implicito, in una tale convinzione, un elemento facilmente individuabile: la posizione turinghiana esprime, forse in modo solo parzialmente consapevole, l'idea che il pensiero *tout court* sia meccanizzabile. Potremmo forse ritenere che Turing non avrebbe accolto la nostra chiosa, ma pare manifesto che le conseguenze dei ragionamenti che egli conduce – e che abbiamo cercato di seguire – si diano in questi termini, perlomeno se valutate in modo complessivo e sino ai loro esiti più remoti.

Quanto appena detto costituisce uno snodo di rilievo prioritario nell'economia del nostro discorso, indirizzandoci convintamente verso un primo abbozzo di risposta alla questione formulata in partenza.

L'analisi del lavoro che Turing conduce prima nell'articolo del 1936 e poi nei suoi lavori più tardi (i cui esiti sono stati sintetizzati, per nostra comodità, guardando all'opera del 1950) ci dice – pur secondo modalità tra loro diverse – che nel suo pensiero è possibile cercare le tracce del successivo sviluppo dei modelli computazionali della mente (esplicitamente formulati, si intende).

Allo stesso tempo, però, è necessario prendere atto della complessità teorica e disciplinare che produce, come risultato, le molte forme di tali modelli. Sarebbe pertanto grande il fraintendimento nel tentativo di attribuire interamente al

⁴⁶ *Ivi*, pp. 444-445.

matematico di Cambridge la paternità su programmi di ricerca lontani dalle sue intenzioni manifeste.

Nell'articolo del 1936 Turing analizza il processo di calcolo e lo riduce ad un insieme ristretto di azioni elementari. Dopodiché, compie un salto che quasi passa sottotraccia: l'estensione (possiamo supporre, irriflessa) di questa analisi dal calcolo al pensiero *tout court*. E dunque non solo il processo di calcolo sarebbe meccanizzabile, ma l'intera intelligenza, siccome essa consiste nello svolgimento sempre di un'azione - o poche azioni finite - alla volta.

Eppure, *On Computable Numbers* e – in modo particolare – la dissertazione di dottorato del 1938⁴⁷ contengono forti elementi a tutela di una posizione più cauta sulle possibilità di una analogia mente-macchina, in primo luogo una considerazione positiva del concetto di *intuizione* non computabile. Essa interverrebbe ogni qual volta – per esempio – un matematico ha cognizione della verità di una qualsiasi proposizione formalmente indecidibile, ossia di verità che non possono stabilirsi mediante esecuzione di un insieme di procedure meccaniche.

Con l'evoluzione del suo lavoro il concetto, che pure sembrava avere un peso cruciale nella sua considerazione dei processi mentali, pare invece scomparire. A far data dalla pubblicazione di *Proposal for the development in the Mathematical Division of an Automatic computing engine (Ace)*⁴⁸, Turing si mostra persuaso che le sole operazioni computabili siano sufficienti a rappresentare ogni funzione e processo eseguiti dal cervello.

Una tale anomalia potrebbe forse essere risolta differenziando gli elementi di stretta derivazione tecnica da quelli che si collocano nel più generale panorama dell'euristica turinghiana, i quali non rispondono tanto ai “fatti” osservabili e accertati (nella fattispecie, le dimostrazioni della logica matematica gödeliana) quanto ai meno nitidi convincimenti del matematico. In altri termini, l'evidenza – prima riconosciuta e poi rigettata – di processi mentali intuitivi che

⁴⁷ A. TURING., *Systems of logic based on ordinals*, in Proceedings of the London Mathematical Society, serie seconda, 45, 1938, pp. 161-228.

⁴⁸ A. TURING., *Proposal for the development in the Mathematical Division of an Automatic computing engine (Ace)*, rapporto all'Executive Committee del National Physical Laboratory del 1945, in B. E. CARPENTER e R. W. DORAN, *A. M. Turing's Ace report of 1946 and other papers*, MIT Press, Cambridge Mass, 1986, pp. 20-105, 1945.

sfuggono alla meccanizzazione apparterebbe ad un orizzonte squisitamente tecnico-matematico; al contrario, l'idea secondo cui l'ambito della calcolabilità e quello del pensiero coincidono perfettamente afferirebbe alla sfera di un paradigma radicato che guida la costruzione di teorie e metodi ed è carico di valori più e meno espliciti.

Si spiegherebbe così la trasposizione di questa immagine dei processi mentali nella controparte "macchinica" e l'elaborazione - sin dagli inizi della carriera - di uno schema del processo di calcolo determinato da "stati mentali" finiti, i quali, sempre dentro i confini di un modello a forte tenore riduzionistico, sarebbero in grado di dare luogo ad esiti potenzialmente infiniti. Il risultato emerge con forza, come abbiamo visto, nell'articolo del 1950: tutte le operazioni mentali sono computabili e l'organizzazione cerebrale da cui sorge l'intelligenza deve essere riproducibile come macchina a stati discreti finiti.

1.3.4 Quale riduzionismo?

La filosofia turinghiana si gioca dinamicamente – così ci pare di vedere – su una gamma variegata di posizioni teoriche, e manifesta mutamenti costanti nell'arco della carriera del matematico. Volendo riportare il disordine di queste posizioni ad uno schema generale, potremmo distinguerne almeno due versioni:

I. Un riduzionismo operativo, o debole

Esso riguarda la prima fase degli studi di Turing, nella quale è certamente incluso l'articolo del 1936, e si dà spesso in forma esplicita. È operativo perché riguarda innanzitutto la possibilità di formalizzazione del processo di calcolo; è un riduzionismo di comodo, procedurale, pertanto è anche debole (non intende cioè primariamente affermare l'identità tra pensiero e calcolo simbolico). In esso, la discretezza degli stati della macchina computazionale non si generalizza alla mente, pur derivando l'analisi del calcolo da quella del pensiero.

II. Un riduzionismo generale, o forte.

Consistente in una effettiva identificazione del pensiero con il calcolo, nella astrazione generale compiuta sulla mente, intesa ora come principio funzionale organizzato secondo stati finiti e processi atomici, costituito da elementi singolari (stato mentale/simbolo processato) stanti tra loro in relazione uno a uno. Qui la discretezza è invece predicata del pensiero in senso ontologicamente forte. Questa pare la filosofia sottesa all'articolo del 1950, sebbene – come vedremo tra poco – tale apparenza si presti a fraintendimenti di vario genere.

È necessario, forse, risolvere il problema della coesistenza di questi approcci: quale gerarchia intendere istituita?

In primo luogo bisogna rilevare che il carattere del lavoro di Turing è innanzitutto tecnico, mirante alla costruzione formale di meccanismi di calcolo. Pertanto, potendo essere chiamato a rispondere delle sue posizioni, il matematico difenderebbe senz'altro la versione operativa del suo riduzionismo.

Eppure, ogni studio accurato della letteratura turinghiana rende manifesta l'oscillazione costante tra gli estremi che abbiamo presentato, tra un riduzionismo debole e uno forte.

Tale fluttuazione è difatti riflessa nelle molte posizioni, spesso contrastanti, espresse a vario titolo dagli interpreti del pensiero turinghiano: Frixione e Numerico⁴⁹, ad esempio, rendono conto della grande difficoltà di inquadramento univoco del problema; anche qui si oscilla tra interpretazioni a sostegno di un riduzionismo del secondo genere (e.g. quelle di Hodges⁵⁰ e di French⁵¹) e interpretazioni contrarie, più facilmente riconducibili al quadro di un riduzionismo moderato o, come già è stato definito, operativo e debole.

È il caso degli argomenti di Copeland⁵² e Teuscher⁵³ che, al contrario dei primi, negano l'appartenenza di Turing al consesso degli studiosi che hanno

⁴⁹ M. FRIXIONE, T. NUMERICO, *Alan Mathison Turing*, in APhEx, n. 7, gennaio 2013.

⁵⁰ A. HODGES, *Alan Turing. The enigma of intelligence*, II ed. Vintage, London, 1992 (trad. it. *Storia di un enigma*, Bollati Boringhieri, Torino, 2003).

⁵¹ R. M. FRENCH, *The Turing test: The first 50 years*, in Trends in Cognitive Sciences, 4 (3), 115-122, 2000.

⁵² J. B. COPELAND e D. PROUDFOOT, *On Alan Turing's anticipation of connectionism*, Syntese, n. 108, pp. 361-377, 1996.

⁵³ C. TEUSCHER, *Turing's connectionism: an investigation of neural network architectures*, Springer-Verlag, London, 2002; C. TEUSCHER (a cura di), *Alan Turing: life and legacy of a great thinker*, Springer, Berlin, 2004; C. TEUSCHER, *Turing's Connectionism*, in Teuscher, 2004, pp. 499-529.

sostenuto «che l'intelligenza potesse essere prodotta attraverso una semplice messa a punto della struttura computazionale di manipolazione di simboli secondo regole precise e immutabili»⁵⁴.

I due autori dello studio propendono per questa spiegazione, visto che al centro del lavoro di Turing sarebbe la nozione, filosoficamente rilevante, di *interferenza*:

«Turing sosteneva per la prima volta esplicitamente una posizione che pochi storici del pensiero scientifico sono stati disposti a riconoscergli: da questo momento in poi si interessò solo a macchine diverse da quelle tradizionali, che erano prive di contatti con l'esterno. I dispositivi ai quali voleva dedicare la sua attenzione erano quelli per cui l'interferenza non fosse l'eccezione, [...] ma la regola.»⁵⁵

L'interferenza non sarebbe altro, dunque, che il contatto con l'esterno, opportunamente considerato da Turing essere alla base di ogni definizione non ingenua dell'intelligenza umana.

Se la simulazione dell'intelligenza attraverso la macchina è il nostro obiettivo, non potremo che considerare ogni uomo come una macchina costantemente soggetta a interferenze esterne, sempre in comunicazione, sempre sottoposta a stimoli di varia natura e condizionamenti ambientali che finiranno per determinarne il comportamento in modi difficilmente prevedibili.

La nozione di interferenza aggiunge un grado di complessità all'analogia tra computazione e pensiero, negando implicitamente la tesi secondo cui la sola manipolazione di simboli mediante regole sia sufficiente a produrre qualcosa come un'intelligenza. L'interferenza pone tra le regole di calcolo e il risultato operativo (della macchina o del pensiero umano) un residuo incalcolabile, non organizzato.

Ne deriva un fatto fondamentale e ulteriore: così concepita, la nozione di intelligenza risulta come un concetto sociale, una procedura di riconoscimento

⁵⁴ M. FRIXIONE, T. NUMERICO, op. cit., p. 541.

⁵⁵ *Ivi*, pp. 538-539.

varata collettivamente e determinata da un sistema di regole socialmente condiviso.

Incrociando le informazioni appena acquisite con quelle già trattate al momento dell'esposizione del Test di Turing – esplicitamente fondato sulla *pretence*, sulla finzione messa in atto dalla macchina – risulta infine più conveniente adottare le posizioni di Copeland e Teuscher difese da Frixione e Numerico:

«L'intento di Turing era quello di proporre un metodo per riconoscere socialmente dei comportamenti intelligenti paragonabili a quelli umani e prodotti da dispositivi meccanici. Egli riteneva lecita anche la domanda “Non possono forse le macchine comportarsi in qualche maniera che dovrebbe essere descritta come pensiero ma che è molto differente da quanto fa un uomo?”»⁵⁶.

Nell'alternanza tra le diverse versioni di riduzionismo che accompagnano il lavoro di Alan Turing, dunque, è facile rintracciare anomalie concettuali e paradigmi contrastanti. Questa è senz'altro la ragione primaria delle molte e diverse posizioni assunte da studiosi, egualmente autorevoli, del pensiero del matematico di Cambridge.

Tuttavia, in base a quanto è stato stabilito, affermiamo che la versione operativa di riduzionismo, enunciata poco sopra, sia quella che meglio si presta a definire la filosofia turinghiana o anche, come l'abbiamo chiamata, la sua euristica.

L'insieme, cioè, tanto delle sue posizioni concettualizzate e riflesse in materia di intelligenza e possibilità di formalizzazione (la sua filosofia), quanto dei metodi di ricerca, dei valori e dei convincimenti che guidano il suo lavoro (la sua euristica) si possono leggere alla luce di un riduzionismo del primo tipo, che è messo in campo con finalità *operative* di astrazione.

Esso risulta necessario alla *simulazione* e *riproduzione* dei meccanismi dell'intelligenza mediante strumenti meccanici adeguatamente programmati, ma non stabilisce l'effettiva sovrapposibilità di pensiero e computazione meccanica.

⁵⁶ *Ivi*, pp. 544-545.

Da ciò risulta che i tentativi di identificazione di pensiero e calcolo, intelligenza e macchina computazionale – i cui modelli analizzeremo nel capitolo secondo – rivendicano una derivazione strettamente turinghiana solo sulla scorta di un fraintendimento, collocando le molte aperture teoriche garantite dal lavoro del padre dell'informatica al di fuori del loro originario campo di concepimento e applicazione.

II.

STRUMENTI DELL'ANALOGIA: IL PARADIGMA RIDUZIONISTICO E I SISTEMI FORMALI

È stato tracciato un abbozzo dei processi concettuali che hanno portato allo sviluppo di una teoria generale della calcolabilità a partire dalla sola analisi di questioni logico-matematiche fondazionali.

È andata così delineandosi una storia di progressive trasformazioni teoriche e di altrettanto sequenziali sconfinamenti di campo, di per sé epistemologicamente rilevante, in quanto evidenza del carattere serendipico della ricerca e del progresso delle scienze in generale.

Si può constatare facilmente la difformità di intenti dei protagonisti degli sviluppi che abbiamo tracciato: Hilbert e colleghi perseguono, come dimostrano anche i celebri *Grundlagen der Geometrie*⁵⁷, l'obiettivo di un aumentato rigore della matematica, da realizzarsi mediante assiomatizzazione, oppure entro il più generale ambito della riflessione logica e metamatematica. Il Problema della Decisione è un caso particolare del lavoro di ricerca matematica concepito in questi termini.

Tuttavia – e qui si realizza la prima trasformazione – la matematica novecentesca compie, con l'*Entscheidungsproblem*, un salto decisivo, finendo per atterrare fuori dal terreno del dibattito fondazionale.

Raccolta dal giovane Turing, la sfida di Hilbert produce un effetto del tutto estraneo alle intenzioni del matematico di Königsberg, innanzitutto perché provata insolubile, poi dando origine ad un intero campo disciplinare, tutto inesplorato, estremamente fecondo di conseguenze epocali sul versante teorico e su quello pratico a un tempo.

⁵⁷ D. HILBERT, *Grundlagen der Geometrie*, Teubner, Leipzig, 1899 (tr. it.: *Fondamenti della geometria. Con i supplementi di Paul Bernays*, Franco Angeli Editore, Milano, 2009). L'opera è una vera e propria riscrittura assiomatica della geometria euclidea. In quanto tale, si colloca a pieno titolo nell'epocale dibattito sui fondamenti. Entro questo ambito si registrano numerosi altri tentativi affini; uno tra tutti è quello del matematico italiano Giuseppe Peano, il quale assiomatizza l'insieme dei numeri naturali nel 1889.

Coincidendo questi problemi con l'atto di nascita dell'informatica, l'impatto pratico della trasformazione in questione è veramente definibile come epocale. Soggiace questo grande impatto pratico, tuttavia, un insieme di rivolte teoriche di enorme spessore, ovviamente prioritarie nella nostra ricerca. La nascita della teoria della computabilità, già discussa in precedenza, ha conferito nuova energia, nuovi strumenti concettuali e una vasta e aggiornata risonanza alle opzioni meccanicistiche lungamente espresse nella storia delle idee, che hanno difatti ripreso a fiorire in maniera decisiva.

Quanto al carico euristico del lavoro di Turing, l'insieme delle ipotesi di lavoro, dei valori, delle intuizioni che lo guidano, lo abbiamo visto costantemente accompagnato da una filosofia di matrice riduzionistica, talvolta nascosta e implicita, talvolta esplicitata in modi differenti e sfumati. L'oggetto di tale riduzione è, informalmente inteso, il pensiero; l'esito della riduzione è la sua assimilazione al calcolo, alle operazioni formali di manipolazione simbolica. Si è accertato, tuttavia, che tra le molte forme del riduzionismo turinghiano, solo quella che abbiamo deciso di chiamare "operativa" o "debole" può adeguatamente essere riferita agli intenti del matematico di Cambridge. Non paiono invece conformi alla sua filosofia le versioni forti del riduzionismo in questione, che postulano la piena sovrapposibilità di pensiero e calcolo nella forma privilegiata del rapporto *input/output*. Di queste sfumature si è cercato di rendere conto nel capitolo da poco concluso.

Dovremo, giunti a questo punto dell'analisi, proseguire nella direzione dei modelli computazionali forti. Essi traggono dagli studi computazionali un impulso evidente, il quale può essere tuttavia interpretato – come abbiamo fatto – nei termini di un sommario fraintendimento degli esiti del pensiero di Alan Turing.

Gli studi computazionali, però, costituiscono solo una parte, sebbene centrale, del processo di introduzione, entro il campo largo delle scienze cognitive, degli approcci riduzionistici e formalizzanti che ne guidano la teoria. A testimoniarlo sta un vasto insieme di esperienze scientifiche di rilevanza essenziale, le quali presenteremo brevemente, e le cui conseguenze per la formazione dell'analogia mente-macchina analizzeremo nelle pagine seguenti.

Entro la storia dei mutamenti teorici che stiamo provando a tracciare è certamente possibile collocare in questo passaggio il secondo sconfinamento di campo, la seconda grande contaminazione concettuale, la prima consistendo nel passaggio dai temi matematico-fondazionali a quelli dell'informatica delle origini.

2.1 Nicolas Rashevsky: la biologia come sistema formale

2.1.1 Scienze teoretiche, scienze sperimentali

Nel pieno solco dell'analogia già prospettata si colloca il lavoro di Nicolas Rashevsky, eminente fisico teorico russo attivo, a partire dagli anni Trenta del Novecento, per lo più presso l'Università di Chicago.

Protagonista di una straordinaria migrazione disciplinare, che può leggersi alla luce di un generale interesse, estremamente vivace a inizio XX secolo, per l'importazione di strumenti e metodi della fisica entro il campo delle scienze biologiche, Rashevsky è generalmente riconosciuto quale fondatore della cosiddetta "biofisica matematica".

Spesso riguardante l'introduzione di strumenti per l'analisi quantitativa dei fenomeni biologici, quella di Rashevsky è invece una migrazione di diverso carattere, motivata dal tentativo di trattare la disciplina in maniera interamente teoretica, con aderenza al lavoro sperimentale pressoché nulla.

L'interesse a descrivere la biologia – ambito disciplinare a trazione osservativa e sperimentale – come sistema squisitamente formale è, in effetti, ragione primaria della diffidenza dei biologi "classici" nei confronti del lavoro del fisico russo. Della sua ricezione sarà importante discutere più avanti.

Addottoratosi in fisica teorica nel 1919 presso l'Università di Kiev, Rashevsky lavorò alle facoltà di fisica di Costantinopoli e poi di Praga, pubblicando i suoi primi articoli sulla *Zeitschrift für Physik*.

All'avvio di carriera operò su una varietà di tematiche d'avanguardia della fisica a lui contemporanea: si occupò di teoria della relatività, di meccanica delle matrici, di fotomagnetismo, di elettrodinamica⁵⁸. Nell'elaborazione di tali

⁵⁸ T. H. ABRAHAM, *Nicolas Rashevsky's Mathematical Biophysics*, Journal of the History of Biology, 37, pp. 333-385, p. 337, Kluwer Academic Publishers, 2004.

problemi si scorge già la tendenza allo sviluppo di trattazioni meramente matematico-formali e la scarsa aderenza alle istanze sperimentali che avrebbero caratterizzato, più tardi, il suo lavoro in biologia:

«[...] his conclusions were often related to the logical consistency of the problem expressed in mathematical terms, and referred to variables solely as mathematical variables, devoid of their meaning within a physical system.»⁵⁹

All'epoca del suo impegno di ricerca presso l'Università di Pittsburgh, lavorando allo studio della dinamica delle soluzioni colloidali, Rashevsky formulò una analogia tra i processi di divisione delle particelle osservabili nei colloidali e quelli della divisione cellulare negli organismi viventi, procedendo così al tentativo di descrivere i meccanismi di separazione delle cellule biologiche attraverso lo sviluppo di sistemi di equazioni differenziali, già utilizzate per la descrizione delle interazioni tra particelle delle soluzioni colloidali.

È essenziale notare che alla radice della descrizione formale dei processi di divisione cellulare stanno un insieme di ipotesi basilari, di semplificazioni e di approssimazioni, senza le quali la descrizione risulterebbe impossibile. L'assunzione di ipotesi di partenza e la formulazione di semplificazioni e approssimazioni riguarda la struttura stessa della nozione di analogia, sulla cui importanza e natura ritorneremo in modo sistematico. Ora possiamo, invece, osservare sinteticamente qualcuno dei trattamenti formali riservati da Rashevsky ad alcuni processi biologici, partendo precisamente da quello riguardante la divisione cellulare. Un loro svelto esame illuminerà le prassi di astrazione, i limiti e le difficoltà che accompagnano ogni condotta formalistica e riduzionistica, consentendoci poi di costruire, su tali evidenze, gli argomenti successivi della presente ricerca.

⁵⁹ *Ibid.*

2.1.2 Sul trattamento formale dei fenomeni biologici

Sulla scorta della già presentata analogia con le soluzioni colloidali, egli avvia la costruzione di uno schema ipotetico utile alla pre-formalizzazione del fenomeno della divisione spontanea di una cellula:

«He conceived of hypothetical aggregates of living cells in a surrounding medium, whose division would depend on the same thermodynamic principles as those that governed spontaneous division in colloidal particles.»⁶⁰

Dunque procede a rappresentarsi la cellula biologica e il suo ambiente circostante in termini che possiamo così sintetizzare: una goccia di liquido A sospeso nel liquido B, a sua volta contenente un assortimento di “sostanze” C, D, E, F, ..., variamente interagenti in modo da dare origine al liquido A. Nel sistema appena prospettato, l'interazione di C, D, E, F generante A causa necessariamente una crescita nella dimensione di A stesso: ad esempio, l'elemento D è convertito, al termine di una reazione, in una nuova goccia di A la quale, non essendo solubile in B, potrebbe unirsi alle quantità di A già esistenti. L'accrescimento del liquido A, Rashevsky trova, dipenderebbe in tal caso: 1) dall'area superficiale della goccia; 2) dalla concentrazione di D nella goccia, secondo la seguente espressione (ricalcata sulla formula classica per il calcolo dell'area superficiale di una sfera):

$$4\pi r_0^2 h(n_0 - n),$$

in cui r_0 è il raggio della goccia, n_0 la concentrazione di D in B, n la concentrazione di D in A e h una costante. Allo stesso modo, la quantità di D trasformata ogni secondo in A dipenderebbe: 1) dal volume della goccia; 2) di nuovo dalla concentrazione di D nella goccia. In questo caso l'espressione è derivata dalla formula per il calcolo del volume di una sfera:

$$\frac{4}{3}\pi r_0^3 qn,$$

⁶⁰ *Ivi*, p. 339.

con q costante e n ancora indicante la concentrazione di D in A.

Rashevsky continua con la formulazione di equazioni differenziali esprimenti la variazione di massa della goccia e , ponendo tale valore in relazione con quello della tensione superficiale della sua cellula idealizzata, determina la soglia di accrescimento alla quale la goccia finirebbe spontaneamente per dividersi⁶¹.

La grande maggioranza degli studi dedicati al lavoro di Rashevsky rileva la consuetudine di approcci simili nella fisiologia degli anni Venti e Trenta del Novecento: il lavoro del fisico russo non sarebbe, pertanto, completamente estraneo alla cornice concettuale entro cui molti scienziati a lui contemporanei usavano aggredire il problema della divisione cellulare.

I criteri d'indagine allora in voga prevedevano un ampio ricorso a metodi fisici e biochimici e al trattamento della cellula in termini di permeabilità delle membrane, di reazione agli stimoli chimici esterni, di interfaccia tra superfici e di tensioni superficiali⁶².

Tuttavia, a differenza dei suoi colleghi, attenti alla verifica sperimentale delle loro descrizioni formalizzanti, gli studi di Rashevsky mancano di riferimento a casi sperimentali specifici, occupandosi solo di sistemi cellulari idealizzati.

Il suo utilizzo dei formalismi matematici non mira all'analisi quantitativa dei processi biologici: egli non applica le sue espressioni a insiemi preesistenti di dati ottenuti sperimentalmente. Al contrario, la matematica è il nucleo pulsante della sua metodologia, che appare formalistica e deduttiva sin dal principio, essendo impiegata per l'idealizzazione e la ri-concettualizzazione dei processi biologici che analizza⁶³. L'eco di tali posizioni risuona chiaramente in un articolo pubblicato, nel 1935, sulla rivista *Nature*, la cui prima sezione è un eloquente intermezzo metodologico:

⁶¹ Cfr. N. RASHEVSKY, *The mechanism of cell division*, in *Bulletin of Mathematical Biophysics*, vol. 1, p. 23-30, 1939.

⁶² Cfr. T. H. ABRAHAM, *op. cit.*, e D. S. HOFFMANN, *The Dawn of Mathematical Biology*, *Controversia*, vol. 9, n. 2, pp. 53-61, 2013.

⁶³ Emblematicamente, Rashevsky pone in esergo all'articolo *Foundations of Mathematical Biophysics*, comparso in *Philosophy of Science*, Vol. 1, n. 2, l'esortazione platonica “Μηδεις αγεωμετρητος εισιτω”, una citazione di Roger Bacon (“He who knows not mathematics cannot know any other sciences; what is more, he cannot discover his own ignorance or find its proper remedies”) e una tratta dai *Principi metafisici della scienza della natura* di Immanuel Kant (“Ritengo, tuttavia, che in ogni particolare scienza della natura si trovi solo tanta scienza effettiva quanta matematica vi è in essa”).

«[...] there is a wealth of literature on the application of mathematical statistics to various biological phenomena [...] but most of this application is restricted on the use of physical apparatus in biological experiment; and very little attempt has been made to gain an insight into the physico-chemical basis of life, similar to the fundamental insight of the physicist into the intimate details of atomic phenomena»⁶⁴.

Analoga elaborazione formale, modellata precisamente sulla sua matematica della divisione cellulare, Rashevsky dedica ad un secondo, fondamentale problema fisiologico: quello della trasmissione degli impulsi nervosi.

In una serie di articoli pubblicati nei primi anni Trenta⁶⁵ sono esposti gli elementi di una teoria matematica della conduzione nervosa basata sulle nozioni di eccitazione, inibizione e gradiente elettrochimico.

Rashevsky analizza le teorie disponibili, le classifica in due gruppi differenti e ne riconosce le assunzioni fondamentali, che riguardano: il ruolo degli ioni in soluzione, la loro distribuzione, il loro movimento quando sottoposti ad una corrente elettrica. Tali teorie, inoltre, presumono che ai fini dell'eccitazione nervosa una soglia critica di concentrazione ionica sia necessaria.

Per parte sua, Rashevsky assume che eccitazione e inibizione delle cellule nervose siano regolate da due differenti fattori: un "fattore di eccitazione", la cui concentrazione chiama e , ed un "fattore di inibizione", la cui concentrazione decide di indicare con la variabile i . Il rapporto e/i sarà dunque responsabile dell'eccitazione: se e/i diventa maggiore di una certa costante (h), ovvero se la concentrazione del fattore di eccitazione è sufficientemente più alta di quella del fattore di inibizione, allora si dà impulso nervoso.

Ad h Rashevsky assegna valore 1 (lo stesso faranno per diretta influenza del maestro, come vedremo più avanti, McCulloch e Pitts nel loro *Logical Calculus*), mentre assume che i valori di concentrazione di e ed i a riposo (quando, cioè, il neurone è in stasi) siano: $e_0 < i_0$.

⁶⁴ N. RASHEVSKY, *Mathematical Biophysics*, Nature, vol. 135, pp. 538-530, 1935, p. 528.

⁶⁵ N. RASHEVSKY, *Learning as a Property of Physical Systems*, Journal of General Psychology, 5, 207-229, 1931; N. RASHEVSKY, *Outline of a Physico-mathematical Theory of Excitation and Inhibition*, Protoplasma, 20, 42-56, 1931; N. RASHEVSKY, *On the Theory of Nerve Conduction*, Journal of General Physiology 14, 517- 528, 1933; N. RASHEVSKY, *Some Physico-mathematical Aspects of Nerve Conduction*, Physics, 4, 341- 349, 1933.

Procede infine a tracciare due equazioni differenziali che descrivono, entro il contesto fortemente idealizzato delle sue premesse, le variazioni di concentrazione di i ed e rispetto alla corrente elettrica:

$$\frac{de}{dt} = KI - \kappa(e - e_0)$$

$$\frac{di}{dt} = MI - m(i - i_0)$$

Dove K , M , k , m sono costanti e I è la corrente elettrica. Per ultima Rashevsky scrive l'equazione che descrive la relazione tra l'intensità della corrente, la durata dello stimolo elettrico e le concentrazioni di e ed i :

$$K_t = \log \frac{KI}{KI - k(i_0 - e_0)}$$

Abbiamo riportato, in linguaggio matematico, alcune delle conclusioni raggiunte da Rashevsky, a esclusiva evidenza di un fatto per noi fondamentale: ogni teoria formale di un sistema complesso si regge, per sua stessa natura, sull'acquisizione di premesse di carattere semplificatorio o persino arbitrario, su procedure di approssimazione, di astrazione e di idealizzazione nonché sull'elaborazione *ad hoc* di enti e concetti utili.

Tale è, qualora gli esempi superiori fossero ritenuti insufficienti, pure il caso dei modelli formali introdotti nel capitolo XII di *Mathematical Biophysics* ove, ai fini dello svolgimento di un calcolo sulla variazione di energia, in un sistema cellulare idealizzato, al momento della trasformazione di una sfera di partenza (la cellula madre) nelle due sfere risultanti (cellule figlie), è necessario postulare un'immaginaria espansione e una relativa immediata contrazione della sfera:

«To calculate the energy change during division, [...] we shall subject the cell to an imaginary cycle, consisting of a gradual

expansion until it is completely dispersed, and a subsequent recondensation into two half-cells»⁶⁶,

Da cui l'introduzione della formula per il calcolo della pressione sulla membrana in relazione all'espansione del volume della cellula:

$$\int_{r_0}^{\infty} F_m dr_0 = \frac{RT\mathfrak{N}q_0B}{4\pi AMr_0}$$

E di quella per il calcolo della variazione di energia nel sistema dopo la condensazione in due cellule figlie:

$$\Delta W_m = -\left(\frac{RT\mathfrak{N}q_0B}{4\pi AMr_0} - \frac{RT\mathfrak{N}q_0B}{6.4\pi AMr_0}\right) = -\frac{3RT\mathfrak{N}q_0B}{32\pi AMr_0}$$

In questo caso persino il lettore matematicamente più inesperto potrebbe afferrare, anche per sola affinità visuale, la derivazione della formula finale per il calcolo della variazione energetica (ΔW_m) da quella superiore, che invece descrive la variazione delle forze sulla membrana (F_m) quando questa è sottoposta all'espansione volumetrica della sfera (indicata dall'integrale da r_0 a ∞).

Tale ulteriore, sintetica esibizione di procedure di calcolo e modellizzazione non costituisce un'incursione accessoria o inessenziale nei tecnicismi formali dell'opera di Rashevsky; al contrario, ci consente di ribadire quanto stabilito in precedenza sui meccanismi sottesi ai processi di formalizzazione – fatto che guida il nostro discorso sulle possibilità e i limiti del trattamento formale dei sistemi complessi, da Turing alle Scienze Cognitive: affinché possa darsi descrizione formale (ovvero finitistica, nel lessico di Hilbert, o effettiva, in quello turinghiano) di un qualsivoglia fenomeno complesso, devono realizzarsi procedure di astrazione, di semplificazione, di idealizzazione, di costruzione di enti e processi funzionali allo sviluppo o alla tenuta descrittiva della teoria.

⁶⁶ RASHEVSKY N., *Mathematical Biophysics. Physico-mathematical foundations of Biology*, vol. 1, 3a ed., Dover Publications, New York, 1960, p. 151.

Inoltre, siccome dei fenomeni complessi, in quanto tali, può darsi lettura multilivellare, le procedure formalizzanti descrivono generalmente, in modo più o meno efficace, soltanto i livelli inferiori del fenomeno medesimo.

Rashevsky stesso mostra una certa insoddisfazione verso il procedimento:

«The imaginary dispersion and recondensation of a cell, used in chapter xii for the calculation of energy-difference between a whole spherical cell and two spherical half-cells, is very unsatisfactory»⁶⁷,

E riconosce le equazioni summenzionate come approssimazioni:

«Rashevsky admitted that these equations were both “first approximations” since for high values of I in this case, e and i become negative. [...] Rashevsky was making approximations – deriving approximate formulae that hold, for example, for only “very small” values of t or m or k .»⁶⁸.

Identica ammissione, sebbene di carattere più generale, Rashevsky avanza nelle introduzioni ai già citati due volumi di *Mathematical Biophysics*:

«This book is naturally full of shortcomings, of which the author is perfectly aware. [...] We have in mind the necessary *abstraction* and *oversimplification* which is inherent to this approach. Objection has been raised against the use of the word “cell” to describe a highly oversimplified conceptual system [...]»⁶⁹.

Alla matematizzazione dei processi cellulari e dei fenomeni relativi alla trasmissione nervosa Rashevsky dedica un numero cospicuo di studi. Si occupa, ad esempio, di: teoria fisico-matematica della mitosi, teoria molecolare della divisione cellulare, forma e movimento delle cellule, crescita cellulare, meccanismi di autoregolazione biologica.

⁶⁷ *Ivi*, p. XII.

⁶⁸ T. H. ABRAHAM, *op. cit.*, p. 346.

⁶⁹ RASHEVSKY N., *Mathematical Biophysics. Physico-mathematical foundations of Biology*, p. IX, corsivi miei.

Ancora: di matematica della comprensione e del pensiero razionale, della visione, dei criteri estetici, dei processi di formazione di concetti astratti eccetera. Giunge ad estendere i propri metodi persino alla trattazione di fenomeni storico-sociali. I risultati in questione sono raccolti nel volume, pubblicato da MIT Press, dal titolo *Looking at History Through Mathematics*⁷⁰. I saggi li collezionati spaziano su una grande varietà di temi, ancora modellati matematicamente: interazione degli individui in società, velocità dei mutamenti comportamentali nelle società storiche, credenze e pregiudizi, effetto dei fattori geografici nello sviluppo delle società e ancora altro.

La tendenza alla trattazione formale di tutte le questioni superiori, a partire da quelle inerenti ai fenomeni cognitivi e ai meccanismi del pensiero razionale, è innegabilmente vicina ai – se non diretta fonte di ispirazione dei – coevi o poco più tardi tentativi di dare origine alla disciplina che andrà sotto il nome di *Artificial Intelligence*, il cui obiettivo consiste precisamente nel catturare il “pensiero” (sul difficile e troppo vago uso della nozione abbiamo già speso più di qualche parola) entro un qualche sistema formale.

Di nuovo, la biofisica cui Rashevsky dà vita fa uso della nozione di matematica in maniera assai diversa da quella dei suoi contemporanei: poco o per nulla incline al lavoro laboratoriale, Rashevsky intende la sua scienza in termini puramente teoretici, come il tentativo di fornire descrizione formale dei processi biologici (stesso trattamento ritiene di poter estendere ai sistemi complessi in generale).

Al contrario, l’uso di strumenti matematici tipico degli ambienti della biologia “classica”, meramente misurativo, è orientato alla sola analisi quantitativa dei fenomeni studiati, da svolgersi su collezioni di dati ottenuti per via osservativa-sperimentale.

⁷⁰ RASHEVSKY N., *Looking at History through Mathematics*, MIT Press, Cambridge, Massachusetts, 1968.

2.1.3 Sistemi complessi e riducibilità ai livelli inferiori

La posizione di Rashevsky è il prodotto, anche in questo caso, di una filosofia di stampo riduzionistico in senso proprio, il cui argomento può esprimersi come segue: siccome i sistemi biologici sono, al grado fondamentale, sistemi fisico-chimici, l'indagine che li riguarda può essere ricondotta ai metodi in uso nella fisica e nella chimica (notiamo – incidentalmente – che se la proposta formalistico-riduzionistica è problematica nel caso delle scienze biologiche, la molto maggiore complessità dei meccanismi cognitivi rispetto a quelli strettamente biologici autorizza, riguardo a tali opzioni nelle scienze cognitive, dubbi di grado ancora superiore).

Pertanto, la costruzione della teoria - a partire da qualche fatto “grezzo” e mediante la stretta deduzione garantita dal trattamento matematico - può essere identicamente condotta nelle scienze esatte e in quelle biologiche. Soltanto l'introduzione dell'analisi formale della matematica, e non invece il mero procedere sperimentale e osservativo, garantirebbe alla biologia l'accesso alle sue leggi fondamentali e alle sue proprietà più nascoste:

« Such an insight is possible only by mathematical analysis. [...] our experiments do not and cannot reveal those hidden and fundamental properties of Nature. [...] we must *infer*, from the wealth of known, relatively coarse facts, to the finer, not directly accessible fundamentals»⁷¹.

Sulla nozione di riduzionismo è precisamente questo aspetto ad aprire prospettive interessanti: la necessità – postulata da Rashevsky – di formulare il discorso biologico in termini fisico-chimici, dunque matematico-formali, si inserisce in un ben più ampio dibattito sulla autonomia epistemologica delle scienze biologiche in generale.

Se, infatti, i metodi del matematico russo hanno suscitato più di qualche perplessità presso la comunità scientifica dei biologi⁷², parallelamente hanno

⁷¹ N. RASHEVSKY, *Mathematical Biophysics*, Nature, vol. 135, pp. 538-530, 1935, p. 528.

⁷² Cfr. T. H. ABRAHAM, *op. cit.*, pp. 355-356: «[...] his exuberant imagination leads him to make extravagant claims. In this respect he is poetic, rather than scientific. In claiming that Nature must act in a certain way he shows immaturity».

raccolto il pieno interesse dei filosofi impegnati nell'opera di unificazione delle scienze e, più tardi, dei ricercatori nel campo della nascente IA.

In Rashevsky i positivisti logici hanno intravisto la possibilità di una riduzione effettiva della biologia alla fisica, tramite cui realizzare il sogno della matematizzazione delle scienze in generale: affinché esse risultino in un unico corpo omogeneo, è necessario che parlino un solo linguaggio, quello elementare – o fondamentale – della fisica.

Ogni fatto, ogni proposizione espressa entro ciascuna disciplina nell'alveo delle scienze, deve poter accedere ad una descrizione fisica, deve corrispondere ad una proposizione del linguaggio della fisica – pena la decadenza della sua stessa scientificità: quella fisicalista giunge così ad essere la dottrina dell'unità delle scienze e il riduzionismo di matrice fisicalista il paradigma che ne guida lo sviluppo teorico.

Non desta sorpresa, allora, che l'enfasi di Rashevsky sull'uso dei metodi della fisica teorica nelle scienze biologiche abbia attratto l'interesse dei grandi protagonisti della scuola viennese.

Nel mese di giugno del 1936, ad esempio, Rashevsky prese parte al Congresso Internazionale sull'unità della scienza tenutosi a Copenaghen. Tra gli altri, vi parteciparono pure Karl Popper, Otto Neurath e Niels Bohr. In quell'occasione discusse un articolo a nettissimo contenuto epistemologico, riguardante primariamente la relazione tra biologia e scienze formali e l'importanza e i limiti dell'utilizzo dei formalismi matematici nella costruzione della teoria dei fenomeni biologici.

A fare da contraltare all'interesse dei positivisti logici sta, tuttavia, l'evidenza della sua ricezione nell'accademia statunitense: come è stato anticipato, la biofisica matematica di Rashevsky godette di considerazione limitata presso la comunità dei biologi, e la sua ricerca smise – entro la metà del secolo – di essere sovvenzionata massicciamente dall'università di Chicago e dalla Rockefeller Foundation⁷³.

Ciononostante, mentre la comunità dei biologi classici conservava il proprio scetticismo, considerando l'approccio di Rashevsky troppo astratto,

⁷³ Cfr. T. H. ABRAHAM, *op. cit.*, p. 369 e ss.

eccessivamente formalistico e lontano dalle tradizioni empiriche della biologia, i pionieri del movimento cibernetico e padri dell'IA, (Marvin Minsky, Warren McCulloch, Walter Pitts, Allen Newell, Herbert Simon, W. J. Rapaport ecc.), abbracciavano entusiasticamente le sue intuizioni. In esse scorgevano il terreno fertile per lo sviluppo delle loro idee sulla replicabilità delle funzioni cognitive mediante meccanismi computazionali e strutture logiche adeguatamente costruite.

Con i metodi di Rashevsky, dunque, i padri dell'Intelligenza Artificiale hanno ereditato, automaticamente, pure i limiti concettuali ad essi correlati, i quali abbiamo provato a mostrare nelle pagine precedenti.

Ciò fornisce un indizio ulteriore a sostegno dell'ipotesi di partenza: che, cioè, sin dalla preistoria dell'IA – convenzionalmente rintracciabile nelle ricerche turinghiane – il tentativo di descrivere i processi del pensiero in termini formali e meccanicamente riproducibili è stato viziato dai problemi sottesi ad ogni approccio formalistico (sempre implicante la triade variabile astrazione/semplicificazione/costruzione di enti e concetti *ad hoc*) e di ogni riduzionismo (che invece soprassiede sulla necessaria lettura multilivellare dei fenomeni complessi).

2.2 McCulloch e Pitts: la cognizione come calcolo logico neuronale

La diretta influenza dell'opera di Rashevsky si registra anche nell'articolo pubblicato, nel 1943, insieme dal neurofisiologo Warren McCulloch e dal logico e matematico Walter Pitts e intitolato *A Logical Calculus of Ideas Immanent in Nervous Activity*⁷⁴. Si tratta del primo tentativo di esibizione di un modello computazionale di neurone.

La teoria sviluppata da McCulloch e Pitts – indirizzata a mostrare, mediante tecniche di formalizzazione logica, il modo in cui i meccanismi neurali potrebbero svolgere funzioni mentali – possiede un peso peculiare nell'economia del nostro discorso: si tratta infatti del primo uso storicamente

⁷⁴ W.S. McCULLOCH, W. PITTS., *A Logical Calculus of the Ideas Immanent in Nervous Activity*, in *Bulletin of Mathematical Biophysics*, 5, 1943; rist. in McCulloch, W.S., *Embodiment of Mind*, MIT Press, Cambridge, 1965.

rilevabile di concetti computazionali nell'indagine sul problema, assai antico, del rapporto mente-corpo e insieme della prima formulazione di una teoria computazionale del cervello e della mente. Il fatto sopra esposto trova svariate conferme in letteratura. Si veda, ad esempio, l'analisi proposta da Gualtiero Piccinini:

«Before McCulloch and Pitts, neither Turing nor anyone else had used the mathematical notion of computation as an ingredient in a theory of mind and brain. [...] McCulloch and Pitt's theory changed the intellectual landscape, so that many could see neural computations as the most promising way to explain mental activities.»⁷⁵

Esiste dunque, ricostruendo la storia dei concetti in questione, una non trascurabile evidenza a sostegno dell'idea che nel lavoro dei due studiosi risieda l'inizio della transizione verso modelli computazionali di spiegazione dei processi mentali.

Come si è mostrato, l'uso della matematica per la modellizzazione dei fenomeni biologici è ampiamente attestato nel clima della ricerca novecentesca: la novità introdotta da McCulloch e Pitts consiste invece nell'uso della nozione logico-matematica di computazione – che abbiamo visto sviluppata da Alan Turing e Alonzo Church negli anni '20 del Novecento – per mostrare come le dinamiche neuronali possano spiegare l'insorgenza di funzioni mentali di più alto livello.

Turing stesso propose una teoria computazionale del cervello – accostandolo ad una “macchina calcolatrice digitale”, sebbene solo in maniera analogica e senza impegnarsi in una giustificazione formale – o nello sviluppo di una teoria computazionale del ragionamento logico (necessità alla soddisfazione della quale, invece, lavorano i due studiosi statunitensi).

⁷⁵ G. PICCININI, *The First Computational Theory of Mind and Brain: A Close Look at McCulloch and Pitts's "Logical Calculus of Ideas Immanent in Nervous Activity*, Synthese, vol. 141, n. 2, pp. 175-215, p. 176, 2004.

Tenteremo di mostrare in quali modi i modelli computazionali dei processi cognitivi si innestino su tali teorie di matrice fisicalista e – ancora prima – quale sia il retroterra storico e filosofico delle ultime.

Uno sguardo al *milieu* teorico di McCulloch e contemporanei rivela la convinzione, da parte di una mole considerevole di studiosi, di dover indirizzare le teorie concernenti la mente proprio verso spiegazioni di tipo neurale:

« He believed that the goal of neurophysiology and psychiatry was to explain the mind in terms of neural mechanisms, and that scientists had not seriously tried to construct a theory to this effect. »⁷⁶

Per inciso, si può ritenere che tale programma abbia ottenuto successo stabile e duraturo, considerando che, soprattutto a partire dall'altrettanto stabile successo delle neuroscienze cognitive, le spiegazioni neurologiche dei fenomeni mentali sono assurte al grado di *standard* teorico – grado tuttora mantenuto.

Ai risultati esposti in *A Logical Calculus of Ideas Immanent in Nervous Activity* hanno contribuito in modo decisivo gli ambienti intellettuali della ricerca neurofisiologica e matematica statunitense degli anni Venti del Novecento.

McCulloch, che – sulla scia del lavoro di Rashevsky – nutriva grande interesse per l'importazione di metodi matematici nell'ambito delle scienze biologiche, tentò difatti alcune primissime applicazioni del calcolo logico alle reti nervose. Tale interesse poté svilupparsi robustamente a partire dalla frequentazione, a Yale, di un ciclo di seminari presso il dipartimento di filosofia: egli conobbe i *Principia Mathematica* di Withehead e Russell e acquisì familiarità con la riflessione teorica di J. H. Woodger, il quale a sua volta tentava l'uso dei metodi logici dei *Principia* nell'assiomatizzazione della teoria biologica⁷⁷.

Un'influenza decisiva ebbe pure, come si è stabilito, la biologia matematica di Rashevsky:

«Rashevsky's method involved making assumptions about the essential features of a biological process, representing those features as mathematical

⁷⁶ *Ivi*, p. 177.

⁷⁷ Cfr. *Ivi*, pp. 179-180.

variables, connecting the variables through equations that satisfy the initial assumptions, and studying the equations to investigate possible mechanisms underlying the biological process.»⁷⁸

Si può scorgere, nella formulazione del lavoro di McCulloch e Pitts, una aderenza pressoché pedissequa ai metodi seguiti da Rashevsky, con la sola differenza – facilmente evincibile dalla comparazione dei risultati, e in più sottolineata da Piccinini – dell'utilizzo della matematica discreta della logica formale al posto di quella continua del calcolo integrale e differenziale⁷⁹.

2.2.1 Atomi psicologici e altri problemi delle teorie riduzionistiche

I primissimi tentativi, da parte di McCulloch, di sviluppare una teoria psicologica di matrice meccanicistica, risalgono agli anni Venti del Novecento. La teoria prevede, ai fini della spiegazione dei fenomeni psicologici, l'introduzione di eventi mentali atomici chiamati “psiconi”, in analogia con gli atomi della fisica:

« My object, as a psychologist, was to invent a kind of least psychic event, or “psychon” that would have the following properties: first, it was to be so simple an event that it either happened or else it did not happen. Second, it was to happen only if its bound cause had happened...that is, it was to imply its temporal antecedent. Third, it was to propose this to subsequent psychons. Fourth, these were to be compounded to produce the equivalents of more complicated propositions concerning their antecedents »⁸⁰.

Il passaggio, oltre ad illuminare le proprietà degli “atomi psichici” postulati da McCulloch, conferisce credito ulteriore alla tesi centrale del capitolo - riguardante i limiti dei processi di riduzione e formalizzazione - a favore della quale si continuerà ad argomentare. In questo caso, lo “psicone” non è introdotto come conseguenza di una procedura empirico-osservativa, ma come ente

⁷⁸ *Ivi*, p. 181.

⁷⁹ *Ibid.*

⁸⁰ W.S. McCULLOCH, *What is a Number, that a Man May Know It, and a Man, that He May Know a Number?*, in *General Semantics Bulletin*, 26/27, pp. 7-18, 1961, rist. in. McCulloch, *Embodiments of Mind*, MIT Press, Cambridge, 1965.

necessario allo sviluppo e alla tenuta di una teoria computazionale della cognizione.

McCulloch, infatti, assegna allo psicone un contenuto proposizionale e, su tale base, tenta la costruzione di un calcolo logico-proposizionale dei fenomeni mentali. Il tentativo originario finirà per naufragare, ma l'idea di applicare la nozione di calcolo al sistema nervoso continuerà ad essere la colonna portante del programma di ricerca del neurofisiologo.

Venuto a contatto, presso il Dipartimento di Filosofia di Yale, con la notazione simbolica e il sistema deduttivo dei *Principia Mathematica* di Whitehead e Russell, McCulloch tentò ancora – con l'ausilio dei nuovi strumenti concettuali – l'applicazione della logica booleana al comportamento delle reti nervose. Suppose, infatti, di poter trasporre la proprietà fondamentale del suo psicone (l'essere in attività oppure in stasi – *i. e.* l'evento si verifica oppure no) nella notazione bivaloriale della logica algebrica di Boole, le cui variabili possono assumere solo i due valori “vero” o “falso”, rispettivamente denotati con 1 e 0. Nella sua versione aggiornata, la teoria identifica lo psicone – ente, di fatto, immaginario – con l'attività di trasmissione di impulsi elettrici di una rete neurale idealizzata. Si noterà il carattere arbitrario della manovra.

Similmente, Walter Pitts, dopo aver studiato logica con Carnap e prima di venire a contatto con i tentativi di McCulloch, divenne membro del gruppo di ricerca di Rashevsky. Sulla scia delle intuizioni del maestro, il quale lavorò lungamente alla descrizione formale dei meccanismi cognitivi⁸¹, poté sviluppare l'idea che le strutture cerebrali svolgano nient'altro che funzioni logiche e che, dunque, il cervello sia precisamente una “macchina logica”. Jerome Lettvin, celebre neuroscienziato statunitense, il quale svolse ricerca con Pitts stesso, afferma:

⁸¹ Citiamo, di nuovo, i saggi raccolti in *Mathematical Biophysics. Physico-Mathematical Foundations of Biology*. Tra gli altri, trattano questioni neurali-cognitive gli articoli: *Mathematical Biophysics of Rational Learning and Thinking*, *Mathematical Biophysics of Abstraction*, *Mathematical Biophysics of Some Mental Phenomena*, *Mathematical Biophysics of Some Simple Neurological Structures*, *Mathematical Biophysics of the Gestalt Problem*, *Mathematical Biophysics of Color Vision and of Flicker Phenomena*, *Applications of the Foregoing to Learning*.

«Quite independently, McCulloch and Pitts set about looking at the nervous system itself as a logical machine.»⁸²

In maniera indipendente i due studiosi giunsero alla convinzione di poter spiegare la natura logico-computazionale dei processi mentali, e di poterlo fare interrogando le architetture neurali ad essi sottese. Tale programma di ricerca era – come abbiamo visto – fortemente ancorato ai paradigmi teorici del contesto di riferimento.

La convergenza delle due ricerche autonome si realizza infine nell'articolo scritto a quattro mani e pubblicato, nel 1943, sul *Bulletin of Mathematical Biophysics* – rivista fondata e diretta da Rashevsky, nonché luogo di pubblicazione pressoché esclusivo delle ricerche entro l'ambito della nascente biofisica matematica.

Si deve subito rilevare che il tentativo di McCulloch e Pitts produce, sul versante filosofico, un problema cruciale, il quale possiamo esprimere nelle seguenti sotto-articolazioni: in primo luogo, si pone la necessità di mostrare *come* i fenomeni neurali possano essere formalizzati mediante l'uso della logica proposizionale; poi si dovrà spiegare come essa possa catturare, quando utilizzata come linguaggio formale utile alla modellizzazione degli impulsi neuronali, la vastissima gamma dei processi mentali; infine – e questo è il fatto di maggiore peso teorico – assumendo come risolti i due compiti precedenti, si adotta automaticamente, rispetto al problema del rapporto mentale-cerebrale, un approccio di tipo monistico e strettamente fisicalista. In altri termini, si elimina *tout-court* la distinzione, assai problematica, tra mentale e cerebrale e si istituisce una teoria tutta sbilanciata sul secondo dei due termini.

Il problema del rapporto tra mentale e cerebrale è un impedimento primario al tentativo di fornire una teoria della mente basata sui meccanismi neurali. McCulloch pare perfettamente conscio del peso della questione, che infatti enuncia in modo esplicito:

⁸²G. PICCININI, *op. cit.*, p. 185.

«We have a dichotomy in medicine, which has grown increasingly. Psychiatric approach on one side, particularly the psychoanalytic approach, has produced one group; the organic approach to the physiology of particular organs and disease processes has made organicists of another group [...]. *I mean the dichotomy between mind and body.*»⁸³

Ciò che si riferisce al “mentale” descrive processi psicologici, coscienza, intenzionalità, idee, eccetera; ciò che si riferisce al “cerebrale” descrive processi corporei, fisici, che riguardano configurazioni molecolari e organiche, scambi biochimici, eccetera.

Un tale abisso tra psicologico e fisiologico non può essere concepito senza che si generino inconsistenze teoriche, quindi deve essere colmato.

Il modo in cui si intende risolvere il problema consiste nel mostrare come un elemento simil-neuronale possa incarnare delle funzioni logico-razionali o cognitive, attribuendogli una struttura causale simile a quella della logica proposizionale.

McCulloch e Pitts istituiscono una diretta corrispondenza formale tra eventi neurali e funzioni logiche (interpretate come “idee”, o particolari funzioni mentali), per supporre infine che essa basti a trasformare i primi nelle seconde.

Anche qui il problema è duplice, e riguarda: 1) l’effettiva possibilità di stabilire un’equivalenza tra gli scambi elettrochimici delle sinapsi e i metodi simbolico-deduttivi del calcolo logico – che è soltanto analogica; 2) l’interpretabilità dei sistemi formali di proposizioni (le regole del calcolo logico) nei termini di fenomeni cognitivi *tout court*. Entrambe le mosse paiono arbitrarie e viziate da fallacia.

Il modello che i due studiosi propongono esprime una concezione fisicalista della mente; si può dunque ritenere – considerando l’uso di concetti computazionali che McCulloch e Pitts mettono in campo – che esso funga da anello di congiunzione tra le teorie riduzionistiche di derivazione turinghiana (e nelle versioni forti, e in quelle deboli), la cui opera di riduzione verte sull’uso della

⁸³ *Ivi*, p. 186, corsivi mei.

nozione di calcolo, e i successivi impianti teorici – *à la* Rashevsky – fondati sulla spiegazione matematico-fisiologica dei fenomeni cognitivi.

Così trova risposta la questione che era stata presentata più in alto, la quale chiedeva di esibire il modo in cui le teorie della famiglia computazionale – spesso ancora contemplanti un qualche tipo di dualismo tra psichico e fisiologico – si siano infine innestate su modelli della mente decisamente anti-dualisti ed organicistici.

2.2.2 *Neuroni artificiali, reti logiche*

Per allargare ulteriormente la nostra analisi dobbiamo presentare – quantomeno informalmente – le reti di neuroni artificiali progettate da McCulloch e Pitts. Un'esposizione sintetica della teoria ci consentirà di ponderare le conseguenze generali della loro proposta e di avanzare alcune considerazioni critiche.

Registrate le necessarie ipotesi di lavoro, sulle quali dovremo tornare più tardi (ad esempio: che l'attività di ciascun neurone segua la legge del “tutto o nulla” - ovvero che sia sempre in attività oppure in stasi -, che la struttura delle reti rimanga invariata nel tempo, che l'attività delle sinapsi inibitorie blocchi gli impulsi dei neuroni che la subiscono di volta in volta, eccetera), e introdotta l'idea centrale del lavoro (che le dinamiche neuronali siano equivalenti a sistemi di proposizioni formali), i due studiosi procedono a tracciare lo schema dei loro sistemi neurali artificiali.

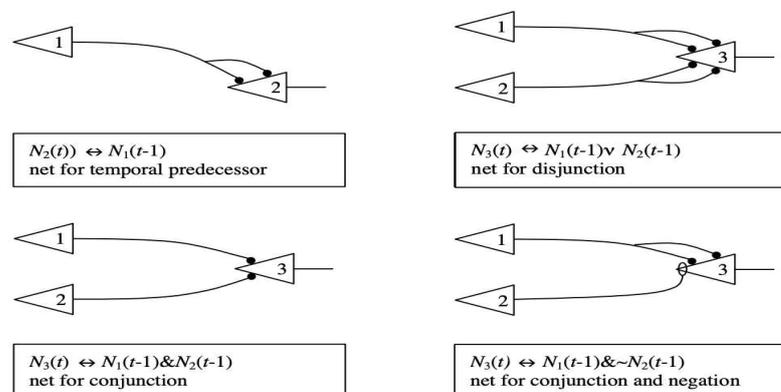
Denotano i neuroni di una rete idealizzata N con i simboli c_1, c_2, \dots, c_n , determinando così le espressioni primitive del sistema, nella forma $N_n(t)$. In analogia con i neuroni naturali, che “scaricano” un impulso elettrico ad una determinata soglia di attivazione, l'espressione $N_1(t)$ indica che il neurone c_1 scarica nell'istante t .

Tali espressioni primitive possono essere combinate mediante connettivi logici, a costituire formule complesse che descrivono il comportamento di più neuroni in istanti diversi, entro una rete. Ad esempio, l'espressione $N_1(t) \vee N_2(t - I)$ indica che o c_1 scarica nell'istante t o c_2 scarica nell'istante $t - I$.

È emblematico, e ancora conforme ai caratteri di arbitrarietà che ci paiono caratterizzare l'operazione dei due studiosi (nonché, a livelli operativi specifici, tutti i processi di formalizzazione) che McCulloch e Pitts ammettano tra i connettivi utili allo costruzione del loro sistema formale: congiunzione ($A \wedge B$), disgiunzione ($A \vee B$), congiunzione con negazione ($A \wedge \neg B$); e che a questi aggiungano, di nuovo secondo le esigenze particolari della loro teoria, un operatore logico peculiare (S) la cui funzione consiste nello spostare indietro nel tempo l'indice temporale di un'espressione della forma $N_n(t)$, come segue: $S[N_n(t)] = N_n(t - 1)$.

Furi dalla notazione restano il connettivo classico di negazione ($\neg A$), il condizionale materiale o implicazione semplice ($A \rightarrow B$) e il bicondizionale o doppia implicazione ($A \leftrightarrow B$), che pure costituiscono l'ossatura di ogni linguaggio logico-formale proposizionale/enunciativo classico. Il fatto non è secondario: sebbene siano molte le estensioni della logica classica che introducono connettivi ulteriori o che rifiutano uno o più di quelli tradizionali, McCulloch e Pitts intendono dimostrare che le funzioni neurali equivalgono formalmente alle operazioni ammesse entro un linguaggio logico di tipo proposizionale. Operano dunque descrivendo, secondo i metodi introdotti più in alto, alcune reti di neuroni che soddisfano le relazioni logiche di congiunzione, disgiunzione, congiunzione con negazione e predecessore temporale, condensate nei diagrammi seguenti:

Figura 1 - Diagrammi delle reti soddisfacenti le funzioni dei quattro connettivi logici di McCulloch e Pitts.



(Fonte: G. PICCININI, *The First Computational Theory of Mind and Brain*, p. 195)

I diagrammi superiori costituiscono i blocchi primitivi che, se opportunamente combinati, consentono di costruire circuiti neuronali che soddisfano determinate funzioni logiche.

Pur non essendo i tecnicismi formali della teoria di interesse del nostro discorso, sarà necessario riferirsi ancora ai diagrammi di McCulloch e Pitts quando affronteremo il problema della legittimità operativa di ogni approccio riduzionistico-formalizzante nella modellizzazione dei sistemi complessi.

2.2.3 Sull'equivalenza tra reti di McCulloch-Pitts e funzioni Turing-computabili

Procedendo nell'esposizione della teoria, i due autori suppongono che le loro reti logiche possano calcolare ogni funzione computabile secondo Turing-Church senza precisare, tuttavia, in che senso esse operino tali calcoli. Verosimilmente, indicavano la possibilità di descrivere il comportamento delle reti ricorrendo al lessico e ai formalismi della teoria della computabilità effettiva di Turing. Il legame tra la loro teoria e la computabilità di Turing è enunciato *en passant*:

«It is easily shown: first, that every net, if furnished with a tape, scanners connected to afferents, and suitable efferents to perform the necessary motor-operations, can compute only such numbers as can a Turing machine»⁸⁴

Quello precedente è il solo passaggio in cui gli autori connettono il loro lavoro con quello turinghiano nel campo della computabilità, e tuttavia è un punto essenziale in prospettiva storico-scientifica: trattasi, infatti, della prima occorrenza storica - esplicita - dell'innesto della computabilità turinghiana su un modello scientifico delle funzioni cerebrali. Con altri, semplificatori toni, potremmo dire: della matematica di Turing nelle scienze cognitive.

⁸⁴ W.S. McCULLOCH, W. PITTS., *A Logical Calculus of the Ideas Immanent in Nervous Activity*, p. 113.

In esso si può, dunque, scorgere l'inizio effettivo – e non soltanto suggerito – del paradigma computazionale che ha guidato lo sviluppo della teoria entro le scienze cognitive in generale.

Peraltro, ciò che le reti possono calcolare non è dimostrato nell'articolo, e neppure si dimostra la loro equivalenza formale con le Macchine di Turing, che è solo enunciata. Poco più avanti infatti, nello stesso paragrafo, si afferma:

«This is of interest as affording a psychological justification of the Turing definition of computability and its equivalents, Church's A-definability and Kleene's primitive recursiveness: if any number can be computed by an organism, it is computable by these definitions, and conversely»⁸⁵

L'affermazione non deve passare sottotraccia, giacché tradisce un'ambizione considerevole: che si intenda, cioè, il *Logical Calculus of Ideas Immanent in Nervous Activity* quale giustificazione psicologica della tesi di Church-Turing. Essa – come si ricorderà – dice, sulla base di alcune considerazioni matematiche intuitive, che ogni funzione calcolabile da un essere umano è altresì computabile da una macchina di Turing.

Stabilire allora l'equivalenza tra reti di McCulloch-Pitts e macchine di Turing significa suggerire, implicitamente, che le reti catturano i limiti dei fenomeni cognitivi di cui i neuroni biologici sono substrato – determinando esse ciò che ogni umano dotato di carta, penna e capacità di calcolo può effettivamente computare.

Trattare il cervello come una macchina di Turing: ecco l'enorme ambizione, e carica di conseguenze, che informa la ricerca dei due studiosi e che informerà variamente, dopo di loro, le scienze cognitive in generale. Anch'esse, infatti, costruiranno il loro statuto intorno al medesimo paradigma computazionale-cerebrale.

McCulloch e Pitts hanno avanzato una proposta di riduzionismo esplicito, riguardante: 1) la possibilità, apparentemente garantita dalla loro teoria, di ridurre i fenomeni psicologici a quelli neurofisiologici, la psicologia alla

⁸⁵ *Ivi.*

neurofisiologia; 2) l'idea che la relazione fondamentale tra i fatti psicologici sia di carattere logico, precisamente proposizionale e bivaloriale:

«Thus in psychology, introspective, behavioristic or physiological, the fundamental relations are those of two-valued logic. [...] the activity is inherently propositional»⁸⁶

A ciò si aggiunge, a formare la caratteristica triade formalismo-riduzione-determinismo, la convinzione che sia finalmente possibile superare ogni ambiguità aperta dalla scissione tra mentale e cerebrale, e che ogni fenomeno afferente alla sfera del mentale (sia esso percettivo, razionale, emotivo ecc.) possa ritenersi rigorosamente (o, per l'appunto, deterministicamente) derivabile dai fatti neurofisiologici:

«[...] both the formal and the final aspects of that activity which we are want to call *mental* are rigorously deducible from present neurophysiology»⁸⁷

Al netto dei toni assertivi, però, parti fondamentali della teoria, e anche varie sue conseguenze ottimisticamente prospettate come risolutive, presentano tratti di arbitrarietà, quando non addirittura carattere dogmatico.

Ne abbiamo proposti esempi diversi, ma è bene riassumerli ancora e, se possibile, espanderne l'elenco.

2.2.4 *Qualche problema ulteriore*

2.2.4.1 *Si fa presto a dire "calcolo"*

Si parta dal titolo: McCulloch e Pitts chiamano *logical calculus* il risultato delle loro ricerche. A ben guardare, però, l'attribuzione è problematica.

Di tale problematicità rende conto il logico Friedrich Fitch che, nella revisione – datata 1944 e pubblicata sul *Journal of Symbolic Logic* – del celebre articolo, spiega perché la loro operazione non possa dirsi “calcolo logico” in senso pieno.

⁸⁶ *Ibid.*, p. 114.

⁸⁷ *Ivi.*

Formalmente, infatti, un calcolo non è che una combinazione di:

1. *Grammatica*. Un insieme di simboli e regole utili alla costruzione di formule ben formate, ovvero di adeguate sequenze simboliche.
2. *Sistema deduttivo*. Un sistema deduttivo è necessario per individuare le espressioni che possono essere intese come assiomi e quelle che possono invece derivarsi dalle prime. Esso deve includere, inoltre, una procedura effettiva per stabilire se una successione qualsiasi di espressioni soddisfi o meno le regole del sistema deduttivo.

McCulloch e Pitts, tuttavia, pur costruendo una grammatica utile alla rappresentazione di una classe specifica di proposizioni, non la corredano di alcun sistema di deduzione.

Si limitano, invece, ad assumere quali assiomi un insieme di ipotesi semplificatorie ed idealizzanti sulla natura dei neuroni – così da definire una classe di reti di neuroni altrettanto idealizzati.

Il problema, che potrebbe apparire come una banale sottigliezza di segno tecnico, è invece essenziale, visto che larga parte del successo della teoria si deve alla sua pretesa di aver *dimostrato formalmente* la piena plausibilità del carattere logico-proposizionale dell'attività delle reti.

2.2.4.2 Ancora sull'equivalenza di reti logiche e macchine di Turing

Identica preoccupazione grava sulla già presentata equivalenza tra *logical calculus* e Macchine di Turing, la quale gli autori danno per assodata, che invece è solo enunciata informalmente.

Dopo la pubblicazione dell'articolo, furono molti gli studiosi ad assumere acriticamente, come *standard* teorico, l'interpretazione dell'attività neurale come fenomeno computazionale, ritenendola fondata su una prova matematica effettiva – che tuttavia non esiste.

Le conseguenze sono autoevidenti: si cominciò a ritenere che il “calcolo” non fosse una sola tra le molte e differenti attività umane, ma che, in qualche modo, fosse la sola, propria attività umana: ogni fatto mentale assunse così

connotati computazionali, e ogni comportamento poté essere pensato nei termini di *output* seguente un calcolo.

Allo stesso modo, la tesi di Church-Turing conobbe uno slittamento di significato e di contesto, passando da indicare quali funzioni possono essere computate a stabilire le potenzialità e i limiti strutturali del cervello umano.

È grazie al (o per causa del) lavoro di McCulloch e Pitts, ci pare di vedere, che si sono potute pensare la ricerca e l'opera di Alan Turing come una esplicita difesa del computazionalismo cognitivo. A questa evidenza si può riferire la confusione, già rilevata nel capitolo precedente, tra le due differenti versioni del riduzionismo turinghiano: quella “forte”, dunque, al contrario di quella “debole” o operativa, non sarebbe altro che una ridefinizione *ex post* del lavoro di Turing, opportunamente forzata sui suoi risultati matematici.

2.2.4.3 *Semplificare, astrarre, idealizzare*

Si è già parlato delle mosse iniziali che garantiscono agli autori la costruzione dell'impianto generale della loro teoria. Essa comincia prendendo in considerazione alcune condizioni fondamentali, assunte in qualità di “assiomi” del sistema formale:

1. L'attività neuronale segue la legge del “tutto o nulla”
2. L'attività delle sinapsi inibitorie previene l'attivazione neuronale
3. Le reti posseggono struttura fissa e invariabile

Tali “assiomi” costituiscono una decisiva semplificazione e una altrettanto marcata idealizzazione delle allora note proprietà dei neuroni biologici.

McCulloch e Pitts riconoscono il carattere arbitrario delle premesse (2) e (3), che non rispecchiano l'assai più complessa natura delle reti neurali biologiche.

Suppongono inoltre, per la realizzabilità stessa del loro calcolo logico, che i neuroni di una rete operino i propri “scambi” informativi entro intervalli temporali fissi. È, questa, una condizione fondamentale: essa produce una “discretizzazione” della dinamica delle reti, la quale consente di descrivere schemi di impulsi neurali mediante la matematica discreta della logica proposizionale.

In altri termini, l'attribuzione di una natura "essenzialmente proposizionale" all'attività dei neuroni, che costituisce il fulcro della teoria, consentendo di assegnare valore cognitivo (le "idee" emergenti dal calcolo logico) alla dinamica delle reti, dipende primariamente da un'opzione arbitraria degli studiosi, che scelgono di impiegare la notazione logico-proposizionale in luogo della matematica continua del calcolo integrale e differenziale.

Che i risultati della teoria dipendessero da particolari scelte di astrazione, da opzioni idealizzanti e da meccanismi semplificatori, d'altra parte, era cosa perfettamente nota agli autori. McCulloch afferma: «Formal neurons were deliberately as impoverished as possible»⁸⁸.

Semplificare, astrarre, idealizzare: sono, questi, meccanismi *assolutamente* indispensabili ad ogni operazione riduzionistica, che, per tali ragioni, sempre risulta in una compressione della complessità del fenomeno su cui agisce, e quindi in uno scollamento di entità variabile tra i suoi risultati formali e la realtà empirica che intende descrivere.

Per tutte queste ragioni, le opzioni riduzionistiche e i formalismi che di queste sono strumento posseggono un preciso quadro di legittima applicabilità, fuori del quale producono problemi maggiori di quelli che risolvono. Sulla questione si tornerà tra poco.

2.2.4.4 Reti sintattiche, reti semantiche

Nella sezione dedicata alle conseguenze epistemologiche della teoria, McCulloch e Pitts sostengono, compiendo uno dei loro caratteristici salti concettuali, che:

«Since that activity [quella del singolo neurone] is inherently propositional, all psychic events have an intentional, or "semiotic", character»⁸⁹

⁸⁸ W. S. McCULLOCH, *Recollections of the Many Sources of Cybernetics*, ASC Forum, VI(2), 5–16, 1974.

⁸⁹ W.S. McCULLOCH, W. PITTS., *A Logical Calculus of the Ideas Immanent in Nervous Activity*, p. 114.

La tesi, in questo caso, si riferisce alla precedentemente citata emergenza di significati primitivi dalla combinazione di funzioni formali.

Si è stabilito che il cardine del *Logical Calculus* è nell'attribuzione di un carattere proposizionale, enunciativo, all'attività delle reti neuronali. Abbiamo dimostrato che: (1) i neuroni, e dunque le reti formali costruite dagli autori, presentano tutti i limiti derivanti dalle operazioni di idealizzazione e semplificazione del loro corrispettivo biologico. Pertanto, se ne allontanano in modo irrimediabile; (2) l'assegnazione, alle reti medesime, di un contenuto logico-proposizionale, è possibile solo sulla scorta di precise scelte metodologiche – e si avvale comunque delle operazioni di cui al punto (1), risultando così come un utile artificio formale, e non invece come una caratteristica plausibile delle reti neurali biologiche.

Da tutto ciò consegue che, nell'attribuire qualità “intenzionali” o “semiotiche” alle reti logiche, McCulloch e Pitts si arrischiano in una scelta che aggrava l'impronta arbitraria della loro operazione: non solo essi riconoscono – senza prova effettiva – un contenuto proposizionale alle reti, ma si spingono ad affermare che da esso sorgono singoli brandelli di significato i quali, ricomposti, renderebbero conto dei fatti psicologici – intenzionali, per l'appunto, dotati di carattere semantico – nella loro interezza.

In altri termini, presumono risolto il problema, tutt'oggi questionato, del rapporto tra sintassi e semantica, tra l'insieme dei significati e l'insieme delle regole formali. Siccome la questione è veramente aperta e dibattuta, non possiamo esprimerci definitivamente su di essa, ma possiamo illustrare velocemente qualche fatto noto.

In primo luogo, i sistemi di regole formali, da soli, non sono in grado di generare significati. Ogni regola istituisce una direttiva che informa sulle modalità di manipolazione simbolica senza, tuttavia, che ai simboli siano assegnati particolari connotati semantici.

Essi emergono, invece, come risultato di operazioni referenziali ad altissima complessità, sempre coinvolgenti fattori extra-formali: convenzionali, contestuali, sociali – in una parola: culturali.

Se un sistema formale esprime un significato di qualche genere, si dovrà allora ritenere che esso gli proviene, per così dire, dall'esterno.

Si consideri, ad esempio, un enunciato classico nel sistema formale logico-proposizionale (lo stesso, pressappoco, che individua le funzioni logiche della teoria di McCulloch e Pitts): $P \rightarrow Q$, rappresentante la regola di implicazione semplice.

Possiamo interpretare, correttamente, “P” e “Q” come variabili proposizionali e la regola intera come l’asserzione: “se P è vero, allora Q è vero”; possiamo costruire la matrice della regola e sapere come manipolarla entro il suo sistema formale di riferimento, ma perché ad essa corrisponda un qualche attributo semantico la semplice sintassi deve essere superata in favore della comprensione del significato di “se... allora...” nel linguaggio naturale, ovvero del senso “esterno” della relazione di implicazione.

Quindi si dovrà possedere una nozione “ingenua” o “naturale” di causalità, di effetto, di necessità, di relazione ecc.

In altri termini, la semantica di qualsiasi linguaggio (compresi quelli formali) sta in relazione con la sua sintassi in quanto elemento esterno o aggiuntivo: è la prima a garantire codifica e decodifica della seconda, e non viceversa.

Per tutte queste ragioni, ci sembra che il ponte che gli autori gettano tra ambito dei significati e ambito delle regole formali sia costruito su fondamenta effimere – addirittura inesistenti: l’argomento è facilmente destinato al collasso.

Tutte queste difficoltà non hanno ostacolato la diffusione della teoria, la quale ha riscosso, pur trascinandosi dietro ogni sua debolezza e inconsistenza, successo unanime in una varietà amplissima di ambienti scientifici – con conseguenze che pondereremo più avanti.

2.2.4.5 *Uno scherzo formalistico*

Alla tesi centrale del nostro discorso, relativa – per così dire – alla “*triviality*”⁹⁰ dei meccanismi di formalizzazione e riduzione, può essere riferito un aneddoto

⁹⁰ L’inglese *trivial* mi sembra, in questo caso, più adeguato al contesto del discorso rispetto a “banale”, sua controparte nella lingua italiana.

dai contorni umoristici, uno “scherzo formalistico” escogitato insieme da Jerome Lettvin e Walter Pitts negli anni trascorsi da membri dell’*entourage* di Rashevsky a Chicago.

Lo scherzo è esposto in un articolo intitolato *A Mathematical Theory of Affective Psychoses*⁹¹, pubblicato nello stesso anno del *Logical Calculus*.

Lettvin e Pitts, seguendo i metodi – ormai a noi noti – di Rashevsky, individuarono due variabili, φ e ψ , per rappresentare rispettivamente “l’intensità emotiva” e “l’intensità di attività” di un soggetto generico.

Utilizzando complessi sistemi di equazioni differenziali, scrissero le relazioni matematiche utili a descrivere le variazioni e le interazioni di φ e ψ al variare del tempo.

Procedettero infine a risolverle e a determinare un’analisi della stabilità e dell’instabilità del sistema individuato dalle variabili.

Sottoposero i loro sforzi all’esame del comitato scientifico del *Bulletin of Mathematical Biophysics*, presieduto da Rashevsky in persona, presentandoli come un modello formale utile a spiegare le dinamiche di insorgenza di certi particolari fenomeni psicotici (sindrome di Kahlbaum, psicosi reattive, disturbo bipolare).

Alla rottura delle condizioni di equilibrio del sistema, regolate dall’interazione tra le due variabili (*ça va sans dire*, individuate con la massima arbitrarietà e in più indicanti grandezze assolutamente immaginarie), la teoria faceva corrispondere la comparsa dei disturbi.

Nessuno, tuttavia, colse gli intenti parodistici dello studio, che prontamente venne pubblicato tra le pagine della rivista.

Così, una canzonatura spiritosa dei metodi del maestro poté diventare un risultato scientifico in piena regola, ovviamente lasciando gli autori assai divertiti.

L’aneddoto illumina un aspetto dominante del nostro ragionamento; l’abbiamo incrociato ripetutamente, ma abbiamo ora occasione di esprimerlo

⁹¹ J. L. LETTVIN, W. H. PITTS, *A Mathematical Theory of Affective Psychoses*, *Bulletin of Mathematical Biophysics*, 5, 139–148, 1943.

senza sottintesi: di ogni fenomeno, sempre si può fornire descrizione formale, purché si operi ad un livello di astrazione adeguato e si applichino su di esso le necessarie semplificazioni.

2.2.4.6 Sulla non-eccezionalità delle descrizioni formali

Lo “scherzo formalistico” di Lettvin e Pitts è certamente un caso limite della regola, la quale tuttavia si applica allo stesso modo – lo abbiamo visto – ai risultati di Rashevsky e di McCulloch.

Similmente le loro teorie richiedono l'introduzione di processi immaginari, di enti e concetti sviluppati *ad hoc*, l'uso di meccanismi di compressione della complessità del fenomeno originale. Ognuno di questi passaggi garantisce la formalizzabilità del fenomeno stesso – che risulta utile, a sua volta, come strumento per la sua riduzione ai livelli inferiori.

È in funzione di tutto questo che si ritiene necessario un ripensamento profondo di ogni approccio che – scorgendo un'eccezionalità nella descrizione formale di qualsivoglia fenomeno – insegue inconcludenti paradigmi teorici di tipo formalistico-riduzionistico, ritenendo bastevole, a garantire certezza e affidabilità scientifica alla teoria, qualsiasi descrizione matematico-formalizzante.

Si è mostrato, invece, che l'impiego di una formalizzazione funzionante non implica alcuna eccezionalità (dunque alcuna automatica *correttezza interpretativa* del fenomeno), potendosi invece svolgere su ogni fenomeno (e su ciascuno in molti modi differenti)⁹² e non essendo neppure qualificabile come processo sterile o neutrale. Al contrario, l'abbiamo vista sempre agire conservando margini variabili di arbitrarietà.

Nel capitolo seguente e ultimo tratteremo le conseguenze dell'adozione del paradigma computazionale entro le scienze cognitive, confrontandole con alcuni importanti argomenti limitativi e infine determinando il campo di legittima applicabilità delle operazioni riduzionistiche.

⁹² Si veda il caso di McCulloch e Pitts, i quali dimostrano che, per ciascuna funzione logica, è possibile costruire una varietà di reti che la soddisfano.

III.

EPISTEMOLOGIA DELL'IA COME MODELLO TEORICO DEI SISTEMI COGNITIVI

3.1 Genesi concettuale del paradigma computazionale

Sono stati raccolti, dall'analisi della nozione di computabilità effettiva (I) e poi dell'uso di strategie formali per la descrizione di sistemi ad elevata complessità (II), elementi in numero sufficiente.

È emersa una chiara insistenza, da parte degli studiosi citati, su strategie di ricerca e modelli teorici di carattere riduzionistico, che dal rinnovato successo delle opzioni meccanicistiche inaugurate dagli studi turinghiani hanno tratto nuova linfa e avallo definitivo.

Ci è parso, tuttavia, che nella generalizzazione che conduce dalla “macchina” come astrazione matematica ristretta al “meccanicismo” – qui inteso come regola generale di interpretazione, per così dire, trans-disciplinare – sia nascosto un errore interpretativo grande e risonante.

La confusione, qui, riguarda la difficoltà a leggere correttamente le implicazioni della matematica della computabilità.

Essa propone, si è visto, almeno due riduzionismi di segno diverso: uno forte, l'altro operativo; l'uso dell'uno o dell'altro come paradigma per la costruzione della teoria non è indifferente – conducendo anzi a conseguenze opposte.

Scegliere la corretta opzione riduzionistica: ecco il compito cruciale. Se ben eseguito, esso fornisce automaticamente gli adeguati limiti applicativi delle descrizioni formali, comunica ove si possa formalizzare, ove si possa ridurre con profitto. Nel caso contrario, si persevererà nella produzione di teoria affetta dalle medesime distorsioni, dagli stessi limiti che abbiamo visto all'opera nel secondo capitolo.

Per individuare i contesti di corretta applicazione, nell'ambito delle scienze cognitive, degli approcci riduzionistici e formalizzanti, possiamo utilmente ripartire dall'analogia tra cognizione umana e intelligenza meccanica.

Essa afferma che l'insieme delle operazioni svolte da una mente umana possa essere adeguatamente formalizzato ed eseguito da una macchina calcolatrice. Qui è la radice del computazionalismo cognitivo: se la mente è formalizzabile, allora essa opera solo e precisamente mediante il calcolo.

L'analogia mente-macchina, esempio eminente di riduzione, sostiene che sia possibile ridurre, a mezzo di procedure formali, la totalità delle operazioni della mente a insiemi di regole dirigenti l'azione di un calcolatore.

Ponendo l'idea computazionale sotto la lente d'ingrandimento, si rileva tuttavia la necessità di complicare la lettura della nozione di intelligenza. Per la sua vaghezza innegabile, correttamente evidenziata da Alan Turing, i propugnatori del computazionalismo cognitivo ne hanno ristretto l'estensione semantica: essa finisce per indicare tutti e soli i compartimenti logico-razionali della cognizione umana, da essa tagliando fuori i restanti, che poco hanno a che vedere con il logico, con il razionale - in una parola: con il formalizzabile.

Tale restringimento semantico non è certamente problematico in sé stesso: può bene immaginarsi ed essere concepito un uso differente della nozione, che includa nella sua definizione certi elementi a discapito di altri.

È, invece, estremamente problematico il salto che da esso conduce al paradigma computazionale della mente: acquisita una definizione ristretta dell'intelligenza (che, lo ripetiamo, è fatta coincidere con i suoi aspetti strettamente logico-razionali, più facili da interpretare come formalizzabili), essa è poi generalizzata, quindi ricondotta alla sua valenza originaria.

Schematicamente, si può riassumere l'operazione nei termini seguenti:

1. La nozione di intelligenza è eccessivamente vaga, quindi non può essere trattata scientificamente;
2. Per (1), la si deve restringere a ciò che di essa pare formalizzabile. L'operazione è legittima;
3. Adottata una definizione parziale di intelligenza, coincidente con il dominio di ciò che in essa è formalizzabile, di nuovo è estesa ad ogni aspetto della cognizione;

4. Si ritorna alla nozione originaria, che ora, tuttavia, intende l'intelligenza in senso ampio come completamente formalizzabile – dunque computazionale.

Al termine del processo la nozione, originariamente concepita come eccessivamente vaga e non formalizzabile, si può intendere *tout court* come formalmente descrivibile.

Pare evidente che il difetto all'origine dell'analogia sia localizzato al punto (3). Non è indebito il restringimento del campo della nozione; è tuttavia illegittima la sua generalizzazione.

Ciò è compatibile col nostro discorso, centrale quanto a rilevanza complessiva, sulla competizione tra riduzionismi di diverso segno: si può certamente impiegare una procedura di riduzione con fini operativi, purché non si pretenda di estendere i suoi risultati alla descrizione pre-riduzionistica del fenomeno.

L'assunzione dell'IA come modello per la descrizione dei fenomeni cognitivi, che è legata strettamente ai suoi successi tecnici, ha avuto un'altra, importante conseguenza: la tendenza a interpretare l'intelligenza nei termini del *problem solving*. Così concepita, essa riguarda innanzitutto la capacità di cercare e applicare procedure meccaniche per la soluzione di problemi.

Non a caso, i primi grandi avanzamenti dell'IA si sono dati intorno alla costruzione di algoritmi per la soluzione di giochi quali, ad esempio, gli scacchi. Allo stesso modo, ogni qualvolta un programma per computer performa meglio di un umano nello svolgimento di compiti simili, la notizia è presentata come un avanzamento decisivo verso la costruzione di qualcosa come una IA generale, o propriamente detta⁹³.

3.2 Forme d'inferenza

Occorre, però, che l'idea stessa della correlazione tra “razionale” e “formalizzabile” sia indagata più da vicino.

⁹³ La casistica è ampia e variegata. Per il prestigio generale degli avvenimenti in questione, si possono citare: *Deep Blue*, calcolatore sviluppato dalla statunitense IBM che, nel 1996, riuscì per la prima volta a battere il campione di scacchi Garry Kasparov; *AlphaGo*, software di *Google DeepMind* plurivincitore contro il campione europeo Fan Hui nel 2015.

Ciò che dell'intelligenza, nozione vaga, può, altrettanto vagamente, essere indicato come razionale, non costituisce in verità un blocco semplice ed omogeneo di funzioni, di abilità e di esperienze.

Per cercare cosa, nella cognizione, possa venire opportunamente accostato al computo di una macchina, alla semplice applicazione di regole di calcolo, è necessario che i processi che la caratterizzano siano esaminati secondo le possibilità e i limiti delle forme logiche che li catturano.

La più classica delle distinzioni, in questo particolare rispetto, è quella proposta da Charles S. Peirce, che individua tre forme fondamentali del ragionamento, il quale può avere carattere deduttivo, induttivo o abduttivo.

3.2.1 *Logic Theorist: formalizzabilità in senso proprio*

Per comprendere il peso della distinzione nello sviluppo della presente ricerca si può fornire un veloce resoconto intorno ad una delle pietre miliari dell'IA simbolica: l'algoritmo *Logic Theorist*.

Alla metà degli anni Cinquanta dello scorso secolo - prima ancora che l'IA come disciplina autonoma dell'informatica fosse inaugurata - Allen Newell, matematico e psicologo, Herbert Simon, economista, psicologo e informatico, futuro Premio Nobel per l'economia, e Cliff Shaw, informatico, scrissero il primo programma mai concepito per lo svolgimento di compiti di *automated reasoning*.

Logic Theorist era un sistema automatico per la dimostrazione di teoremi - una sorta di matematico virtuale: riuscì a provare in totale 38 dei 52 teoremi contenuti nel secondo capitolo dei *Principia Mathematica* di Whitehead e Russell, in più di un caso fornendo prove più eleganti o più dettagliate di quelle proposte dagli autori.

I successi tecnici di *Logic Theorist* ebbero, comprensibilmente, una risonanza decisiva in tutti gli ambienti; sembravano, infatti, corrispondere a un comportamento umano complesso, replicando e superando in eleganza le sottigliezze intellettuali necessarie ad ogni matematico per trovare la prova di un teorema. Herbert Simon si esprime entusiasticamente sulla portata filosofica dell'invenzione:

«[We] invented a computer program capable of thinking non-numerically, and thereby solved the venerable mind/body problem, explaining how a system composed of matter can have the properties of mind»⁹⁴

Nonostante l'ottimismo estremo di Simon e Newell, la ragione del successo del programma risiedeva in un fatto piuttosto semplice: i teoremi che dimostra e le procedure logiche secondo le quali la macchina è programmata si collocano nell'alveo della logica di tipo deduttivo.

I ragionamenti nella forma deduttiva coincidono con il dominio della logica classica, che include a sua volta una sconfinata varietà di elementi, dai sillogismi aristotelici alla geometria euclidea, sino agli algoritmi di IA simbolica di cui *Logic Theorist* fu il primo esempio.

Il nucleo di ogni ragionamento deduttivo consiste nel derivare, da alcune premesse di partenza (valide, nelle matematiche, come assiomi), le necessarie conclusioni.

Il grande vantaggio offerto dalla deduzione è il suo assoluto rigore formale: essa può essere trattata matematicamente e le sue conclusioni, quando ben dedotte, considerate garanzia della correttezza del ragionamento. Il limite evidente della logica deduttiva, come è noto, è la sua analiticità, ossia il fatto che le conclusioni dell'operazione formale siano già contenute nelle premesse. Esse sono soltanto rese esplicite, non producendo alcuna nuova conoscenza.

I ragionamenti catturati dalla logica deduttiva sono, si diceva, formalizzabili in senso pieno: in questo risiedeva il successo del programma di Newell e Simon. Questo anche, quindi, motiva la convinzione di poter replicare il pensiero umano nei termini di un sistema formale, con la sua grammatica e le sue regole di calcolo.

Erik J. Larson, che nel suo *The Myth of Artificial Intelligence*⁹⁵ segue la distinzione proposta da Pierce, ne fa l'ossatura di un lucido argomento storico-

⁹⁴ D. CREVIER, *AI: The Tumultuous Search for Artificial Intelligence*, BasicBooks, Harper Collins Publishers, New York, 1993, p. 46.

⁹⁵ E. J. LARSON, *The Myth of Artificial Intelligence. Why Computers Can't Think the Way We Do*, The Belknap Press of Harvard University Press, Cambridge, Massachusetts, 2021.

tecnologico sulle possibilità e i limiti dell'IA come modello della cognizione umana.

Il cardine dell'argomento è che, se è vero che i comportamenti e le abilità rispondenti a schemi logico-deduttivi possono intendersi come formalizzabili (dunque replicabili tecnicamente), lo stesso non vale per quelli che cadono sotto il dominio delle operazioni di carattere induttivo e – in misura ancora maggiore – abduttivo.

3.2.2 Induzione e abduzione: formalizzabilità in senso analogico

Nella storia della ricerca in IA, l'aumentata disponibilità di dati garantita dalla nascita del *World Wide Web* ha favorito un cambio di paradigma nello sviluppo di sistemi "intelligenti", che ha dunque cominciato ad occupare lo spazio delle logiche di carattere induttivo, ovvero statistico-probabilistico:

«[...] as the web exploded, the volumes of data available for so-called shallow or statistical methods made deductive or rule-based systems seem less useful, and clunky»⁹⁶

La caratteristica tipica dell'inferenza induttiva è l'uso della generalizzazione costruita su grandi moli di dati empirico-osservativi. L'impiego di metodi induttivi è infinitamente più vantaggioso in termini di produzione di nuova conoscenza – difatti è ubiquo nelle scienze – e tuttavia le conclusioni che consente di trarre dai dati di partenza hanno valore probabilistico. Esiste sempre cioè, in termini di principio, la possibilità che un'ipotesi ottenuta induttivamente, o una teoria costruita su un insieme vasto di dati osservativi, venga falsificata da un'evidenza contraria, perché le osservazioni sono costitutivamente incomplete.

Anche questo è un fatto evidente nelle scienze, il cui avanzamento procede precisamente secondo queste modalità: l'avvicinarsi di ipotesi e modelli teorici è manifesto nelle scienze cognitive in modo particolare, ove, a causa della complessità dei fenomeni presi in esame, è più facile imbattersi in risultati osservativi che rompono le regolarità precedentemente conosciute e costringono a costanti e ulteriori lavorazioni sulla teoria. Persino un semplice

⁹⁶ *Ibid.*, p. 113.

mutamento di paradigma sperimentale, di frequente, basta a spezzare l'efficacia descrittiva di un modello teorico.

Ancora, pure la collezione di evidenze non si configura come atto neutrale, essendo invece sempre (e variamente) intrisa di teoria.

Per tutte queste ragioni, i procedimenti logico-induttivi stanno - rispetto a quelli di pura deduzione - in una inferiore gerarchia di formalizzabilità.

Sono, i secondi, espressione eminente del significato di sistema formale, ossia di apparato che unisce, come si spiegava nel capitolo precedente, una grammatica (un alfabeto simbolico, una sintassi, delle formule ben formate) e un insieme di regole di calcolo.

Nel solo contesto dei sistemi formali è ammissibile l'uso della nozione di dimostrazione: essa non è che una successione finita di formule (di nuovo, effettiva nel lessico di Turing), ricavata dalla ripetuta applicazione delle regole di calcolo agli assiomi fondamentali del sistema.

Una simile struttura, gerarchicamente incentrata sul rapporto tra assiomi, formule ben formate e procedure di calcolo, è ciò che garantisce a *Logic Theorist* e a tutti gli altri prodotti della programmazione simbolica la loro efficacia operativa.

È evidente allora che l'assenza di un identico rigore nei contesti ove l'inferenza induttiva si rende necessaria non ci consente di intenderli come egualmente formalizzabili. Possiamo parlare, in questo caso, di sistemi *semi-formali*.

Le cose si complicano ancora se si passa a considerare il problema dell'abduzione. Un'inferenza di tipo abduttivo si dà nella forma:

Premessa 1	B
Premessa 2	$A \rightarrow B$ (forse)
Conclusione	A (forse)

Ovvero: si osserva un fatto (B) e si cerca una ipotesi plausibile (A) che, se vera, potrebbe spiegare B. Per esempio:

Premessa 1	Il ladro ha sporcato il pavimento di fango
Premessa 2	Chiunque entri dal giardino, sporca il pavimento di fango
Conclusione	Il ladro è entrato dal giardino (forse)

Qui ogni schema rigoristico salta⁹⁷, sia esso formale o semi-formale, e si approda nel campo delle congetture fondate sull'esperienza, della selezione delle spiegazioni plausibili, della ricerca di analogie e via dicendo.

Affinché qualcosa di simile si realizzi, l'agente che produce l'inferenza deve tuttavia essere dotato di un sistema vastissimo di conoscenze semantiche sul mondo, e in più deve poterle inserire in una rete che renda conto delle connessioni che le legano.

Larson tratta il problema esplicitamente, usando come evidenza il fallimento di alcuni grossi progetti di ricerca cui egli stesso ha preso parte. Emblematico, ad esempio, è uno studio del Dipartimento della Difesa statunitense, che ha immesso grandi risorse finanziarie nel progetto di costruzione di "*commonsense knowledge bases*"⁹⁸ algoritmiche. Obiettivo dell'operazione era fornire tali sistemi di calcolo di una comprensione generale del mondo, di un "buonsenso" naturale simile a quello utilizzato dagli umani di ogni epoca e luogo per orientarsi nelle loro decisioni. La macchina avrebbe dovuto imparare, per esempio, che "gli umani vivi hanno la testa, [...] l'acqua rende le cose bagnate"⁹⁹, che i coccodrilli non possono partecipare alle corse ad ostacoli e via dicendo.

Fornite di tale cognizione dei fatti normali dell'esperienza – questa era la speranza – le macchine avrebbero potuto operare le stesse congetture ed ipotesi, mettere in atto gli stessi ragionamenti abduttivi che gli umani costantemente producono. L'ottimismo degli ideatori del progetto, però dovette scontrarsi con diversi ordini di problemi; il primo:

⁹⁷ Si intende che viene meno la possibilità di rendere conto di un argomento elaborato per abduzione nei termini di un sistema formale in senso tecnico.

Per quello che si è detto in (II), invece, operando agli adeguati livelli di astrazione è possibile descrivere formalmente (in senso non tecnico) qualsiasi argomento. Senza approdare a sistemi formali propriamente detti, si sono tentati molti approcci alla formalizzazione dei ragionamenti abduttivi; cito, a titolo di esempio, logica epistemica e ALP (Abductive Logic Programming), che tuttavia sono ancorate a paradigmi deduttivi o induttivi.

⁹⁸ E. J. LARSON, op. cit., p. 175.

⁹⁹ *Ivi.*

«[...] most of what we know is implicit. We bring our knowledge into consciousness, making it explicit, only when circumstances require it, like when we are surprised or have to think through something deliberately»¹⁰⁰

E ancora, strettamente connesso al precedente, il secondo:

«[...] the “tip of the iceberg.” All our implicit knowledge might be necessary for some inference or other, but the total amount is vast. The knowledge base of an ordinary person is unbelievably large, and inputting and representing it in a computer is a gargantuan task»¹⁰¹

In altri termini, neppure il più banale dei compiti di ragionamento sull'esperienza quotidiana può essere risolto senza che siano codificate moli enormi di materiale conoscitivo apparentemente insignificante. L'idea di poter afferrare complessivamente la nozione di intelligenza pur aggirando il nucleo forte del problema dell'abduzione si dimostra semplicemente disperata: non si può codificare integralmente tutta l'informazione che si possiede e si manipola, che anzi è per lo più nascosta.

Essa, peraltro, deve essere opportunamente rappresentata, ovvero organizzata in una rete gerarchica che tenga conto del rapporto intercorrente tra le sue parti.

Anche questo è problematico:

«The very attempt to organize knowledge so it's computationally accessible for real-time reasoning inevitably leaves out items that will be required given some scenario or other. There doesn't appear to be any recipe for constructing a knowledge base for intelligence»¹⁰²

Perché condizioni simili possano darsi, lo vedremo tra poco, è necessario un salto oltre ogni spiegazione formale o parziale dell'intelligenza – ridotta e idealizzante – per approdare finalmente alla sua comprensione complessiva, multilivellare, non compressa.

¹⁰⁰ *Ivi.*

¹⁰¹ *Ibid.*, pp. 175-176.

¹⁰² *Ibid.*, pp. 178.

La riduzione dell'intelligenza al razionale è, dicevamo, l'*escamotage* che ne autorizza un'interpretazione nel senso della formalizzabilità. Questo ci pare essere uno dei pilastri delle filosofie computazionaliste, che tuttavia peccano di semplicismo pure nella considerazione dello statuto stesso del razionale, il quale non apre sempre e automaticamente la strada della formalizzazione. Dall'intelligenza si passa alla ragione, dalla ragione ai soli meccanismi deduttivi: solo così se ne può sostenere la piena formalizzabilità, che è dunque ottenuta mediante progressivi restringimenti di campo, mediante considerazione sempre più esclusiva di casi particolari, infine indebitamente estesa alla nozione di partenza.

Al termine del processo un nuovo paradigma – quello computazionale – emerge. Esso pretende di includere (ovvero di definire) anche le parti di nozione che al momento della sua costruzione erano rimaste escluse dall'analisi. Il difetto dell'argomento è autoevidente.

3.2.3 Sui limiti matematici dei sistemi formali: Gödel e Lucas

Difficoltà aggiuntive provengono, nella pretesa di estendere le descrizioni formalistico-riduttive all'intero insieme dei meccanismi cognitivi, dalle fondamentali limitazioni evidenziate dalle scoperte di Gödel, in seguito utilizzate come premesse di un forte argomento anti-computazionalistico.

In modo particolare, il filosofo e logico britannico John Lucas, all'inizio degli anni Sessanta, ha sondato le implicazioni generali dei fatti dimostrati da Gödel, rapportandole all'allora fortissimo paradigma computazionale.

Il tema è stato già introdotto nel corso del primo capitolo: la prospettiva di una completa formalizzabilità della matematica appariva, sino agli anni Trenta del Novecento, ancora perfettamente percorribile. Tuttavia, la pubblicazione dell'articolo gödeliano *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*¹⁰³ ha segnato un repentino decadimento di

¹⁰³ K. GÖDEL., *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatshefte für Mathematik und Physik, vol. 38, 1931, pp. 173–198. Ristampato con traduzione inglese a fronte in Feferman, *Kurt Gödel Collected Works vol. 1: Publications 1929-1936*, Oxford University Press, New York, 1986, pp. 144–195.

tali ottimismo formalistici, avendo dimostrato che in ogni sistema formale consistente (la consistenza del sistema indica che in esso non possono provarsi proposizioni false) e sufficientemente potente esiste sempre almeno una proposizione *vera* che non si può provare dentro il sistema. Tale proposizione è l'enunciato "gödeliano".

La cesura fra ciò che è *vero* e ciò che si può provare mediante manipolazione meccanica di stringhe simboliche suggerisce, secondo Lucas, l'insostenibilità del paradigma computazionale come modello generale dell'intelligenza.

Siccome una macchina o un algoritmo di IA, se costruiti secondo schemi simbolici (fondati, cioè, sull'inferenza di carattere deduttivo) sono diretta incarnazione di un sistema formale, allora anch'essi saranno gravati dalle medesime limitazioni – individuate da Gödel – che interessano ogni sistema formale.

Considerando, invece, che *noi* possiamo cogliere la verità dell'enunciato gödeliano pur vedendone la non-provabilità, si dovrà riconoscere l'impossibilità di pensare l'insieme della cognizione umana come meccanismo formale:

«But it is enough to show that the machine is *not the same* as a mind. True, the machine can do many things that a human mind cannot do: but if there is of necessity something that the machine cannot do, though the mind can, then, however trivial the matter is, we cannot equate the two, and cannot hope to have a mechanical model that will adequately represent the mind»¹⁰⁴

L'argomento gödeliano di Lucas pare conforme alla nostra ipotesi sul restringimento della nozione di intelligenza – e anzi la spinge più in avanti: è la sua considerazione ridotta che consente l'elaborazione di un modello computazionale del pensiero, e tuttavia, secondo Lucas, pure limitando l'intelligenza al contesto della pura deduzione, l'analogia computazionale, l'opzione meccanicistica, rimangono insostenibili.

¹⁰⁴J. R. LUCAS, *Minds, Machines and Gödel*, 1959, rist. in *Philosophy*, vol. 36, n. 137, Cambridge University Press, Cambridge, 1961, p. 113.

Ad ogni modo, come noi, anche l'autore di *Mind, Machines and Gödel* riconosce come indebito il restringimento dell'intelligenza alla sola forma logica della deduzione – sebbene nel contesto delle possibili obiezioni al suo argomento centrale:

«Human beings are not confined to making deductive inferences [...] a fair model of the mind would have to allow for the possibility of making non-deductive inferences»¹⁰⁵

L'argomento proposto da Lucas ha inaugurato un dibattito di lungo corso, tutt'ora aperto. I due poli della discussione riguardano, in buona sostanza, la possibilità che esso abbia – o meno – veramente dimostrato l'estendibilità dei teoremi gödeliani al problema dell'analogia computazionale.

Avversato sin dall'inizio a molti livelli e poi rielaborato dall'autore, con sguardo retrospettivo, negli anni Novanta (LUCAS J. R., *Minds, Machines and Gödel: a retrospect*, in Millican & Clark, *Etica e Politica*, Clarendon Press, Oxford, 1996), l'articolo è comunque divenuto un riferimento classico nel campo dell'interazione tra IA, scienze cognitive (e relativa epistemologia) e filosofia della mente.

Al di là delle posizioni in gioco nel dibattito, e anche oltre la sua correttezza in generale, l'argomento di Lucas ci riguarda da vicino perché incrocia l'opinione a favore di cui ragioniamo: la sostenibilità del paradigma computazionale – il quale emerge in seguito ad una riduzione semantica della nozione di intelligenza – non può essere valutata in termini assoluti, ma solo conformemente al livello di astrazione a cui decidiamo di operare.

Per comprendere la portata del problema, possiamo valutare brevemente le conseguenze dell'analogia computazionale nella produzione di teoria, per sondare i possibili esiti di un uso “modellistico” dell'IA nelle scienze cognitive e, infine, individuare i contesti di corretta applicazione dei due differenti riduzionismi rilevati nella nostra analisi.

¹⁰⁵ *Ibid.*, p. 119.

3.3 Effetti del paradigma formalistico-riduzionistico nelle scienze cognitive

Dall'inizio della sua storia, e persino durante la sua "preistoria", lo sviluppo dell'IA come disciplina autonoma si è mostrato connesso ad un interesse generale per la comprensione del pensiero umano. Abbiamo fatto coincidere la preistoria dell'idea computazionale con l'avvento della matematica di Hilbert, e la sua storia vera e propria con le ricerche di Alan Turing e – in seguito – dei grandi pionieri di Chicago, Darmouth, Princeton, Stanford: Newell, Simon, McCarthy, Minsky, Von Neumann e via dicendo, ma anche, ovviamente, i fondamentali Rashevsky, Pitts, McCulloch, che abbiamo accertato essere gli ispiratori dell'uso degli approcci matematici e dunque formalizzanti nelle scienze biologiche.

La nostra posizione, però, è più che conservativa: se passassimo a considerare senza troppi filtri rigoristici la volontà umana di replicare, per mezzo della tecnica, la vita in generale, scopriremmo una storia ben più lunga e articolata e vedremmo in essa una sconcertante onnipresenza dell'idea computazionale – più spesso meccanicistica – del pensiero e delle funzioni biologiche (quella computazionale essendo una nozione più ristretta del meccanicismo, e specificata intorno al problema del calcolo automatico).

Martin Davis, lo sappiamo, la fa risalire alla *characteristica universalis* di Leibniz; Daniel Crevier, nella sua ormai classica storia dell'Intelligenza Artificiale propone una rassegna di esperienze miranti alla replicazione artificiale della vita che parte dall'Egitto del IX secolo a.C. e attraversa tutti i secoli, per giungere allo scorso e poi ancora in avanti verso quello presente.

L'interesse per la comprensione profonda del pensiero, del linguaggio, dell'intelligenza, è andato sempre di pari passo con la volontà di ricrearli artificialmente. Così è, senza dubbio, anche per i grandi avanzamenti nel campo dell'IA a partire dallo scorso secolo: essa solleva questioni radicali sullo statuto dell'intelligenza umana, sul linguaggio e sulle sue sorgenti, sulle modalità di codifica ed elaborazione delle informazioni nella mente, sulla natura della memoria e della percezione, sul rapporto tra corporeo e mentale e via dicendo.

Per tutte queste ragioni non deve sorprendere, ad esempio, la perfetta coincidenza delle date di nascita dell'IA a Darmouth e delle Scienze Cognitive al MIT (entrambe ufficialmente sorte nel 1956), definendo le seconde un campo in cui una vasta congerie di discipline particolari (la filosofia, la psicologia, le neuroscienze, la linguistica, l'antropologia e – ovviamente – l'intelligenza artificiale) lavorano, cooperando oppure scontrandosi, al tentativo di comprendere i molti misteri della cognizione umana.

L'IA, in altri termini, ha costituito sin da subito uno dei propulsori principali dello sviluppo teorico nelle scienze cognitive, svolgendo in esse il ruolo di modello; il coevo affiorare dell'analogia mente-macchina, resa molto più robusta dai successi dell'IA stessa, ha prodotto un definitivo riverbero del paradigma computazionale sullo studio dei fenomeni della cognizione. Esso ne ha, pertanto, orientato la teoria in molti modi, garantendole molti benefici e altrettante distorsioni.

3.3.1 *Pars construens*

All'inizio degli anni Quaranta dello scorso secolo, con un particolare tempismo rispetto alle grosse acquisizioni teoriche di Turing, Rashevsky, McCulloch e Pitts, lo psicologo britannico Kenneth Craik pubblicò la sua opera più nota, che decise di intitolare *The Nature of Explanation*¹⁰⁶. In essa, accanto a più generali discussioni di natura epistemologica, proponeva un'idea che abbiamo già incontrato: comprendere qualcosa significa poterne produrre un modello funzionante. La “natura della spiegazione” consiste nell'equazione tra la comprensione di un fenomeno e la sua replicabilità artificiale – o quantomeno la sua descrivibilità in termini di modello.

Craik operò pionieristicamente secondo tale prospettiva, costruendo i primissimi esperimenti psicologici che prevedevano lo sviluppo del relativo modello teorico attraverso l'uso di calcolatori analogici.

Il suo approccio, decisamente guidato dall'analogia computazionale, fu raccolto con profitto da molti degli studiosi che vennero dopo di lui. Spicca, tra questi, il lavoro di Ulric Neisser, il cui testo fondamentale, *Cognitive Psychology*, valse a

¹⁰⁶ K. J. W. CRAIK, *The Nature of Explanation*, University Press, Macmillan, 1943.

fondare un modo inedito di intendere la scienza psicologica. La neonata psicologia cognitiva ha mosso i suoi primi passi e ottenuto i suoi molti risultati precisamente nello spazio aperto dal paradigma computazionale, dall'analogia tra sistema meccanico-formale e fenomeno cognitivo.

Sin dall'inizio della loro storia, dunque, le scienze cognitive hanno interpretato l'intelligenza, alla luce dell'analogia, come un sistema atto alla manipolazione di regole, come un insieme di procedure di calcolo, di codifica, di immagazzinamento, di recupero.

Aver operato secondo tale paradigma ha prodotto vantaggi innumerevoli; possiamo vederne qualcuno nelle righe seguenti.

In primo luogo, i processi di riduzione e formalizzazione che rendono operativa, e fertile di teoria, l'analogia mente-macchina, hanno reso possibile lo sviluppo di *modelli computazionali* dei fenomeni cognitivi.

Trattasi di descrizioni formali, spesso in senso non tecnico (cfr. par. 3.2.2, nota 97), il cui obiettivo centrale è la rappresentazione o la simulazione di alcuni dei processi della cognizione umana mediante sistemi di regole simboliche, algoritmi e reti di dati.

Tali descrizioni possono operare a livelli tra loro estremamente differenti e rappresentare l'andamento di interi processi, come nel caso dei modelli di Treisman per l'attenzione selettiva¹⁰⁷ o di Baddeley ed Hitch per la memoria di lavoro¹⁰⁸, oppure emulare reti di connessioni neurali che - sulla scorta dei risultati di McCulloch e Pitts - svolgono compiti di apprendimento automatico, visione artificiale, riconoscimento di pattern, elaborazione linguistica.

Ancora, elaborare modelli computazionali dei fenomeni cognitivi consente di testare l'affidabilità delle teorie, di produrre previsioni e confrontarle con i dati empirici, di simulare condizioni variate di ogni genere e valutare la risposta del sistema.

¹⁰⁷ A. M. TREISMAN, *A feature-integration theory of attention*, Cognitive Psychology, Vol. 12, Issue 1, 1980, pp. 97-136.

¹⁰⁸ A.D. BADDELEY, G. HITCH, *Working memory*, in G.H. BOWER, *The psychology of learning and motivation: Advances in research and theory*, Vol. 8, pp. 47-89, New York, Academic Press, 1974.

Eloquente esempio di questa possibilità è, tra le altre, la cosiddetta. “Teoria Sensoriale-Funzionale” di Farah e McClelland (1991) elaborata per rendere conto dell’asimmetria, sperimentalmente rilevabile, nei deficit di riconoscimento per categoria (i soggetti che ne soffrono hanno prestazioni peggiori nel riconoscimento degli oggetti animati e migliori nel riconoscimento di quelli inanimati).

Per testare le ipotesi centrali della teoria, gli autori hanno elaborato un modello computazionale basato su di esse; dopodiché sono intervenuti sulla sua struttura, disattivando alternativamente alcune delle unità visive-semantiche o alcune delle unità funzionali del modello (sul preciso significato delle due nozioni ci concediamo di sorvolare).

Hanno potuto così determinare che il danneggiamento delle capacità visive peggiorava le prestazioni di riconoscimento degli oggetti animati molto più di quelle relative agli oggetti inanimati – e quindi valutare la conformità delle ipotesi esplicative ai dati provenienti dall’osservazione. Per testare le loro ipotesi, i due studiosi non hanno dovuto danneggiare le strutture cerebrali di qualche sfortunato soggetto sperimentale: è bastato loro attivare o disattivare parti di codice di un programma per calcolatori. Il vantaggio non è secondario.

In secondo luogo, il successo dell’IA, la quale offre massicci strumenti per l’analisi e l’elaborazione statistica dei dati sperimentali, ha introdotto nelle scienze cognitive un approccio che, nel lessico più aggiornato, possiamo chiamare *data-driven*.

L’importanza dell’analisi sistematica dei dati sperimentali emerge soprattutto al livello dell’indagine neuro-cognitiva, interessata primariamente all’individuazione del sostrato neurale dei fenomeni che caratterizzano l’intelligenza umana. Qui la ricerca di regolarità statistiche, di mappe, schemi e tendenze entro grosse collezioni di dati ha garantito una precisione d’analisi che non conosce precedenti.

In casi innumerevoli l’enfasi sull’elaborazione dei dati ha consentito fondamentali riarrangiamenti teorici.

Ad evidenza di questo fatto possiamo utilizzare ancora la teoria, da poco citata, di Farah e McClelland: le evidenze provenienti dall’analisi di grandi moli

di dati raccolti al livello dei fenomeni neurali ha consentito di superarne le difficoltà fondamentali. La teoria è troppo “rigida”, in modo particolare, sulla categorizzazione delle caratteristiche che organizzano l’immagazzinamento nella memoria delle conoscenze di tipo semantico, presupponendo che ogni informazione semantica possa essere categorizzata secondo le sue proprietà sensoriali (colore, forma, grandezza, suono eccetera) oppure funzionali (l’uso che se ne fa).

Secondo Cree e McRae¹⁰⁹, invece, due sole proprietà non sono sufficienti; hanno proposto infatti un approccio “a caratteristiche multiple”.

L’analisi delle zone di attivazione, nel cervello, al momento della codifica di informazioni semantiche, ha infine dato ragione al loro approccio, individuando una collocazione diffusa proprio secondo le molteplici caratteristiche del materiale informativo: l’informazione relativa al colore attiva un’area, quella relativa alla forma un’altra e così via.

La possibilità di affinare la teoria è dipesa, in questo caso, dalla grande capacità d’analisi dei dati aperta dall’uso della computazione.

In generale, nonostante i loro limiti manifesti, le procedure di riduzione e formalizzazione che sono all’opera dentro il paradigma computazionale hanno concesso alle scienze cognitive una presa su fenomeni altrimenti estremamente elusivi: consentono, diminuendone la complessità generale, astraendo e semplificando, di formulare teorie, ipotesi e spiegazioni.

3.3.2 *Pars destruens*

Alle grandi opportunità che il paradigma computazionale offre alle scienze cognitive si deve affiancare l’insieme dei difetti che genera.

Abbiamo già valutato le criticità generali alle quali, potenzialmente, ogni descrizione riduzionistica e formalistica dei fenomeni complessi apre la strada. È stata messa a fuoco, allo stesso modo, la natura problematica dell’operazione con cui si fa sorgere l’analogia computazionale, che sembra consistere in un

¹⁰⁹ G. S. CREE, K. MCCRAE, *Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns)*, *Journal of Experimental Psychology: General*, 132(2), 2003, pp. 163–201.

restringimento di campo e in una sua successiva, repentina ed indebita ri-estensione.

Si devono adesso ponderare le conseguenze precipue di questi fatti, di nuovo riferendole al problema della produzione di teoria nella scienza della cognizione. Si può cominciare dall'ultima delle questioni presentate nel paragrafo precedente. Se un approccio *data-driven* alla costruzione di teoria le garantisce i benefici che abbiamo conosciuto, è però anche suscettibile di generare distorsioni rilevanti. Di esse discute Larson al paragrafo emblematicamente intitolato *Dead-end Research*:

«We have seen that Big Data AI is not well-suited for theory emergence. On the contrary, without existing theories, Big Data AI falls victim to overfitting, saturation, and blindness from data-inductive methods generally»¹¹⁰

Una metodologia orientata alla sola raccolta e analisi di dati ma povera di audacia teorica e di ipotesi creative non è il modo più efficace per produrre nuova conoscenza.

Analisi più dettagliate, svolte su moli sempre maggiori di dati neurali sempre più puntuali, più circoscritti e localizzati, non producono necessariamente una migliore comprensione delle architetture cognitive e dei fenomeni che le sorreggono.

Nel caso della teoria presentata precedentemente, ad essere feconda sul piano della scienza è stata la *combinazione* dell'approccio orientato ai dati con lo spessore concettuale e teorico di un'ipotesi.

La capacità di produrre congetture ed ipotesi guadagna dalle grandi possibilità di analisi dei dati aperte dall'IA un prezioso ancoraggio al reale; la sola "forza bruta" dei dati sperimentali, senza la profondità di un'idea rimane arida di teoria: essa non sgorga dai soli dati, ma dipende, per così dire, da un atto creativo.

Il pericolo ultimo è una svalutazione del contributo intellettuale, nelle scienze, dello scienziato in carne ed ossa, che può correre il rischio di diventare – con

¹¹⁰ E. J. LARSON, op. cit., p. 267.

una metafora – l’operaio che gira la manovella di un sistema computazionale. Il secondo – e non il primo – sarebbe in questo caso il motore di ogni avanzamento scientifico. In realtà, lo abbiamo visto, le cose non stanno in questi termini.

Un secondo ordine di problemi riguarda la cosiddetta “validità ecologica” dei modelli computazionali dei fenomeni cognitivi. Essa valuta il quale misura il contesto della sperimentazione possa essere rappresentativo della realtà “naturale” del fenomeno. Il problema, che si dà comunque anche nella sperimentazione psicologico-cognitiva classica, è di molto accentuato quando ad essere testati sono modelli computazionali invece che soggetti in carne ed ossa, ambienti virtuali e simulati, semplici e controllabili, al posto della grandiosa complessità di quelli reali.

In questo caso, come in ogni altra circostanza in cui – ai fini della comprensione – siamo costretti a ridurre, semplificare, idealizzare, dobbiamo porci il problema della generalizzabilità delle nostre conclusioni, per evitare di incorrere nell’errore classico che abbiamo individuato all’inizio del capitolo (cfr. par. 3.1).

Veniamo infine al problema cruciale: quando il paradigma computazionale trascende i suoi margini “operazionali”, così li abbiamo chiamati – quando cioè smette di essere interpretato secondo le sue possibilità e secondo i suoi limiti effettivi, esso rischia di produrre una infruttuosa riduzione della complessità dei processi cognitivi, che finisce per intenderli come *costitutivamente* algoritmici, semplici e computazionali, quando invece la loro natura algoritmica e computazionale è il risultato di una precisa opzione filosofica, in modo particolare quella con cui si sceglie di *isolare* – entro un problema generale ad elevato quoziente di complessità – alcune delle sue parti, e dunque di renderne operativa la lettura ad alcuni determinati livelli di astrazione.

Questa è, di nuovo, un’opzione metodologica, una scelta filosofica precisa e non una semplice ed inderogabile necessità dettata dall’evidenza.

Nell’analisi dei fenomeni cognitivi, essendo essi caratterizzati da una complessità straordinaria, si deve necessariamente operare a livelli di astrazione diversificati.

Un fatto qualsiasi che li riguardi – un comportamento, per esempio – può essere spiegato al livello delle interazioni sociali, a quello dello stato psicologico individuale, a quello dei singoli sottoprocessi della cognizione (percezione, memoria, linguaggio, attenzione eccetera), o ancora al livello delle attivazioni neurali. Ciascuna di queste spiegazioni è istruttiva, ma lo è – per l'appunto – al suo proprio livello di astrazione. Per ciò che abbiamo rilevato al termine del secondo capitolo, nessuna di queste spiegazioni è quella “corretta” in termini assoluti; lo è semmai alla sua propria scala, e l'atto con cui determiniamo la scala a cui operare è interamente filosofico.

Interpretare il paradigma computazionale rigidamente, al contrario, ci impedisce di vedere la multilivellarità del fenomeno dell'intelligenza.

Seguendone gli esiti riduzionistici sino alle conseguenze più remote ci imatteremo in assurdit  autoevidenti, certamente inaccettabili anche dentro il paradigma computazionale cos  inteso.

Dovremmo concludere, ad esempio, che siccome le relazioni collettive che danno luogo ai fenomeni storici sono il risultato dell'interazione di individui, che questi a loro volta sono determinati dai meccanismi della cognizione, che ancora dipendono da quelli cerebrali e poi neurali, e cos  via sino alla scala dei fenomeni subatomici, che una spiegazione quantistica dei fatti socio-economici sia pi  attendibile di una di carattere sociologico.   evidente, qui, che il difetto risiede nella confusione tra i piani della spiegazione.

In modo simile, l'uso indebito del paradigma computazionale conduce ad una visione integralmente meccanicistico-deterministica della cognizione, che comprime alcuni degli elementi centrali ma meno riducibili del fenomeno (esperienza soggettiva, coscienza, emozione, ambiente/contesto/cultura) e li esclude dal contesto della trattabilit  scientifica.

3.4 Riduzionismo e anti-riduzionismo in ottica operativa

Abbiamo aperto le discussioni del presente capitolo dichiarando un preciso proposito: stabilire l'esistenza di riduzionismi di segno diverso, comprendere quale di essi possa venire utilmente inteso come metodo scientificamente fecondo. Si   rilevata la necessit , ai fini di una distinzione tra “correttezza” e

“scorrettezza” dei paradigmi riduzionistici, di specificare il problema in relazione al contesto (o meglio al livello) della riduzione stessa.

Per tutto ciò che abbiamo visto sulla struttura profonda dei meccanismi di riduzione e sull’uso dei sistemi formali nella descrizione dei fenomeni complessi – e anche per la mole considerevole delle esperienze scientifiche che abbiamo preso in considerazione, dalla matematica di Hilbert e di Turing alla biofisica formale di Rashevsky, dall’uso della nozione di calcolo logico in McCulloch e Pitts agli impieghi che di essa hanno fatto i padri dell’IA – possiamo individuare il discrimine tra riduzionismi efficaci e riduzionismi, per così dire, velleitari, precisamente nella nozione di *operatività*.

In un’indagine epistemologica sullo statuto del paradigma computazionale come modello teorico dei sistemi cognitivi, l’opzione riduzionismo/anti-riduzionismo pare troppo astratta e generale.

Abbiamo verificato che, in misura pressoché identica, l’uso di procedure di riduzione garantisce avanzamenti e produce distorsioni potenziali.

La necessità di adottare opzioni riduzionistiche è connaturata ad ogni descrizione scientifica del mondo: l’abbiamo vista all’origine della rivoluzione inaugurata da Turing, il cui modello matematico di macchina deve la vita ad un’analisi riduzionistica della nozione di intelligenza; è emersa, pur con ogni sua limitazione, nella straordinaria eleganza formale della biofisica matematica di Rashevsky e nella costruzione del primo modello di neurone formale da parte di Pitts e McCulloch; compare nelle procedure di descrizione dei fenomeni cognitivi, a moltissimi livelli di astrazione.

Per queste ragioni, difendere una posizione anti-riduzionistica *tout-court* avrebbe il valore di un vero e proprio auto-sabotaggio.

Allo stesso modo, si deve riconoscere come inconcludente ogni posizione riduzionistica “forte”, per la quale, cioè, la sola fonte di validità scientifica di una conclusione è il suo essere stata formulata al minimo grado di astrazione possibile (come, più in alto, nel caso della “sociologia subatomica”).

La competizione tra riduzionismi di segno opposto si risolve ponendoli sempre in ottica operativa: per valutare la loro adeguatezza, bisognerà primariamente guardare ai loro *scopi*.

Declinato in prospettiva operativa, ogni approccio riduzionistico mostra la sua legittimità: sul versante tecnico o ingegneristico dell'IA, valutare l'ampiezza complessiva dell'uso della nozione di intelligenza non è certamente rilevante, considerato che lo scopo è la costruzione – operativa, per l'appunto – di un artefatto funzionale ad uno scopo.

Qualora l'intento fosse, invece, di fornire un modello computazionale della cognizione, allora il grado di complessità a cui si opera sarebbe tutt'altro che irrilevante, e il livello di astrazione a cui si colloca la teoria che ne deriva dovrebbe essere isolato con la cura adeguata.

Un simile criterio di discernimento della legittimità della riduzione ci pare riflettere in modo pedissequo la distinzione che abbiamo individuato – nel contesto della filosofia turinghiana – tra riduzionismo debole o operativo e riduzionismo generale o forte.

Allora, come dicevamo, scegliere l'opzione riduzionistica adeguata diventa il compito cruciale.

L'approccio riduzionistico legittimo fornisce automaticamente i limiti applicativi delle descrizioni formali. Esse sono valide quando non determinano una compressione di principio della complessità del fenomeno su cui agiscono: una descrizione formale che tenga conto del grado di astrazione a cui opera e che non intenda generalizzarsi a tutti gli altri. Ecco tracciato il contesto della legittimità della formalizzazione.

La riduzione operata in senso forte, al contrario, smette di essere legittima per il semplice fatto di confondere i piani della spiegazione – peraltro in colpevole disosservanza del principio di non-eccezionalità delle descrizioni formali che abbiamo individuato più in alto.

Un esempio ulteriore ci dica, su questo, l'ultima parola: i difetti, in termini di semplificazione e idealizzazione – addirittura di “fantasia” – che affliggono la teoria di Pitts e McCulloch non determinano il benché minimo decadimento della legittimità del suo uso operativo. Esso è stato condotto con il massimo profitto da Von Neumann, il quale ha usato il calcolo neuronale dei due colleghi per lo sviluppo delle sue tecniche di *design* logico dei programmi per calcolatore.

Al contrario, se è in gioco una pretesa diretta di comparazione tra il modello computazionale di neurone e il neurone biologico, tra il calcolo logico proposizionale e il pensiero come fenomeno di alto livello, la riduzione del secondo al primo è problematica – in ultima istanza illegittima per tutte le ragioni che abbiamo elencato.

Riteniamo di aver acquisito evidenze in numero sufficiente; possiamo a questo punto trarre le conclusioni necessarie.

CONCLUSIONI

Il presente lavoro di tesi ha preso le mosse da una domanda sullo statuto dell'intelligenza umana e, in modo particolare, sulla sua riproducibilità tecnica. Ci si chiede, cioè, se essa possa essere catturata entro un qualche genere di sistema formale o semi-formale.

Il problema è stato affrontato massicciamente negli ultimi cinque decenni, soprattutto in corrispondenza delle cicliche fasi apicali dell'interesse intorno all'Intelligenza Artificiale. Essa è stata uno dei motori fondamentali dei processi che hanno condotto all'emersione di una analogia mente-calcolatore o anche, come l'abbiamo chiamato, di un paradigma computazionale della cognizione.

Siccome la tradizione scientifica che lo riguarda è robusta e ben articolata, abbiamo affrontato il problema in una prospettiva particolare, quella degli approcci riduzionistici e dell'uso dei sistemi formali nella descrizione dei fenomeni complessi.

Facendo luce sulle connessioni che legano i programmi di ricerca sui fondamenti della matematica con la nascita della teoria della computabilità effettiva si è potuto rintracciare il percorso che l'analogia computazionale ha seguito prima di affiorare in maniera definitiva.

Nelle ricerche di Turing, dalle quali siamo partiti, tale percorso si è mostrato prima attraverso una timida analisi riduzionistica della nozione di calcolo e più tardi nella forma esplicita dell'analogia: Turing ha pensato, certo non per primo, ma con strumenti massimamente solidi, che la riproduzione artificiale dell'intelligenza umana fosse impresa possibile e scientificamente necessaria.

Complicando l'indagine sul lavoro del matematico inglese e adottando una prospettiva di "filologia" del pensiero, abbiamo potuto individuare nella filosofia di Turing una biforcazione teorica rispetto al problema del riduzionismo.

Esso ci è apparso declinato in forma forte oppure operativa: nel primo caso si postula un riduzionismo di genere classico, seguendo il quale si argomenta a favore di una descrizione del fenomeno che lo "schiaccia" sui livelli inferiori i quali, per loro natura costitutiva, sono passibili di un migliore trattamento

formale; nel secondo caso la riduzione è convenzionale: essa descrive – sì – riducendo, astraendo, semplificando, ma lo fa con fini operativi. Un genere simile di riduzione è in gioco, per esempio, nella costruzione di sistemi che simulano lo svolgimento di un'operazione intellettuale tipicamente umana; può anche avere un ruolo nella produzione di teoria, purché se ne rilevino i limiti precisi.

Nella seconda sezione ci siamo mossi più convintamente verso alcune delle altre sorgenti fondamentali del computazionalismo.

Mediante l'analisi di alcuni dei risultati fondamentali di Rashevsky, il quale ha traslato, nel campo delle scienze biologiche, i metodi d'indagine di profilo formale che hanno sempre interessato le scienze – cosiddette – esatte, si sono potute valutare alcune delle conseguenze tipiche dei tentativi di formalizzazione dei sistemi complessi.

Sotto il profilo teoretico, nel contesto della nostra indagine, l'importanza di Rashevsky sta nella grande luce che i suoi metodi matematici gettano sulle caratteristiche generali dei processi di formalizzazione.

Identicamente stanno le cose anche per l'opera di McCulloch e Pitts: il loro rilievo è certamente storico, avendo esso a che vedere con la prima occorrenza mai documentata dell'uso della nozione di computabilità nel contesto di un'indagine simil-biologica. Elaborando il primo modello di neurone artificiale, i due studiosi hanno innestato definitivamente la nozione turinghiana di computabilità su problematiche di tipo cognitivo.

Anche i loro metodi formali manifestano insieme grandi potenzialità tecnico-operative e ingenti distorsioni e grosse limitazioni se estesi ad un'indagine teorica di respiro complessivo.

Siamo passati infine a valutare le conseguenze del paradigma computazionale nel contesto delle scienze cognitive.

Si è visto che, per la vaghezza della nozione di intelligenza, i trattamenti scientifici che l'hanno riguardata hanno operato una visibilissima contrazione del suo significato, finendo per indicare esclusivamente ciò che in essa ha a che vedere con il razionale.

A ben guardare, però, anche il “razionale” non identifica un processo semplice ed unitario. Lo si può intendere almeno in senso tripartito: deduttivo, induttivo e abduttivo.

La distinzione è cruciale: grandi difficoltà, infatti, vengono all’IA dall’apparente sfuggevolezza delle inferenze di natura abduttiva, tanto restie ai tentativi di formalizzazione quanto incastonate al centro della considerazione tipica dell’intelligenza naturale. L’abduzione sembra essere, cioè, precisamente il meccanismo della distinzione tra cognizione naturale e cognizione artificiale. Tale argomento ha individuato un ulteriore, apparentemente insormontabile limite del formalismo.

Proseguendo in questa direzione, abbiamo dovuto rilevare – servendoci di evidenze tratte soprattutto dalla storia della psicologia cognitivista – che l’influenza del paradigma computazionale nelle scienze cognitive ne ha guidato lo sviluppo teorico insieme profittevolmente e nocivamente.

Esso è radicato all’origine stessa delle scienze cognitive, i cui risultati non possiamo fare a meno di leggere proprio alla luce dell’analogia mente-macchina. Qui l’uso di approcci semplificanti, di idealizzazioni, di scomposizioni di fenomeni complessi in sistemi e sottosistemi particolari ha dato i suoi frutti, e pertanto non si può sposare un anti-riduzionismo integrale

Combinando queste evidenze con quella delle molte distorsioni che il paradigma computazionale ha introdotto nelle scienze cognitive, abbiamo individuato l’*operatività* della descrizione quale criterio fondamentale di demarcazione tra riduzioni e formalizzazioni legittime o illegittime.

Questo risultato ci pare fondamentale, perché nel tracciare un’epistemologia delle descrizioni formali nel campo della cognizione il bilanciamento tra operazioni vantaggiose e operazioni svantaggiose per la buona tenuta descrittiva della teoria non è compito marginale – essendo anzi una sorta di “imperativo” epistemologico.

Ne concludiamo – conformemente ai nostri intenti di partenza – che i problemi causati da un uso improprio dell’analogia dipendono dall’incapacità di intendere le letture parziali che essa produce e di confinarle dentro i suoi limiti appropriati.

In essi l'uso dell'analogia è perfettamente legittimo, e in più anche necessario, funzionale e fecondo.

Come abbiamo già detto, ci pare di poter stabilire che non sia l'analogia in sé a generare incongruenze teoriche, bensì il suo straripamento, la sua indebita applicazione a tutti i livelli ai quali si può leggere la cognizione.

Nella discussione di questi argomenti si deve includere una considerazione importante: sino ad ora abbiamo usato la nozione di paradigma con la massima scioltezza, quasi mostrando di ritenere che la sua adozione segua una scelta consapevole.

Essendo invece un paradigma, in questo particolare contesto, nient'altro che l'insieme dei metodi, delle procedure e delle conoscenze possedute, dunque una cornice entro cui alcune spiegazioni sono rese più plausibili e alcune domande più facilmente poste, ebbene, per questi motivi lo si deve intendere quale guida *irriflessa* alla produzione di teoria.

Qui emerge forse il valore fondamentale di una ricerca come quella che qui si è tentata: organizzata come una disamina critica del paradigma, che ne ritraccia le sorgenti, ne disvela i meccanismi, i presupposti e le implicazioni, essa può forse servire a mitigare la coazione che tutti questi fatti esercitano, a vederli all'opera e a bilanciarne le conseguenze.

Lo studio condotto nelle pagine precedenti si affianca ad una letteratura vastissima, non potendo vantare pertanto particolari ambizioni di esclusiva tematica o di originalità generale.

Comunque, l'attenzione circoscritta all'indagine sulla generalizzazione delle descrizioni formali, delimitando un'area più specifica, ci pare possa aprire uno spazio di novità, per quanto modesto.

Se ne può immaginare un'espansione ulteriore, specificata intorno ai metodi formali che sono in uso nel campo dell'IA e che tengono ancora in piedi il paradigma computazionale. Esiste infatti, anche per la velocità alla quale si muove l'intreccio tra tecniche di formalizzazione e indagine sullo statuto dell'intelligenza, una sovrabbondanza di materiale su cui esercitare analisi parallele a quelle condotte nelle pagine precedenti. Si potrebbe ambire a far emergere, anche in questo contesto, le tendenze semplificatorie che

caratterizzano i trattamenti formali e che da essi fanno derivare insieme vantaggi descrittivi e distorsioni interpretative.

Ci pare di aver infine chiarito alcune posizioni – segnatamente quelle relative ad una lettura differenziata dei fenomeni complessi sulla base del livello operativo della descrizione – che nel dibattito scientifico rimangono confuse nonostante la grande disponibilità di letteratura.

BIBLIOGRAFIA

- AA. VV., *Vocabulaire de sciences cognitives*, Paris, PUF 1998 (*Dizionario di scienze cognitive*, Roma, Editori Riuniti, 2000).
- ABRAHAM T. H., *Nicolas Rashevsky's Mathematical Biophysics*, *Journal of the History of Biology*, 37, pp. 333-385, Kluwer Academic Publishers, 2004.
- BADDELEY A.D, HITCH G., *Working memory*, in G.H. Bower, *The psychology of learning and motivation: Advances in research and theory*, Vol. 8, pp. 47–89, New York, Academic Press, 1974.
- BOOLE G., *An investigation of the Laws of Thought on which Are Founded the Mathematical Theories of Logic and Probabilities*, Walton & Maberly, London, 1854; rist. Dover, New York, 1958 (trad. it. *Indagine delle leggi del pensiero su cui sono fondate le teorie matematiche della logica e della probabilità*, Einaudi, Torino, 1976).
- ARBIB M. A., *La mente, le macchine e la matematica*, Boringhieri, Torino, 1968.
- BARRY COOPER S., VAN LEEUWEN J., *On Computable Numbers, with an Application to the Entscheidungsproblem – A Correction*, in Alan Turing, *His Work and Impact*, Elsevier, 2013, pp. 13-115.
- BODEN M. A., *Mind as machine: a history of Cognitive Science*, 2 voll., Oxford University Press, Oxford, 2006.
- BODEN M., *The Philosophy of Artificial Intelligence*, Oxford University Press, Oxford, 1990.
- BOOLOS G., JEFFREY R., *Computability and Logic*, 2nd ed., Cambridge University Press, Cambridge, 1980
- CHURCH A., *Review: On Computable Numbers, with an Application to the Entscheidungsproblem*, in *The Journal of Symbolic Logic*, Vol. 2, No. 1, 1937, pp. 42-43.
- COLBURN T., *Philosophy and Computer Science. Explorations in Philosophy*, 1999, ISBN 1-56324-991-X.
- CORDESCHI R., *The Discovery of the Artificial: Behavior, Mind, and Machines Before and Beyond Cybernetics*, Dordrecht, Springer, 2002.

- CORDESCHI R., TAMBURINI G., *Alan Turing e il programma di ricerca dell'Intelligenza Artificiale*, in Hosni H. (ed.), *Menti e macchine. Alan Mathison Turing a cento anni dalla nascita*, Edizioni della Normale, Pisa, 2015, pp. 87-126.
- CREE G. S., MCCRAE K., *Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns)*, *Journal of Experimental Psychology: General*, 132(2), 2003, pp. 163–201.
- CREVIER D., *AI: The Tumultuous Search for Artificial Intelligence*, BasicBooks, Harper Collins Publishers, New York, 1993, p. 46.
- CRAIK K. J. W., *The Nature of Explanation*, University Press, Macmillan, 1943.
- DAVIS, M., *Il calcolatore universale*, Adelphi, Milano, 2012.
- DAVIS, M., *The undecidable basic papers on undecidable propositions, unsolvable problems and computable functions*, Raven press, New York, 1965.
- DI FRANCESCO M., *Introduzione alla filosofia della mente*, Carocci Editore, Roma, 2003.
- FITCH F., *Review of McCulloch and Pitts 1943*, *Journal of Symbolic Logic*, 9(2), 49–50, 1944.
- FLORIDI L., *Philosophy and Computing: An Introduction*, 1999.
- FLORIDI L., *Philosophy of Computing and Information: 5 Questions*, 2008.
- FLORIDI L., *The Logic of Information. A Theory of Philosophy as Conceptual Design*, Oxford University Press, Oxford, 2019.
- FODOR J., *The Language of Thought*, Thomas Y. Crowell, New York, 1975.
- FRIXIONE, NUMERICO, *Alan Mathison Turing*, in *APhEx*, n. 7, gennaio 2013.
- GALVAN S., *Gödel e il modello computazionale della mente*, in *Rivista di Filosofia Neoscolastica*, XCVI, Vita e Pensiero – Pubblicazioni dell'Università Cattolica del Sacro Cuore, Milano 2004, pp. 145-174.

- GALVAN S., *Pensiero e logica*, in Rivista di Filosofia Neoscolastica, vol. 79, n. 1, Vita e Pensiero – Pubblicazioni dell'Università Cattolica del Sacro Cuore, Milano, 1987, pp. 118-128.
- GÖDEL K., *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatshefte für Mathematik und Physik, vol. 38, 1931, pp. 173– 198. Ristampato con traduzione inglese a fronte in Feferman, *Kurt Gödel Collected Works vol. 1: Publications 1929-1936*, Oxford University Press, New York, 1986, pp. 144–195.
- HILBERT D., ACKERMANN W., *Grundzüge der theoretischen Logik*, Springer, Berlin, 1928.
- HILBERT D., *Grundlagen der Geometrie*, Teubner, Leipzig, 1899 (tr. it.: *Fondamenti della geometria. Con i supplementi di Paul Bernays*, Franco Angeli Editore, Milano, 2009).
- HODGES A., *Alan Turing: The Enigma*, New York, Simon & Schuster, 1983, trad. it. *Storia di un enigma. Vita di Alan Turing 1912-1954*, Bollati Boringhieri, Torino, 1991.
- HOFSTADTER D. R., *Gödel, Escher, Bach: un'Eterna Ghirlanda Brillante*, Adelphi, Milano, 1984.
- KLEENE S., *Turing's Analysis of Computability, and Major Applications of It, The Universal Turing Machine: A Half-Century Survey*, ed. R. Herken, Kammerer and Unverzagt, Hamburg, 1988.
- LARSON E. J., *The Myth of Artificial Intelligence. Why Computers Can't Think the Way We Do*, The Belknap Press of Harvard University Press, Cambridge – London, 2021.
- LETTVIN J. L., PITTS W. H., *A Mathematical Theory of Affective Psychoses*, Bulletin of Mathematical Biophysics, 5, 139–148, 1943.
- LIETO A., *Cognitive Design for Artificial Minds*, Routledge, Taylor & Francis, London, 2021.
- LUCAS J. R., *Minds, Machines and Gödel*, 1959, rist. in Philosophy, vol. 36, n. 137, pp. 112-127, Cambridge University Press, Cambridge, 1961.
- LUCAS J. R., *Minds, Machines and Gödel: a retrospect*, in Millican & Clark, Etica e Politica, Clarendon Press, Oxford, 1996.

- *Mathematical Logic and the origin of modern computers*, in “Studies in the History of Mathematics”, Mathematical Association of America, Washington, 1987.
- MCCULLOCH W. S., PITTS W., *A Logical Calculus of the Ideas Immanent in Nervous Activity*, in “Bulletin of Mathematical Biophysics”, 5, 1943; rist. in McCulloch, W.S., *Embodiment of Mind*, MIT Press, Cambridge, 1965.
- MCCULLOCH W. S., *Recollections of the Many Sources of Cybernetics*, ASC Forum, VI(2), 5–16, 1974.
- MENDELSON E., *Introduzione alla logica matematica*, Bollati Boringhieri, Torino, 1a ed. 1972, rist. 2017.
- MITCHELL M., *L'intelligenza artificiale. Una guida per esseri umani pensanti*, Giulio Einaudi Editore, Torino, 2022.
- MOOR J. H. (a cura di), *The Turing Test. The Elusive Standard of Artificial Intelligence*, in *Studies in Cognitive Systems*, vol. 30, Springer Science+Business Media, 2003.
- NATALE S., *Macchine ingannevoli. Comunicazione, tecnologia, intelligenza artificiale*, Giulio Einaudi Editore, Torino, 2022.
- *Philosophy of mathematics: selected readings* / edited by Paul Benacerraf, Hilary Putnam
- PICCININI G., *The First Computational Theory of Mind and Brain: A Close Look at McCulloch and Pitts's “Logical Calculus of Ideas Immanent in Nervous Activity*, *Synthese*, vol. 141, n. 2, pp. 175-215, p. 176, 2004.
- PUTNAM H., *Minds and Machines*, in Hook S., *Dimensions of Mind*, New York University Press, pp. 57-80, New York, 1960; repr. in Putnam, *Philosophical Papers, vol. 2: Mind, Language and Reality*, Cambridge University Press, Cambridge, 1979 (*Mente, linguaggio e realtà*, trad. it. di Roberto Cordeschi, Adelphi, Milano, 1987).
- RAPAPORT W. J., *The Philosophy of Computer Science*.
- RASHEVSKY N., *Foundations of Mathematical Biophysics*, *Philosophy of Science*, vol. 1, n. 2, 1934, pp. 176-196.
- RASHEVSKY N., *Looking at History through Mathematics*, MIT Press, Cambridge, Massachusetts, 1968.

- RASHEVSKY N., *Mathematical Biophysics. Physico-mathematical foundations of Biology*, voll. 1-2, 3a ed., Dover Publications, New York, 1960.
- RASHEVSKY N., *Mathematical Biophysics*, Nature, vol. 135, 1935, pp. 538-530.
- RASHEVSKY N., *The mechanism of cell division*, in Bulletin of Mathematical Biophysics, vol. 1, 1939, pp. 23-30.
- RESCORLA M., "Church's Thesis and the Conceptual Analysis of Computability", Notre Dame Journal of Formal Logic 48, 2007, pp. 253-280.
- RUSSEL S. J., NORVIG P. (a cura di), *Artificial Intelligence. A Modern Approach*, 4a edizione, Pearson, 2021.
- SIEG W., *Calculations by Man and Machine: Conceptual Analysis*, in Reflections on the Foundations of Mathematics, ed. W. Sieg, et. al., Association for Symbolic Logic, Natick, 2002
- SIEG W., *Hilbert's Programs: 1917-1922*, The Bulletin of Symbolic Logic, Vol. 5, No. 1, Association for Symbolic Logic, 1999, pp. 1-44.
- SOARE R., *The history and concept of computability*, in Handbook of computability theory, Elsevier, 1999.
- TREISMAN A. M., *A feature-integration theory of attention*, Cognitive Psychology, Vol. 12, Issue 1, 1980, pp. 97-136.
- TURING A., *Computing Machinery and Intelligence*, Mind, vol. 59, n. 236, 1950, pp. 433-460.
- TURING A., *On Computable Numbers, with an Application to the Entscheidungsproblem*, in Proceedings of the London Mathematical Society, serie seconda, 42, 1936, pp. 230-67; errata corrige in 43, 1937, pp. 544-46.
- TURING A., *Systems of logic based on ordinals*, in Proceedings of the London Mathematical Society, serie seconda, 45, 1938, pp. 161-228.
- TURING A., *Proposal for the development in the Mathematical Division of an Automatic computing engine (Ace)*, rapporto all'Executive Committee del National Physical Laboratory del 1945, in B. E. CARPENTER e R. W. DORAN, *A. M. Turing's Ace report of 1946 and other papers*, MIT Press, Cambridge Mass, 1986, pp. 20-105, 1945.

