L'IA e la semantica

Alberto Peruzzi

The development of Artificial Intelligence (AI) has led to many controversies of philosophical interest involving issues of logic, computer science, software engeneering. The paper goes through the main steps of this development, deals with their philosophical relevance and sketches the unsolved problems, with special emphasis on the AI models of semantic competence. Some criticisms raised to such models in AI, either in its "classical" or "neural networks" format, are examined. On the ground of the discussion of these criticisms, a view is proposed which avoids the inherited gap between the project of AI and naturalistic philosophy and suggests that filling the gap also provides a filter in order to select a definite form of naturalism.

Keywords: *artificial intelligence*, *semantics*, *logic*, *connectionism*, *naturalism*.

1. *Che senso ha?*

L'Intelligenza Artificiale (IA) è una disciplina rinascimentale. Ovviamente non nel senso che sia stata sviluppata nel Rinascimento. Nasce infatti con un seminario organizzato da John McCarthy nel 1956 al Darmouth College, nel New Hampshire. Ma nel senso che è tanto un'arte quanto una scienza il cui oggetto di studio è definito da ciò che riesce a fare. Cosa vuol fare l'IA? Sistemi intelligenti artificiali. Cosa studia l'IA? Sistemi intelligenti artificiali. I computer, con i loro programmi, ne sono l'esempio canonico.

La storia raccontata nel film di Steven Spielberg che s'intitola *AI* (*Artificial Intelligence*) è quella di un bambino artificiale, che si comporta in maniera simile a un bambino in carne e ossa; invece di essere il risultato di potenzialità scritte nel suo DNA, è il risultato di un meraviglioso programma per calcolatore.

Supponendo d'intenderci su cosa sia un sistema, l'idea è dunque quella di realizzare sistemi 'artificiali', perché non esistono già in natura, ma 'intelligenti', cioè, in grado di fornire prestazioni paragonabili a quelle che consideriamo tali anche se non sappiamo definirle bene. Sappiamo, comunque, riconoscere diversi gradi di intelligenza: un sasso non è intelli-

gente, un gatto è più intelligente di una lucertola. Una scala grossolana? Potremmo difenderci: non si può definire esattamente l'essenza di qualcosa che un'essenza non ce l'ha. Oppure potremmo azzardare un criterio per riconoscere l'intelligenza quando è presente ed eventualmente misurarla. Quale? Un criterio basilare è dato dalla capacità di svolgere un compito, risolvendo un problema. Certo, non basta. Bisogna anche tener conto della varietà dei compiti che si è in grado di svolgere. L'albero che ora vedo nel giardino svolge ottimamente il compito di stare in equilibrio, e meglio di qualunque bipede implume. Un topolino messo in un labirinto riesce a trovarne la via d'uscita prima di un uomo. Un delfino di sei mesi sa muoversi nell'acqua meglio di un bambino di sei anni. Una rondine sa risolvere il problema di volare senza bisogno di protesi metalliche. Ma intelligenza non è soltanto *sapere-come* fare qualcosa; è anche *sapere-che* qualcosa è così e non cosà. Non è soltanto un insieme di procedure, ma anche di conoscenze espresse in proposizioni dichiarative.

Da sempre, gli esseri umani hanno giocato la carta del linguaggio verbale per sancire una distanza incolmabile tra essi e tutti gli altri organismi – a quanto pare, per *sapere-che* ci vuole un quid molto raro, chiamato coscienza. Ma per scomodare la coscienza bisogna che ne valga la pena. Se al computer Hal 9000 di *2001: Odissea nello spazio* fosse stato chiesto di descrivere cosa stava succedendo e avesse risposto immancabilmente «Om» (essendone cosciente), che avreste pensato della sua 'intelligenza'? In un linguaggio che si rispetti, stati di cose diversi sono rappresentati in proposizioni diverse. (D'altra parte non dobbiamo neanche fare come i saggi di Laputa: la sintassi ci fa risparmiare etichette.) E quante proposizioni ci sono in un linguaggio *naturale* come l'inglese o l'italiano? Potenzialmente, un numero infinito. Fossero l'una totalmente autonoma dall'altra, il pensiero sarebbe come un gigantesco, e caotico, album fotografico. Ma l'intelligenza sta anche nel collegare fra loro proposizioni diverse, oltre che nell'averne tante a disposizione. Gli uomini primitivi non erano capaci di svolgere certi compiti, ma, prova e riprova, sapendo che se A allora B e se B allora C ..., hanno saputo comprendere le dipendenze tra fatti e costruire strumenti che li aiutassero a svolgere molti compiti.

Anche confinando l'attenzione ai nostri simili, ci sono diversi tipi d'intelligenza, corrispondenti ai "domini" cognitivi: l'intelligenza di Mozart e quella del Tenente Colombo, quella di Kasparov e quella di Leonardo. Se di ciò non si tien conto quando si va a misurare il famigerato QI, il valore che si ottiene è un po' come una misura della felicità basata sul numero di sorrisi al giorno (non irrilevante, ma non decisivo). Comunque, la caratteristica esclusiva dell'intelligenza *umana* è stata indicata nel pensiero razionale, il quale, per essere riconosciuto, ha bisogno di essere espresso e comunicato. Il mezzo deputato a questo scopo è

appunto il linguaggio e non è solo un medium per codificare in un formato opportuno tante informazioni di tipo diverso: è anche un *organon* del pensiero. Serve a *rappresentare* le nostre conoscenze in un formato che ne favorisce la memorizzazione, ne ottimizza il recupero e serve a combinarle tra loro in maniera da ottenerne di nuove.

Perciò si è ritenuto che un sistema di IA dovesse in primo luogo essere un sistema che dispone di un linguaggio e che ne manipola in modo appropriato i simboli. 'In modo appropriato' significa: conformemente a un preciso insieme di regole. Che tipo di regole? Dovendo corrispondere a procedure 'razionali' di elaborazione delle informazioni, cosa potrebbero essere se non regole *logiche*?

La logica è una disciplina antica, solo che per poter essere implementata su un hardware diverso dal nostro cervello ha dovuto aspettare che si capisse come tradurre i processi logici nei termini di un calcolo e che fossero disponibili le risorse tecnologiche adatte per far eseguire tale calcolo a una *macchina*. I sistemi dell'IA classica sono principalmente macchine logico-linguistiche. Di macchine, come congegni prima meccanici e poi elettrici, nel corso della storia se ne son costruite tante. Solo recentemente sono state costruite macchine logico-linguistiche. Premesso che non tutte le regole logiche e ancor meno tutti i linguaggi si prestano allo scopo, buona parte della ricerca in IA si è occupata di come automatizzare la logica e di come costruire un linguaggio (artificiale) opportuno per la codifica, la trasmissione e il recupero delle informazioni.

Gli ostacoli di natura tecnica, sotto il profilo matematico e ingegneristico, non sono mancati. S'è cominciato a dubitare che i progetti, quanto mai ambiziosi, dell'IA potessero essere candidamente dilazionati (ai computer della prossima generazione). Anzi, sono state fatte obiezioni di principio al progetto stesso dell'IA. Anche se ridimensioniamo le aspettative iniziali sulle magnifiche sorti e progressive dell'IA, il nostro ambiente di vita si sta comunque popolando di prodotti che derivano da ricerche di IA impensabili solo qualche anno fa; sarebbe troppo facile ridurre il significato dell'IA a un insieme di ricette orientate al mercato.

Oggi, semmai, non c'è più un unico paradigma-guida nel campo dell'IA e, in particolare, l'enfasi sull'architettura dei linguaggi e l'idea che si debba puntare a una 'rappresentazione' delle conoscenze corrispondente a un sofisticato assetto ipotetico-deduttivo non è più un tratto condiviso. Andrew Brooks ha costruito al MIT degli insetti artificiali col preciso scopo di mostrare che l'IA può e deve fare a meno di sistemi rappresentazionali. Si sta cercando di ridurre la distanza dell'IA dall'intelligenza naturale non partendo dal carro (linguaggio e logica) ma dai buoi, e così l'attenzione si sposta verso sistemi che siano in grado di auto-organizzare la propria 'intelligenza' disponendo unicamente di *pattern* comportamentali pre-linguistici.

I computer sanno fare velocemente e bene cose che noi facciamo lentamente e commettendo errori, mentre non sanno fare cose che per noi sono banali. Eppure, ogni volta che un tipo di simulazione artificiale fallisce, impariamo per via *sperimentale* qualcosa di prezioso sull'intelligenza naturale – qualcosa che sarebbe stato difficile imparare in altro modo. E sapendo come *non* costruire un sistema di IA potremo costruirne di più efficienti. È presumibile che robot intelligenti si diffonderanno sul pianeta. C'è chi dice fra meno di un secolo. I problemi *filosofici* posti da una simile eventualità sono numerosi. È possibile che sistemi di IA superino l'intelligenza umana e finiscano per trattarci come animali domestici? Dovremo considerare una macchina responsabile delle sue azioni? Di fatto, non abbiamo ancora costruito macchine che vagamente ci somiglino. Gli ostacoli non mancano e non è neanche detto che siano superabili. Per capire quali siano conviene partire da quelli che abbiamo superato.

2. *Un po' di storia*

Gli esseri umani hanno messo a punto due tipi basilari di sistemi simbolici con i quali organizzare in un codice discreto e sequenziale la struttura del continuo spazio-temporale e di tutto ciò che è immerso in esso: il linguaggio verbale e il calcolo numerico. Che si possa parlare dei numeri e dei calcoli eseguibili con essi è ovvio. Meno ovvio è che si possa codificare in termini numerici tutto quanto è esprimibile in un qualsiasi linguaggio. E qui sta appunto la chiave di ciò che fanno gli odierni computer. L'idea di una codifica digitale di informazioni che normalmente coinvolgono rappresentazioni analogiche, però, non è nuova. Risale nientemeno che ai pitagorici.

Sono passati più di duemila anni, durante i quali l'idea si è fatta strada, superando numerosi ostacoli pratici e teorici, aggirandone altri e ignorandone, in modo opportunistico, altri ancora. Per fare un esempio, la scoperta degli irrazionali sembrava aver definitivamente frustrato le speranze della discretizzazione, ma lo slittamento tipico della filosofia moderna dall'*ordo rerum* all'*ordo idearum* permise di riaprire i giochi. Dopo l'equazione tra pensare e calcolare affermata da Hobbes, il divino Leibniz iniziò l'algebrizzazione sia della struttura dei concetti sia della struttura dei ragionamenti. Il controllo sulla correttezza di un sillogismo avrebbe potuto finalmente essere realizzato con una procedura del tutto 'meccanica', analoga alla prova del 9 per le ordinarie divisioni. Chi però riuscì a dare una compiuta sistemazione all'idea di un calcolo algebrico della logica fu George Boole: e ciò avvenne intorno alla metà dell'Ottocento. Nello stesso periodo si ebbe, grazie a Charles Babbage, anche un primo contributo all'implementazione di un calcolo simbolico. Se la mac–

china calcolatrice di Pascal consentiva di eseguire solo operazioni aritmetiche, la macchina analitica di Babbage consentiva di rappresentare algoritmicamente una gamma molto più ampia di dati, servendosi di una tecnologia che anticipava l'impiego delle schede perforate.

A parte quest'anticipazione, è stato solo per mezzo del rigoglioso sviluppo della logica matematica tra la fine dell'Ottocento e gli anni Trenta del Novecento che si è riusciti a dare un assetto stabile alla codificazione formale dei tipi più diversi di inferenze, ben più complesse di quelle rappresentabili in termini di algebre di Boole, perché entrano in gioco relazioni e quantificatori (come nell'implicazione 'Se c'è qualche individuo che ama tutti, allora ognuno è amato da qualcuno'). Si veniva, così, precisando l'architettura di linguaggi rigorosi che si differenziavano dalle lingue parlate in misura più marcata di quanto fosse mai avvvenuto con il linguaggio in cui erano (e sono) scritti i testi usuali di matematica. Il linguaggio logico di riferimento è il linguaggio della logica del primo ordine, in cui i quantificatori vincolano solo variabili individuali. (In esso non è esprimibile l'implicazione 'Se c'è una relazione che qualcuno ha con tutti, allora è riflessiva'). Questo e altri linguaggi logici si presentavano come il formato definitivo in cui avrebbe dovuto esprimersi tutta la conoscenza e non solo quella matematica, ma certo il primo passo era la rifomulazione in essi di tutta la matematica. Oltre a fornire la matrice di questi linguaggi e ad elencare i principi di costruzione dell'edificio matematico, la logica si candidava a fornire la base più solida su cui l'edificio avrebbe pouto poggiare. Ma come garantire in maniera incontrovertibile che la solidità dell'edificio non sarebbe mai stata minacciata dall'insorgere di contraddizioni?

Fu proprio dai problemi incontrati nel tentativo di fornire tale garanzia che trassero origine quelle ricerche che avrebbero portato a una esplicita teoria generale degli algoritmi di calcolo. David Hilbert sperava che si sarebbe riusciti a garantire la coerenza (noncontraddittorietà) dell'aritmetica servendosi esclusivamente di manipolazioni finitarie dei simboli formali presenti nel linguaggio aritmetico – manipolazioni che a loro volta sono rappresentabili nella stessa aritmetica. Ciò richiedeva (i) la considerazione di tutte le possibili dimostrazioni che l'aritmetica era in grado di ospitare e (ii) la prova che nessuna di queste portava a una contraddizione. Nel 1931, Kurt Gödel dimostrò due teoremi di 'incompletezza', in base ai quali il programma 'formalista', così come tracciato da Hilbert, non poteva essere realizzato. Il primo teorema stabiliva l'esistenza di proposizioni aritmetiche che non si possono né dimostrare né refutare nell'aritmetica; il secondo stabiliva che l'aritmetica non può dimostrare la propria coerenza. E tutto ciò dipendeva dal fatto che il linguaggio aritmetico (proprio come una qualsiasi lingua naturale) è capace di rappresentare al proprio interno la sua stessa sin-

tassi, cioè, di codificare numericamente tutte le proprietà logico-sintattiche di ogni possibile discorso sui numeri. Per ottenere i due teoremi, Gödel fu indotto a isolare e a precisare formalmente il concetto intuitivo di procedura meccanica (algoritmo, funzione computabile) dando avvio a quella teoria delle funzioni 'ricorsive' che di lì a poco avrebbe trovato anche altre formulazioni, tra le quali va segnalata quella introdotta da Alan Turing e che da lui prende nome: la teoria delle 'macchine di Turing'.

Una macchina di Turing non è qualcosa di fisico, bensì è un congegno ideale di computo. (Nei riferimenti bibliografici ho indicato alcuni testi che forniscono un'ottima descrizione di come è strutturata una macchina del genere e di come essa equivalga, quanto a potenza di calcolo, a una funzione ricorsiva). Grazie al fatto che esistono macchine di Turing *universali*, cioè, in grado di simulare le computazioni di qualunque altra macchina di Turing, questa teoria ha catalizzato una gran mole di ricerche, costituendo il paradigma teorico per la descrizione di ciò che un reale computer può fare.

Turing formulò anche un test, su cui i filosofi hanno versato un fiume d'inchiostro. In breve, il test di Turing consiste nel mettere a confronto le risposte di un essere umano, come campione di sistema intelligente, con quelle fornite da un sistema artificiale S: se un giudice imparziale non è capace di discernere le risposte dell'uno da quelle dell'altro, il test è superato da S e conseguentemente non si potrà negare intelligenza a S.

Con tutto questo siamo ancora sul piano teorico. Ma già negli anni Quaranta, McCulloch e Pitts avevano sfruttato la descrizione logica dei circuiti elettrici in termini di algebre di Boole per modellare le reti nervose. A partire dai prototipi iniziati nel 1941, nel 1946 fu realizzato il primo efficiente calcolatore, l'ENIAC, alla cui costruzione parteciparono sia 'John' von Neumann, che aveva lavorato al programma di Hilbert, sia Norbert Wiener, che in quegli stessi anni stava gettando le basi di una nuova disciplina: la *cibernetica*, come scienza dei processi di autoregolazione mediante cicli retroattivi (feedback) – pensate al funzionamento di un termostato.

Mancava un ultimo tassello al mosaico. Claude Shannon e William Weaver pubblicarono proprio allora il saggio con cui nasceva la *teoria dell'informazione*. Per precisare la ridondanza necessaria alla comunicazione efficiente di un messaggio lungo un canale soggetto a rumore, si definiva la 'quantità d'informazione' del messaggio come inversamente proporzionale alla sua probabilità e la si faceva corrispondere al numero di alternative binarie (0/1) richieste per definire il messaggio. Di qui il *bit* (acronimo di *bi*nary digi*t*) come unità di misura per calcolare l'informazione. Il codice numerico d'elezione diventava dunque quello binario

invece di quello decimale perché gli si potevano correlare i due stati fisici basilari di ciascuna porta di un circuito (on/off). Nel linguaggio-macchina di un computer si trovano infatti lunghe liste di 0 e di 1, elaborate in conformità a precisi programmi, i quali non sono altro che casi particolari degli algoritmi già impostati teoricamente da Gödel e Turing. Per gestire efficientemente i programmi, però, gli informatici avevano bisogno di linguaggi più comodi, di alto livello, e siccome tutto ciò con cui si aveva a che fare erano liste di liste ... di liste di simboli, uno dei primi linguaggi (di 'programmazione') ad affermarsi fu il LISP (LIst Processing), introdotto da John McCarthy. In seguito sono stati formulati molti altri linguaggi di programmazione, uno dei quali, il PROLOG (PRogramming LOGic) s'ispira alle caratteristiche del linguaggio della logica del primo ordine.

Avendo a disposizione questa discreta gamma di risorse informatiche e le tecnologie ingegneristiche di supporto, si potevano realizzare sistemi fisici che, comportandosi in conformità a uno o più programmi specificati in linguaggi opportuni, simulassero prestazioni cognitive. Cominciava il cammino dell'IA vera e propria.

Il lavoro dei ricercatori in IA si è ben presto diviso in aree, ciascuna delle quali affronta una specifica tematica e richiede metodi risolutivi differenziati. Le principali aree dell'IA hanno ormai denominazioni consolidate in inglese e così le elencherò: *problem solving*, *planning*, *learning*, *automatic theorem proving*, *language understanding*, *knowledge representation*, *robotics*. Evitando ogni tecnicismo, in quanto segue ne discuterò le idee di fondo e ne segnalerò quelle difficoltà che hanno alimentato il dibattito filosofico sull'IA. Qui mi limito ad aggiungere solo una cosa: nell'ultimo decennio, i numerosi problemi incontrati nel tentativo di mettere a punto sistemi effficienti di IA hanno spinto ad abbandonare il quadro teorico classico, costituito dall'architettura sequenziale, 'alla von Neumann', dell'elaborazione, e a sostituire questo quadro con uno radicalmente diverso, basato sull'architettura *in parallelo* tipica di quelle si chiamano 'reti neurali'. Cercherò di richiamare brevemente l'idea che guida la costruzione di reti neurali artificiali e di accennarne la rilevanza filosofica, senza nascondere i loro limiti.

3. *Problemi e strategie*

La simulazione artificiale di un problema richiede un sistema di simboli in cui codificare i dati e la sua (eventuale) soluzione; e richiede una manipolazione di questi simboli conforme a un algoritmo di calcolo, seguendo il quale, passo dopo passo, si generi alla fine una codifica della soluzione cercata. Allorché si dispone di un pacchetto di tecniche di problem solving si può impostare la simulazione di *piani* in cui tali

tecniche vengono sfruttate in serie. Ciò è d'interesse immediato per la psicologia: l'importanza dei piani di comportamento era già stata messa in evidenza dai primi cognitivisti. Benché il *planning* non sia un'area nettamente distinta da quella del *problem solving*, convolge aspetti nuovi, come l'anticipazione degli effetti collaterali derivanti dall'esecuzione di una strategia e come l'ottimizzazione delle risorse al fine di conseguire un obiettivo.

È facile costruire un sistema artificiale che segua le regole degli scacchi e riconosca quando c'è una situazione di scacco; altra cosa progettare un algoritmo (programma) che sia in grado di giocare decentemente. Negli ultimi vent'anni sono stati fatti passi da gigante in questa direzione. Ha fatto epoca la notizia che il campione del mondo, Kasparov, aveva perso una partita contro il sistema Deep Blue (IBM). Tuttavia, esiste un efficace generatore di mosse che *garantisca* la vittoria contro qualunque avversario? No, e i giochi in cui è possibile specificare in anticipo una strategia vincente, guarda caso, ci attirano ben poco. Nel caso degli scacchi, nonostante il fatto che i pezzi, le case della scacchiera e le mosse in ciascun momento possibili siano in numero finito, l'insieme delle alternative indotte da una sequenza di mosse è incredibilmente grande. Più grande del numero stimato degli atomi dell'universo.

Passando dagli scacchi al linguaggio, supponiamo di riuscire a fare un programma che elenchi tutte le frasi della lingua italiana (sequenze di mosse nel 'gioco' della grammatica) in ordine alfabetico – nell'ipotesi di aver ibernato il lessico italiano in qualche dizionario, cioè in un *data base* finito. Fra quelle frasi c'è anche l'insieme di tutte le terzine della *Divina Commedia*, ma potreste aspettare un'eternità prima che il programma arrivi a una data terzina!

Quando le possibilità di combinare tra loro, secondo ben precise regole, un insieme finito di elementi crescono a dismisura, si ha un'*esplosione combinatoria*. Gli scacchi sono un esempio della rapidità con cui l'esplosione si può verificare, eppure sono un sistema fatto di poche cose. Più in generale, anche se sappiamo che *c'è* una soluzione a un dato problema, e anche se sappiamo che *c'è* un algoritmo che genera la soluzione, sapere tutto questo può non essere di alcun aiuto pratico a trovarla. Non è esaminando una a una tutte le possibili alternative che procediamo nella maggior parte delle nostre attività. Riferito al linguaggio: se per capire il significato di un enunciato devo predisporre un registro di tutti i significati possibili, non farò mai in tempo a capirne uno.

E qui c'è già una piccola morale: per un comportamento efficiente non occorre essere onniscenti e non basta seguire le leggi della logica. Come facciamo allora a cavarcela con l'esplosione combinatoria? Organizziamo la ricerca della soluzione restringendo le alternative (potando

l'albero delle opzioni) e a questo scopo ci serviamo di un'*euristica:* una o più strategie di ricerca per scartare rapidamente alcuni dei cammini possibili (nello *spazio* del problema). Se non trovate il vostro portafogli, restringete il numero dei possibili posti dove cercarlo. E se non vi ricordate il nome di battesimo di Cavour, non lo cercate nei libri di astronomia.

Nelle simulazioni si è partiti dalle strategie di ottimizzazione per ridurre al minimo la probabilità di 'perdere' (non riuscire a risolvere un problema) e massimizzare la probabilità di 'vincere' (riuscirvi). E così si son scoperte strategie diverse da quelle che di fatto seguiamo e perfino più efficienti per certi scopi. Le strategie artificiali possono anzi gettar luce sulle caratteristiche del tipo d'intelligenza che ci è propria, proprio come abbiamo imparato qualcosa sull'efficienza della forma degli uccelli dopo aver progettato la forma degli aerei. L'importante è che tali strategie siano codificabili in termini di qualche algoritmo (procedura ricorsiva) e implementabili sui circuiti di un calcolatore mediante programmi scritti in un linguaggio opportuno.

L'euristica si collega in modo naturale con valutazioni di rilevanza. Sono state studiate diverse *logiche della rilevanza*. Si tratta di logiche in cui, per poter asserire che il condizionale 'se *p* allora *q*' è valido, *p* deve essere rilevante per stabilire *q*. (Nella logica standard, invece, basta che *q* sia valido perché il condizionale 'se *p* allora *q*' sia a fortiori valido, qualunque sia l'ipotesi *p*). In realtà, ciò che si suppone essere rilevante non resta sempre tale e ciò che non era rilevante lo può diventare. Per problemi complicati, è utile avere a disposizione più euristiche oppure garantirsi che l'euristica adottata sia flessibile. Ma quando si lavora con più euristiche occorrono criteri per *pesarle*, cioè, per comparare tra loro le diverse strategie di ricerca rispetto al compito da svolgere, rispetto alle risorse disponibili, ecc. La rilevanza non va considerata soltanto in rapporto a come sono trattate le informazioni entro una data euristica, ma si ripropone anche al meta-livello, allorché elaboriamo un piano confrontando le diverse euristiche. Perciò, la formalizzazione logica della rilevanza richiede che vi si includa la metalogica e allora le cose si fanno meno nitide. Benché siano eleganti costruzioni, dubito che le 'logiche della rilevanza' conseguano gli ambiziosi scopi loro assegnati.

L'interesse per il problem solving si iscrive in una bimillenaria tradizione filosofica, che ha fatto della padronanza del metodo un ingrediente primario della razionalità, qui non importa se privilegiando metodi deduttivi o induttivi. Il metodo, però, non è mai il primum. Le tecniche di problem solving partono sempre da un quesito che si suppone *dato*. Non meno importante, specie a fini formativi, è la capacità di porre, identificare, inventare o scoprire, un problema. Il progresso della cultura, nella storia delle scienze e delle arti, così come nella maturazione di

ciascun essere umano, è dipeso e continua a dipendere dalla capacità del *problem raising*, del sollevare problemi. Se anche riuscissimo a costruire sistemi artificiali in grado di risolvere un qualunque problema che sia risolvibile, non avremmo ancora riprodotto una mente capace di porre nuovi problemi. Piuttosto, non varrebbe forse la pena impegnarsi a educare anche questa capacità con l'ausilio di sistemi di IA?

Ma che cosa ci aspettiamo poi da una simulazione? Oltre alle note illusioni ottiche, ci sono errori inferenziali ricorrenti in cui gli esseri umani cadono. Inizialmente, l'IA si è proposta di simulare le 'buone' prestazioni cognitive degli esseri umani, non di ripeterne gli errori – uno dei motivi per cui i calcolatori sono entrati prepotentemente nella nostra vita. È fin troppo facile addurre esempi di virtù proprie della nostra mente che un computer non riesce a riprodurre, dimenticando che gli esseri umani hanno anche vizi dai quali i computer sono immuni. Semmai, il tremendo sospetto è che, in un sistema di IA, il miglioramento di certe nostre prestazioni vada a scapito di altre e, più specificamente, a scapito proprio di quelle che ci mettono in grado di porre nuovi problemi.

Come s'è detto, il problem solving è collegato alla pianificazione. Se c'è la possibilità di scegliere tra piani alternativi, entra in funzione un componente essenziale della nostra vita: prendere *decisioni*. Sarebbe dunque opportuna un'educazione a delimitare lo spazio del problema e a fornire un'ampia gamma di euristiche, perché una decisione efficace comporta spesso la modifica dello spazio delle soluzioni. Tutti sappiamo come questa modifica sia difficile, data la tendenza a conformarci alle strategie acquisite. È il caso del problema seguente: con 6 fiammiferi formare 4 triangoli equilateri. La soluzione esiste ma non nel piano (cui ci sentivamo confinati a cercarla).

Quando il dominio degli oggetti e l'insieme degli stati di cose cui gli oggetti possono dar luogo sono fissati, si possono definire sistemi di IA che risolvano problemi concernenti il dominio dato. I cosiddetti *sistemi esperti* sono specializzati nel trattamento di informazioni relative a una ristretta cerchia di questioni, formulate in un linguaggio rigidamente definito e concernenti un prefissato dominio di cose e loro proprietà. Ci sono sistemi esperti impiegati nella diagnostica medica, nella ricerca di idrocarburi, nella progettazione architettonica, ecc. Ciascuno di essi tratta un micromondo e vi resta confinato. Questo stesso confinamento al dominio in cui sono, appunto, 'esperti' ne ha favorito la definizione e il successo commerciale. Ma specializzazione può voler dire anche ottusità e qui sta infatti il loro limite, specie quando si usano per fare modelli della comprensione del linguaggio. I sistemi esperti pagano l'acutezza in particolari compiti con la miopia nei confronti di ciò che va oltre il dominio. I filosofi amano denunciarne la scarsa flessibilità, dimenticando che un non minore difetto sarebbe la totale flessibilità.

4. *Piccoli mondi*

Nell'area della *pattern recognition*, uno dei primi programmi, relativo a un micromondo di solidi geometrici è stato SHRDLU: un ambiente fatto di blocchi colorati, su cui operano più *parser* specializzati, con una correlazione tra un numero ristretto di nomi e forme, così come tra un numero ristretto di verbi e azioni da compiere su questi solidi. Il programma è scritto in LISP. L'analizzatore sintattico stabilisce se un insieme di simboli in ingresso corrisponde alle regole della grammatica per gli enunciati che SHRDLU è in grado di riconoscere; l'analizzatore semantico gestisce i dati su come si presenta lo "scenario" di volta in volta; infine c'è un minisistema che deduce informazioni dai dati forniti e stabilisce se le richieste fatte sono coerenti con il micromondo. Domandina: il sistema ha davvero a che fare con un piccolo mondo di solidi o manipola soltanto simboli? Se manipola soltanto simboli, non ha in realtà una *semantica* ....

SHRDLU fu inventato da Winograd all'inizio degli anni '70 e da allora sono stati realizzati numerosi programmi per riconoscere configurazioni comuni (tavoli, volti, ...), tenendo conto, a differenza di SHRDLU, anche del modo in cui due forme si congiungono e non semplicemente della posizione relativa. L'analisi delle giunture è stata poi estesa a intere scene, evidenziando i vincoli che riducono il numero delle possibilità interpretative di una figura in scene diverse. La comprensione dei vincoli che la fisica impone sulle giunture è sufficiente a scartare combinazioni incoerenti di linee e superfici. Tuttavia, quando ci si è spinti a simulare la rappresentazione della conoscenza che abbiamo del "mondo del senso comune", quest'approccio si è appesantito, richiedendo l'esecuzione di calcoli giganteschi. Non è solo che i programmi necessitano di un gran numero di dati, corrispondenti alla *background knowledge* inglobata nel senso comune. È che prima bisogna esplicitare tale conoscenza. Come se non in un sistema d'assiomi? Lo stampo prediletto a tale scopo è risultato quello della logica del primo ordine, per la quale esiste una semantica rigorosa (quella tarskiana) in cui le nozioni di riferimento e verità trovano precisazione formale. Così facendo, gli studiosi di IA hanno riscoperto le difficoltà inerenti al *formato* delle conoscenze ciò che più diamo per scontate. A differenza dei filosofi, non sono però rimasti paralizzati dalla complessità di ciò che vediamo e di cui parliamo nella vita quotidiana e non hanno visto il peccato originale nella modellizzazione.

Complessità che si rivela in un quesito: se la conoscenza è il prodotto di un insieme di regole meccaniche organizzate in programmi, come si fa a applicare queste regole in modo da ottenere qualcosa che sia diverso da quanto è stato finora ottenuto seguendole? Come si fa a integrare la 'produttività' meccanica del sistema di regole con la creatività?

Gli esseri umani sono in grado di eseguire quest'integrazione; le arti e le scienze hanno una storia punteggiata di innovazioni rispetto a un preesistente sistema di regole. Nel nostro stesso linguaggio la creatività si manifesta senza bisogno d'essere coscienti e si manifesta *solo* relativamente a un sistema di regole. Un'IA che volesse davvero riprodurre il modo in cui sappiamo gestire i significati dovrebbe garantire la possibilità di crearne di nuovi, senza rinunciare a vincoli sulla loro gamma.

Stiamo girando intorno al *frame problem*, cioè al problema della cornice-contesto, formulato da McCarthy nel 1969. Applicando una trasformazione a un elemento di un sistema, lo stato del sistema cambia, ma in un modo che non è facilmente circoscrivibile, perché il cambiamento subìto da un elemento ha effetti collaterali su altri elementi, anche esterni al sistema, che a loro volta hanno effettti retroattivi. Come si fa a tener sotto controllo la catena di questi *feedback*? Nei modelli di IA si parte da un insieme finito di *pattern* basilari di dati e si sfrutta un insieme finito di *pattern* di trasformazione. Con vincoli troppo rigidi, abbiamo un'intelligenza analoga a quella di un robot da catena di montaggio. Senza vincoli s'arriva alla casualità o alla paralisi. Nel caso del linguaggio: o per comprendere una nuova frase tiriamo a indovinare oppure bisognerebbe già comprendere tutto il linguaggio, e allora bisognerebbe comprendere qualunque altra cosa: assurdo.

Dobbiamo accettare quest'aut-aut? È ovvio che per comprendere una frase occorre disporre di informazioni aggiuntive a quelle espresse, com'è ovvio che per *comprendere* tali informazioni bisogna avere a che fare col mondo. Non ne segue che siamo tenuti a misurarci con la totalità infinita delle espressioni e dei fatti possibili. Possiamo rivedere una delle nostre aspettative, purché ne teniamo fisse altre (non sempre le stesse). Così la semantica 'naturale' riesce a essere un sistema relativamente stabile globalmente nonostante i cambiamenti locali – e questa è una caratteristica dei sistemi *aperti* in fisica.

Eccone una spia rivelatrice: quando si fa un ragionamento, si danno per buone molte cose. Accettiamo per *default,* cioè fino a prova contraria, che valgano certi presupposti e non ci curiamo di esplicitarli. Questa comoda pratica non garantisce la validità logica delle inferenze. Be', nella vita quotidiana non ci aspettiamo garanzie assolute. Ciò che è interessante è che sono stati messi a punto diversi sistemi di IA che incorporano proprio quest'idea.

Siete certi che uscendo di casa troverete la città come l'avete lasciata ieri? No, nel frattempo buona parte di quel che c'era potrebbe essere scomparso a causa di un'invasione aliena. Allora, uscendo, starete attenti al minimo passo che fate? No, non rivedrete la vostra aspettativa di trovare tutto tale e quale, *a meno che* non abbiate acquisito informazioni (attendibili) circa i cambiamenti. Ma il fatto di rivedere un'assunzione

non si propaga a catena in una revisione di tutte le vostre precedenti informazioni. L'adozione di stereotipi di oggetti e di situazioni evita dunque l'esplosione combinatoria ma ha un prezzo: rende il sistema abbastanza rigido, cioè, intollerante ai cambiamenti. L'adozione di una strategia-modello è tanto utile a velocizzare la ricerca di soluzioni quanto d'ostacolo a risolvere problemi d'altro tipo. Perché non costruire un sistema in grado di impostare e risolvere tutti i problemi possibili? Questa era stata una delle speranze iniziali dell'IA, ben presto ridimensionata entro i confini di un dominio specifico, cui appunto si rivolge un sistema esperto.

Per questa via, a partire dall'ipotesi che il significato sia una rappresentazione simbolica e che la sua comprensione consista nell'esecuzione di una procedura conforme a un programma, si è aperta la strada per la simulazione, mediante computer, della semantica. Sono stati avanzati diversi tipi di 'semantica procedurale' che descrivono l'attività di comprensione/interpretazione di enunciati sotto forma di procedure di *information processing*.

Un'architettura che ha ricevuto particolare attenzione è stata quella delle reti semantiche. Il lavoro originale di Quillian in cui vennero proposte risale al 1968. Anche le reti semantiche sono formulate in LISP e corrispondono a grafi formati da un insieme di vertici (nodi) occupati da voci lessicali e da un insieme di archi, detti *link*, che collegano un vertice a un altro, esprimenti proprietà di un nodo o relazioni tra nodi. Purtroppo, i modelli a rete semantica presentano diversi difetti, di carattere teorico e applicativo. Accanto ai modelli a rete, ne sono stato elaborati altri, che però fanno sì che anche la più semplice analisi semantica richieda complicati calcoli (e tempi sempre più lunghi). A un certo punto si è cominciato a dubitare che la rappresentazione computazionale di un enunciato possa mai garantire, e spiegare, la comprensione del significato.

Le reti semantiche danno modo di precisare alcuni aspetti di un'idea filosofica che va ancora per la maggiore: l'*olismo*. Ha già fatto capolino con l'aut-aut di cui sopra e, per ragioni che qui non posso discutere, la considero un'idea sventurata. Mi limito a un esempio. Gli olisti dicono che un qualunque costrutto concettuale non si riduce ai suoi componenti, ma la sua identità è definita dalla rete dei suoi link con altri costrutti. Supponiamo che la vostra auto non funzioni. Viene il meccanico, la rimorchia in officina e dopo qualche giorno vi chiama dicendo: il sistema d'accensione è a posto, la trasmissione funziona, il motore non ha problemi, ... ma la macchina non va, evidentemente perché qualche suo rapporto col resto dell'universo è cambiato. Accettereste una simile diagnosi? *Ovviamente*, la composizionalità (il tutto è funzione delle parti e di come si combinano) ha senso solo se i componenti sono quelli della taglia giu-

sta e se sono combinati nel modo giusto. Prendendo l'insieme degli atomi che formano le parti dell'auto e ridisponendoli a caso, è improbabile che l'ammasso risultante sia guidabile su una strada. Forse dobbiamo tener conto di ciò che è esterno all'auto per capire come funziona l'auto? *Ovviamente*, le strade sono fatte perché le auto ci possano viaggiare e le auto sono fatte per viaggiare sulle strade. La buona filosofia fa ritrovare presunte banalità, la cattiva filosofia le svende come spiegazioni.

Ricordate la 'domandina'? La semantica in termini di reti non spiega il riferimento e le condizioni di verità. Lo stesso vale per la semantica in termini di atomi di significato o in termini di un sistema di postulati. L'approccio logico alla semantica ha esplicitato e precisato le nozioni di riferimento e verità servendosi di modelli matematici, i quali hanno occultato gli aspetti procedurali nella *comprensione del significato* – quegli aspetti che l'IA privilegia.

Perché non accontentarsi di un'idea più ecumenica della semantica? Affiancare dimensione referenziale e dimensione 'intensionale', l'esterno e l'interno, ambiente naturale-culturale e ambiente mentale. Ma è possibile descrivere separatamente queste due dimensioni? È possibile risolvere i problemi connessi al riferimento e alla verità senza tener conto delle strutture della mente, e viceversa? Non basta dire no: se una dimensione non si riduce all'altra, come interagiscono? La mia ipotesi è che conoscenza del significato e conoscenza del riferimento siano *parametriche* e che i valori dei parametri pertinenti non possono essere anticipati una volta per tutte; eppure, in ogni contesto d'uso, significato e riferimento di qualunque espressione sono 'insaturi' quanto basta. Ammettendo quest'ipotesi, è a portata di mano una soluzione del classico puzzle: (1) il significato del tutto è funzione del significato delle parti, (2) il significato di un enunciato è dato dalle sue condizioni di verità, ma (3) la presenza, nel linguaggio, di operatori epistemici e modali sembra obbligarci al sacrificio di (1) o di (2). Soluzione: la determinazione del significato del tutto è un procedimento di *back and forth* tra significato delle parti e parametri, che porta a un'equilibrazione dei rispettivi vincoli. E ciò si applica anche ai modelli basati su *frame* e *script*.

La nozione di *frame* è stata introdotta da Minsky nei primi anni '70. Un *frame* è una struttura di dati che rappresenta un oggetto o una situazione stereotipici, codificandone le caratteristiche consuete e le aspettative che ne vengono indotte. Per esempio, un *frame* per SEDIA codifica le informazioni che associamo a una tipica sedia e un *frame* per PASSEGGIATA IN MONTAGNA le informazioni su pendenza del terreno, tipo di paesaggio, temperatura, abbigliamento – insomma ciò che si richiede per capire un discorso su una passeggiata in montagna.

Su questa base si sviluppa un ragionamento per default, che assume *fino a prova contraria* le informazioni codificate nel dato *frame* (o in

più di uno). Se l'informazione espressa da *p* è memorizzata come componente di un *frame*, l'enunciato *p* è assunto come vero *a meno che* si verifichi qualcosa di diverso da quanto ci si aspettava. Le potenziali revisioni non sono tutte allo stesso pari: i componenti di un *frame* sono strutturati gerarchicamente. I nodi più alti sono occupati da informazioni assunte come quasi-costitutive dell'oggetto o situazione (per esempio, una sedia è un solido, una passeggiata in montagna è un'azione prolungata), i nodi meno alti da informazioni assunte per default ma più rivedibili. Insomma, la rivedibilità (flessibilità) decresce dai nodi più bassi a quelli più alti. Una sedia a rotelle va ancora bene, mentre un oggetto su cui sedersi che non avesse schienale non sarebbe considerato una sedia. Un computer munito del *frame* relativo può elaborare una buona rappresentazione simbolica del significato di frasi concernenti l'uso comune delle sedie.

Un *frame* può anche contenere (sub)*frame* e far parte di scenari e piani complessi d'azione. Nel 1977, Schank ha introdotto il concetto di *script* (copione, sceneggiatura), per rappresentare in forma di programma stereotipi di eventi in sequenza, quali ingredienti indispensabili alla comprensione delle frasi. Mentre i 'modelli' della semantica formale elaborata all'interno della logica sono insiemi di dati strutturati *chiusi*, come mondi a se stanti, rigidamente definiti in ogni loro particolare, gli *script* lasciano spazio a flessibilità e indeterminatezza.

La valutazione di programmi del genere si fa in base alle *inferenze* che essi traggono da un testo e in base alla qualità delle *parafrasi* che sono in grado di generare. Ora, il compito di analizzare un comune testo e parafrasarlo richiede strutture cognitive aperte a *tutte* le possibilità interpretative (infinite)? Se fosse così, l'esplosione combinatoria avrebbe la meglio, bloccando l'azione non meno della comprensione. Le idee di *frame* e *script* limitano le opzioni di volta in volta significative e consentono di indagare sperimentalmente il carattere *dinamico* della competenza semantica. Tuttavia vanno incontro alle stesse difficoltà notate in rapporto ad altri modelli: la plasticità che *frame* e *script* consentono è sempre quella predefinita da un programma, perciò si ripropone il problema di come determinare ciò che resta invariato quando c'è qualcosa che varia. Se rompo un vaso pieno d'acqua su una sedia, non cambia né il colore né la forma della sedia, però la sedia e il pavimento si bagnano, mentre il soffitto no. Banale? Per un sistema di IA, no. Per noi sì, perché facciamo affidamento a un repertorio di altri taciti presupposti. La nostra comune *background knowledge* non è certo immune da rigidità, eppure ammette un ampio *bricolage*. Nel regno animale, se da un lato la scarsa flessibilità dà pure un vantaggio adattativo, in non poche circostanze può essere fatale. A differenza di quella darwiniana, la selezione dei programmi è fatta dal mercato: pensate ai sistemi di OCR. Gli esseri

umani riescono a destreggiarsi in ambienti mutevoli con una continua riequilibrazione di costanza e variazione, ottusità e apertura, rigidità e flessibilità. Come riprodurre artificialmente un'efficienza così finemente modulata?

5. Conoscenze e programmi

Se l'idea della mente come elaboratore di informazioni in forma simbolica ha i suoi limiti, i limiti non ne cancellano i meriti. Ora, la rivoluzione tecnologica associata, prima, allo sviluppo e alla diffusione dei personal computer e, poi, alla loro connessione in internet sta portando anche a una rivoluzione nel modo di organizzare la conoscenza e *... nel modo di pensarci come esseri pensanti*.

L'IA rende i classici problemi dell'epistemologia e della filosofia del linguaggio *sperimentali*. Con ciò cambiano anche i termini della classica questione circa il rapporto fra logica e psicologia. Una volta si diceva: c'è uno iato incolmabile tra lo status *de iure*, normativo, della logica, e lo status *de facto*, descrittivo, della psicologia. È difficile continuare a dirlo quando l'aspetto normativo entra direttamente nella progettazione dei sistemi di simulazione e l'aspetto descrittivo entra nella valutazione dell'efficacia di tali sistemi. L'IA ha fatto sì che le questioni *de iure* sulla conoscenza umana siano diventate indagabili mediante questioni *de facto* su modelli computazionali.

Anzi, potrebbero non esserci reali situazioni psicologiche in grado di richiedere l'impiego di alcune strutture logiche, epistemiche e semantiche – e per motivi etici non possiamo sperimentare sull'uomo. Il computer può allora rivelarsi *indispensabile* per far passare le dottrine filosofiche dallo stato di idealizzazioni astratte allo stato di modelli per una scienza sperimentale virtuale. Infatti, oltre che da banco di prova per testare teorie del significato e della conoscenza, l'IA funge da *generatore di metodologie* cognitive, tramite le quali apprezzare la specificità delle risorse usate dagli esseri umani (o da altri organismi).

Per esempio, la distinzione tradizionale fra sapere-come e sapere-che non dice nulla su come si realizzi una loro efficace integrazione. Né le dottrine semiotiche né la filosofia analitica del linguaggio che ha puntato sugli strumenti della logica sono riuscite a spiegare in maniera soddisfacente quest'integrazione. S'è cominciato a fare dei passi avanti allorché ci si è posti il problema di quale linguaggio adoperare per gestire al meglio la rappresentazione della conoscenza in un sistema artificiale – un linguaggio 'procedurale' o un linguaggio 'dichiarativo'?

Teorie e problemi di filosofia della scienza si manifestano allora come teorie e problemi di *ingegneria della conoscenza*. Le questioni relative alla conferma di un'ipotesi possono essere impostate per via simulativa;

e la stessa *formazione* di una teoria è simulabile mediante modelli computazionali (d'apprendimento e d'euristica), così come la formazione delle galassie o la formazione di cicli stabili in un sistema chimico. Del resto, l'analogia tra sviluppo della competenza del bambino in un dominio cognitivo e formazione di una teoria ha sollecitato l'impiego dell'informatica in psicologia (le teorie psicologiche come *meta*-teorie della cognizione). E ne è seguito un incremento di rigore nei modelli dello sviluppo mentale. Per essere implementata su calcolatore, infatti, una teoria deve prima essere assiomatizzata – col che si ottiene già un chiarimento di quel che essa *dice* – ma poi si deve anche elaborare un modello (parziale) della teoria *come programma*. Che il programma funzioni non garantisce la coerenza della teoria, ma se il programma non gira bene, ciò è una spia che nella teoria qualcosa non quadra. Abbiamo così un modo abbreviato per trovare quali tipi di situazioni possono rivelare la falsità della teoria — relativamente, beninteso, alle caratteristiche del dominio inserite nel modello. In particolare, la teoria che sta alla base di un modello dell'apprendimento può essere studiata costruendo sistemi intelligenti che godono delle caratteristiche teorizzate. Perciò potremo trovare programmi che incorporano principi empiristi o razionalisti, innatisti o comportamentisti.

In questo senso un sistema di IA risulta *filosoficamente* economico: è un veloce generatore di controesempi oltre che un generatore di metodologie. La codifica di una teoria in forma di programma permette di testare anticipatamente le previsioni della teoria, sondare *quali cambiamenti teorici* possono eliminare i controesempi e valutare l'adeguatezza di ciascuno di questi cambiamenti. Tutto, naturalmente, dipenderà dalle risorse del linguaggio scelto, da che cosa è stato considerato essenziale alla teoria, cioè tale da ritrovarlo in *qualunque* suo modello, e da quale "fetta" del mondo reale è stata rappresentata in uno specifico modello. Le opzioni al riguardo non sono date a priori. Se alla fine i conti non tornano, la colpa potrebbe non essere né del programma né della teoria ma di un errore di valutazione da parte del progettista. D'altronde, anche queste valutazioni possono essere riviste e corrette, passando da un circolo vizioso a una spirale virtuosa.

È opportuno ripeterlo: i calcolatori sanno svolgere bene compiti che troviamo difficili e non sanno svolgere bene compiti che troviamo facili. È proprio questo fatto *curioso* che ci aiuta a capire l'architettura della mente umana, più di quanto ci abbiano aiutato a capirla le teorie della conoscenza e del linguaggio elaborate in passato dai filosofi. L'insuccesso dei vari tentativi di traduzione automatica ha messo in chiaro i fattori contestuali da tener presenti; e la chiusura dei micromondi ha mostrato che il mondo reale non può essere facilmente scomposto in frammenti tra loro indipendenti. Direte: forse ci si poteva arrivare anche senza i

computer. Sì, ma ora ne abbiamo una prova sperimentale. In più abbiamo capito con precisione *dove* nascono i guai e ci siamo fatti un'idea meno vaga circa la comune intelligenza pratica, suggerendo le domande alle quali l'IA è chiamata a rispondere.

Questo fatto *curioso* è spesso visto come un paradosso: i calcolatori progettati per simulare le nostre capacità sono utili solo per capire che siamo diversi da loro e per capire *quanto* siamo diversi! Non direi che è un paradosso e non mi sembra il caso di trarne conclusioni *di principio* (tanto meno, se olistiche). Senza dubbio, i programmi finora elaborati nel campo dell'IA al fine di dar conto dell'intelligenza *naturale* hano numerosi difetti. Siamo legittimati a inferirne l'*impossibilità* di simulare in maniera efficace le capacità cognitive? Prima di rispondere a tale quesito conviene indicare alcune differenze di struttura formale tra menti e macchine.

I linguaggi naturali in cui trova espressione la nostra attività di pensiero sono diversi dai linguaggi formali dei logici e dai linguaggi di programmazione che gli informatici usano per far funzionare un computer. Ci sono, però, alcuni aspetti basilari dell'architettura di un linguaggio e della sua semantica che fanno da ponte. Come avevano già capito Gödel e Turing, i programmi sono a loro volta configurazioni simboliche, perciò si possono trattare come oggetti, anzi: come oggetti strutturati in forma di tipi di dati (*data types*). Se quest'idea è stata all'origine della diffusione del LISP come linguaggio di programmazione funzionale, l'esigenza di sistemare in maniera rigorosa l'impiego del LISP e dei suoi nipoti ha riportato all'attenzione il linguaggio della 'teoria dei tipi', nata con Russell all'inizio del Novecento. Che cosa mancava a quella teoria? Innanzitutto, la possibilità di avere più di un tipo *base* di dati (per esempio, INT per gli interi e BOOL per i valori di verità), poi la considerazione delle *funzioni basilari* tra un tipo e l'altro, e infine il fatto che la loro totalità può parimenti diventare uno speciale oggetto del discorso. Ebbene, la semantica per linguaggi che colmano queste lacune è formalizzabile in termini di una recente branca della matematica, la *teoria delle categorie*, la quale raffina in misura significativa le risorse della teoria degli insiemi. Si possono trattare i programmi come *oggetti* in opportune categorie. (L'odierna didattica della matematica è tutta centrata sull'insiemistica, quindi non c'è da stupirsi che quanto ho appena detto sia poco apprezzato dagli stessi informatici).

Che un programma possa diventare oggetto di se stesso è decisivo per i modelli elaborati all'interno dell'area che prende il nome di *knowledge representation*, proprio perché mettono in evidenza il ruolo dell'autoriferimento e del ragionamento auto-epistemico. A partire dalla problematica relativa al duplice 'Teorema di incompletezza' di Gödel, il contributo della logica a questo riguardo è stato fondamentale. La filo-

sofia della mente se n'è servita a man basse: la logica avrebbe finalmente dimostrato i limiti della razionalità, nel senso che ogni progetto mirante ad assiomatizzare la razionalità era un sogno (o un incubo) *irrealizzabile*. Non starò a elencare i motivi che si possono addurre per legittimare questa morale – e non si riducono al solo teorema di Gödel. Piuttosto, vorrei ricordare (1) che la conoscenza dell'*impossibilità* di qualcosa è una conoscenza positiva e dovrebbe suggerire rispetto per ciò che *si può* fare con strumenti limitati (la termodinamica ci ha fatto capire che il moto perpetuo non è possibile e ce ne siamo serviti per costruire frigoriferi); (2) che il modo in cui viene solitamente letto il teorema di Gödel dipende da presupposti che non siamo obbligati a condividere.

6. *IA e logica*

Ci sono diversi temi in relazione ai quali si è determinato un fecondo commercio di idee fra IA e logica. Uno dei primi a essere affrontato è quello che riguarda, oltre all'automazione del controllo sulla correttezza di un'inferenza, l'automazione delle dimostrazioni di risultati già noti (conseguiti da esseri umani) e l'uso del calcolatore come strumento per ottenere nuovi teoremi. Forse l'esempio più famoso è costituito dal 'Teorema dei quattro colori', che fu ottenuto nel 1976 grazie appunto all'impiego di un calcolatore cui si demandò l'incarico di vagliare un numero enorme di casi, alcuni dei quali particolarmente complicati, al fine di poter stabilire che ogni mappa piana, opportunamente suddivisa in regioni può essere colorata con non più di quattro colori (regioni distinte adiacenti devono avere colori diversi). Questo ricorso al computer ha sollevato un'accesa polemica: fino a che punto potremmo fidarci del computer qualora non fossimo concretamente capaci di controllare, passo dopo passo, il lavoro deduttivo da esso svolto? È ancora una *dimostrazione* un processo inferenziale che non siamo in grado di fare nostro?

Se già queste domande ci costringono a tematizzare ulteriormente le proprietà generali che ascriviamo di norma al 'dimostrare', l'IA ha anche contribuito a evidenziare caratteristiche specifiche del ragionamento umano. Infatti, la necessità di servirsi di assegnazioni per default in sistemi di IA ha condotto a riconoscere l'importanza di una logica dell'*a meno che*, che sia non-monotòna (da $p \rightarrow r$ non scende più che $p \wedge q \rightarrow r$) e autoepistemica.

Inoltre l'IA ha fornito nuove motivazioni teoriche ed applicazioni concrete per la teoria, introdotta da Lofti Zadeh negli anni Sessanta, che tratta gli insiemi *fuzzy* (sfumati) e la relativa logica. La *fuzzy logic* tiene conto della vaghezza dei concetti mediante valori di verità intermedi nell'intervallo reale tra 1 (vero) e 0 (falso), dunque in maniera molto diversa dalle valutazioni probabilistiche. In effetti, già in comuni compiti di categorizzazione, la distinzione degli elementi x di un dominio tra

quelli che sicuramente hanno una proprietà *P,* cioè $[x \in P] = 1$, e quelli che sicuramente non ce l'hanno, cioè $[x \in P] = 0$, non sempre è così netta, e neppure corrisponde a una distribuzione di probabilità: il valore di $x \in P$ varia piuttosto nell'intervallo unitario in maniera più o meno continua. Non che i raffinamenti all'analisi del linguaggio scaturiti da queste (e altre consimili) linee di ricerca ispirate alla logica siano la soluzione di tutti i problemi inferenziali in situazioni di incertezza e di variabilità dei dati. Anzi, in molti ricercatori si è fatta strada la convinzione che il privilegio accordato a un'architettura puramente logico-linguistica sia manchevole sotto più rispetti. L'equazione pensiero = linguaggio, fatta propria da quasi tutte le varianti di cognitivismo e poi passata direttamente nell'IA "classica" sta ormai mostrando i suoi limiti.

Il *frame problem* riguarda quali cambiamenti sono provocati da un'azione e cosa invece resta come prima. La via comunemente seguita per risolverlo consiste nel mettere a punto un modello dinamico e, poiché si ha di mira la simulazione dei contesti pratici relativi all'uso del linguaggio naturale, ci si ritrova a fare i conti con la fisica ingenua. Nel 1978 Pat Hayes fece circolare un ambizioso 'manifesto della fisica ingenua' in cui esponeva il progetto di formalizzare tutte le conoscenze proprie del senso comune circa l'ambiente fisico, usando come linguaggio quello della logica classica del primo ordine. L'IA avrebbe così catturato l'*ontologia* del senso comune e l'avrebbe anche rigorizzata grazie alla semantica tarskiana: il *background* da cui tutti i vari micromondi traggono il loro senso avrebbe finalmente trovato una codifica appropriata.

Questo progetto ha assorbito le energie di molti ricercatori. Dag Lenat si è impegnato a dotare un computer di uno dei più giganteschi *data base* mai immaginati, concernente i più familiari domini, dalla meccanica dei macro-oggetti ai fatti storici, dallo sport alla letteratura, mettendo a punto il sistema CYC (da En-CYC-lopedia). Come dicevo, oggi si ha la sensazione che questo tipo di progetti sia giunto a un'*impasse* (semmai, si cerca di ricostruire ontologie dedicate). McDermott, che ne era stato un convinto promotore, ha ammesso il fallimento, scorgendovi la nemesi del logicismo iniziato da Frege e Russell e poi diventato abito mentale di molti filosofi analitici. Anche se, naturalmente, la logica non è il logicismo e di linguaggi logici non c'è solo quello più sfruttato in IA (il linguaggio del primo ordine), questa novella 'critica della ragione pura' si estende anche a varianti più sofisticate di logicismo. Be', se i ragionamenti concreti delle persone non trovano una resa adeguata nella logica, l'IA prenderà un'altra strada. Ma non è così facile, perché i limiti delle prospettive centrate sull'analisi logica sono anche limiti delle prospettive 'cognitiviste', che trattano il pensiero come un linguaggio.

Senza una ben definita integrazione fra strutture epistemiche di tipo diverso (di percezione e d'azione) che tenga conto della genesi di tali

strutture in un sistema aperto (che interagisce con un ambiente non meno strutturato), è difficile credere che l'IA possa offrire qualcosa di più che strumenti ausiliari, fortemente vincolati a un prefissato dominio.

Se queste relazioni strutturali non sono primariamente di carattere logico-deduttivo, allora tutto si fonda su qualche forma d'induzione? Neppure. Una volta versato il bicchier d'acqua sulla sedia, il bambino *non deduce* che tutto è rimasto com'era (non può più bere dal bicchiere l'acqua versata), né che tutto è cambiato (se ha sete, andrà a cercare la bottiglia d'acqua dove l'aveva lasciata). Se dovesse dedurre prima di agire, non gli basterebbe tutto lo spazio di memoria pur reso disponibile dai suoi miliardi di neuroni per codificare i presupposti dell'inferenza, eseguirla in tempo utile e comportarsi in conformità alla conclusione. Un'induzione sarebbe di nuovo un ragionamento (probabilistico) non meno complesso di una deduzione. Morale: non basta aggiornare la logica (negazione per default e simili) né arricchire la deduzione con qualcosa di pur sempre inferenziale (abduzione e simili). Possiamo quindi fare a meno della logica? Neanche. È solo che, lasciata a sé, la logica gira a vuoto. La revisione di una credenza non è né puramente logica né indipendente dalla logica. I fattori che contano dipendono dalla struttura assegnata agli oggetti, agli eventi e alle interazioni con essi. Quest'assegnazione non è né arbitraria né determinata dalla logica, ma va esplicitata per prima.

Molti hanno erroneamente creduto che le conoscenze del senso comune siano rappresentabili e regimentabili come fossero una (immensa) *teoria* e che una teoria sia nient'altro che un sistema linguistico: da ciò sono dipesi alcuni insuccessi dell'IA classica ma la radice dell'errore sta già nella deificazione del linguaggio celebrata nella filosofia del Novecento. Anche il "logicismo" in IA non è che una forma particolare di funzionalismo e il funzionalismo non è che una variante del formalismo, così come inteso in filosofia della matematica. C'è chi ha detto che il modo in cui ci rappresentiamo la nostra attività di pensiero e i suoi più stabili prodotti in formato simbolico non è che una meravigliosa realtà virtuale. Può darsi, ma l'efficacia di questa o di qualunque altra realtà virtuale è parassitica rispetto a preesistenti schemi della spazialità che sono oggetto della topologia e riguardano configurazioni *dinamiche*.

Non c'è modo di simulare una teoria del contenuto senza un modello del processo. Un'ontologia senza una dinamica non funziona. E ciò vale anche quando ontologia e dinamica sono riferite a quegli oggetti che chiamiamo 'proposizioni'. Per superare i difetti della scollatura che invece persiste in buona parte dell'IA attuale, penso che la progettazione di un sistema intelligente debba correlare le risorse logiche agli invarianti tra un dominio e l'altro, trattando le proposizioni come un sottodominio immerso in ciascun dominio-ambiente i cui stati esse codificano.

Rappresentiamo la realtà (e la finzione) rappresentando anche le conoscenze che ne abbiamo. E quest'autorappresentazione bene o male funziona perché non è una rappresentazione statica di conoscenze statiche. Solo tenendo conto di ciò, il gap tra sistemi naturali e artificiali potrà essere ridotto. Ma tenerne conto non basta. Un concetto, così come ogni oggetto simbolico, non è contrapposto agli oggetti che rappresenta; e le procedure di manipolazione di simboli non sono che schemi di manipolazione di oggetti. Le potenzialità di un'analisi costruttiva e variazionale, come quella resa possibile dalla teoria delle categorie, possono essere apprezzate, se non per altro, almeno per l'integrazione che questa consente di ottenere fra aspetti denotazionali e procedurali dell'*information processing*. Se c'è una sintassi che possiamo comprendere, è quella che deriva per astrazione dagli schemi topologico-dinamici della corporeità. Ma allora la mente non può più essere separata nettamente dal corpo.

Tutto questo è detto senza staccarsi definitivamente dal piano 'rappresentazionale'. C'è invece una linea alternativa di pensiero, legata al connessionismo, che se ne stacca in maniera ben più decisa.

7. *Reti neurali*

Il connessionismo è un tipo di 'filosofia' cresciuto con il progressivo potenziamento delle reti neurali artificiali, a partire dalla metà degli anni Ottanta. Una rete del genere è composta da un insieme di nodi interconnessi (o unità) ciascuno dei quali ha una soglia d'attivazione; inoltre, se un nodo *a* è connesso a un nodo *b* da un *link*, questo è unidirezionale, è unico, e ha associato un *peso* che misura il contributo dell'attivazione di *a* all'attivazione di *b*. I nodi iniziali di ogni sequenza di link sono quelli sensibili a segnali di input esterni alla rete; quelli finali forniscono l'output della rete.

A parte le analogie della prima cibernetica fra auto-organizzazione del cervello e sistemi basati su feedback, il primo modello si può far risalire agli assembramenti (o assemblee) cellulari, studiati da Donald Hebb nell'immediato secondo dopoguerra. Partendo dall'idea secondo cui nell'apprendimento si viene selezionando una opportuna matrice di pesi, una prima regola di selezione è: se le unità A e B sono eccitate simultaneamente, la forza di connessione tra A e B aumenta. La "Regola di Hebb" stabilisce, appunto, che se il prodotto delle forze (pesi) è positivo la connessione diventa eccitatoria, e se è negativo inibitoria.

I primi tentativi in questa direzione si erano scontrati col fatto che, se una rete ha solo due strati di unità (dunque, si riduce a unità di input e di output) i tipi di 'computazione' che riesce a eseguire sono nettamente inferiori a quelli consentiti da un'architettura sequenziale classica. La scienza cognitiva d'ispirazione funzionalista aveva dunque buon gioco.

Tuttavia, da quando si è cominciato a costruire reti con numerosi strati intermedi di unità 'nascoste', capaci di un massiccio parallelismo, il quadro è cambiato. Il principale modello di reti neurali è stato delineato dal gruppo di ricerca *PDP,* che deve il suo nome alle iniziali di *Partial Distributed Processing*.

Indubbiamente, una rete neurale somiglia al reale funzionamento del cervello più di quanto gli somigli una macchina di Turing. Anche i neuroni hanno una soglia di attivazione; analogamente alle sinapsi, le connessioni tra nodi possono essere eccitatorie o inibitorie; e una rete, a differenza di un programma di computer, è plastica, proprio come il cervello. Inoltre, ammessa una sufficiente ridondanza, il carattere distribuito della codifica compensa meglio eventuali danni subiti da singoli nodi. Il pattern complessivo di attivazione di una rete è descrivibile come composizione funzionale a partire dallo stato di attivazione di ciascun nodo e la risultante è tipicamente espressa da una funzione non lineare. La flessibilità di una rete può essere sfruttata mediante algoritmi di learning, supervisionato o no.

Nel primo caso, alla rete si forniscono esempi del pattern sul quale sintonizzarsi e via via si correggono, in base all'algoritmo dato, i pesi sui link in modo che l'output converga a quello 'giusto'. Il più diffuso di questi algoritmi è quello di *backpropagation*, introdotto da Paul Werbos nel 1974, che riduce progressivamente lo scarto dell'errore nell'output partendo dalle unità dello strato finale e risalendo all'indietro di strato in strato.

Operando correzioni che si propagano in senso inverso a quello dei segnali nella rete, quest'algoritmo retrogrado ha l'inconveniente di ricorrere a un istruttore esterno. Si potrebbe passare dall'esterno all'interno del sistema solo se la rete cui l'algoritmo si applica fosse una sottorete di una rete ricorsiva (con autoaccesso). A prescindere dalla difficoltà di riprodurre in questo modo gli stessi effetti dell'algoritmo retrogado, il guaio è che, per determinare l'errore, si è presupposto di sapere quale fosse il risultato 'giusto'. Ma l'intelligenza di noi esseri umani sta anche nelle capacità di autoapprendimento, le quali non sono una nostra risorsa esclusiva: sono noti esempi d'apprendimento autonomo in numerose specie. (Peraltro, i connessionisti hanno contribuito alla nascita di quel nuovo ambito di ricerca in IA che va sotto il nome di 'vita artificiale', in cui si può simulare un'ampia gamma di strategie composite di learning).

L'apprendimento non-supervisionato corrisponde a reti che si autoorganizzano: le regolarità su cui sintonizzarsi sono lasciate estrarre alla rete stessa. Perché ciò si realizzi, bisogna che l'ambiente sia carico d'informazione e che non sia progettato da un ... genio maligno – d'altronde, se così fosse, la rete come farebbe a sopravvivere? Gli algoritmi di questo tipo funzionano in base a un principio di coerenza per risonanza: unità adiacenti tendono a rinforzarsi o a inibirsi a vicenda.

Ora, questi vari modi di organizzare una rete neurale si prestano anche a simulare, più o meno bene, il comportamento simbolico sequenziale ma certo in una rete non c'è alcun insieme di *regole* (logiche) per elaborare *rappresentazioni* nel senso tradizionale. Si va piuttosto da un'architettura totalmente distribuita a una sempre più localizzata modularmente. Quanto alla variegata fenomenologia cui si presta la coppia locale/globale, i relativi modelli sono di natura geometrica, non logica; e per minimizzare l'errore non ci serve la classica teoria della computabilità ma l'ancor più classica Analisi.

8. *Dinamiche subsimboliche*

Ciascuna rete neurale si può descrivere come un sistema dinamico *S*, la cui traiettoria nello spazio delle fasi *F(S)* è determinata da una funzione evolutiva. I valori di questa funzione dipendono dallo stato di attivazione di ciascun nodo e dai pesi sui relativi link. Pertanto la dimensione di *F(S)* può essere enorme se la rete è costituita da un ampio numero di nodi, disposti in più strati e fortemente interconnessi. Anche se la dinamica che interessa è discréta (le transizioni di stato sono funzioni di un tempo discretizzato e tra due punti vicini in *F(S)* non c'è necessariamente un punto intermedio), si possono ottenere buone approssimazioni servendosi della matematica del continuo. La funzione evolutiva si esprime in una legge (solitamente non lineare) di moto in *F(S)*, ma ciò che più conta è che ogni punto (stato, fase) di questo spazio abbia un bacino d'attrazione. Un attrattore globale è un punto verso il quale tende l'intero sistema: se esiste, corrisponde alla stabilità strutturale del sistema – rispetto a eventuali perturbazioni delle soglie di attivazione e dei link. La topologia che ne risulta indotta su *F(S)* dà informazioni essenziali per una descrizione qualitativa dell'evoluzione che il sistema subisce nel tempo, nel corso dell'esecuzione di un compito cognitivo.

Una simile impostazione ha dato buoni risultati nell'area della pattern recognition, nella simulazione dei processi decisionali e in compiti di memorizzazione e categorizzazione. Il punto è se la dinamica di una rete possa catturare tutto ciò che descriviamo in termini di rappresentazioni simboliche e regole ricorsive per la manipolazione dei simboli. Può l'ordine logico emergere autonomamente dal funzionamento parallelo di un sistema i cui stati non sono disposti come quelli di una macchina di Turing? Sintassi e semantica possono essere ascritte a una rete in cui non c'è alcun vero simbolo?

Senza dubbio, è un notevole vantaggio teorico disporre di sistemi in grado di sviluppare da sé capacità di alto livello a partire da pochi schemi distribuiti, ma se poi non riusciamo a spiegare ciò per cui un filosofo

come Ernst Cassirer definì l'uomo come *animal symbolicum,* cioè, la multiforme e finissima organizzazione logico-linguistica dei concetti, il vantaggio non ci aiuta a progredire verso un sistema di IA paragonabile alla mente umana.

Alle precedenti domande Jerry Fodor e Zenon Pylyshyn hanno dato una risposta negativa in un articolo del 1988 che ha innescato una polemica dai toni molto accesi.

Fodor e Pylyshyn mettevano in dubbio due cose: (1) che il modello della cognizione basato sull'architettura delle reti neurali sia davvero diverso dal modello classico allorché si propone come *esplicativo* della competenza sintattica e semantica, e (2) che, se è davvero diverso, non riesce a spiegare quel che deve spiegare. La forma del loro ragionamento è dunque quella di un sillogismo disgiuntivo: o il connessionismo non differisce dall'approccio classico (cognitivista) se non nel modo di implementare i programmi, oppure ne differisce. Nel primo caso non è una vera alternativa, nel secondo lo è ma ripropone un associazionismo ormai superato. Ergo, il connessionismo non può essere quella rivoluzione che i suoi profeti vogliono farci credere. (NB: talvolta si legge che il contrasto fra IA classica e connessionismo è quello tra un'architettura top-down e una bottom-up, ma proprio le obiezioni di Fodor e Pylyshyn fanno capire che ciò è improprio, se solo si riflette che la prima è, in buona parte, composizionale, la seconda no).

I connessionisti hanno contestato questo sillogismo disgiuntivo dicendo che le reti neurali possono avere un'architettura molto più ricca del vecchio associazionismo e non implementano soltanto programmi di macchine 'classiche' pur essendo in grado di fornire prestazioni equivalenti. In particolare, Smolensky ha replicato che una rete neurale può reggere il peso di una struttura combinatoria, in cui un intero sia funzione delle parti che lo compongono, benché la causalità stia tutta nell'attivarsi o no dei nodi della rete, e non nelle 'rappresentazioni' che attribuiamo al suo comportamento. Per i cognitivisti (e per l'IA classica) le rappresentazioni sono simboli formali e i processi cognitivi sono l'esecuzione di programmi, per i connessionisti le rappresentazioni sono tutt'al più reinterpretabili come vettori che specificano parzialmente lo stato di un sistema dinamico, e i processi cognitivi sono l'evoluzione di un sistema fisico governata da equazioni del tipo di quelle che governano flussi e campi.

È prematuro dire se, *in linea di principio,* i tipi di computazioni eseguibili da una rete connessionista, che si evolve in base a principi dinamici (e probabilistici), siano o no più estesi dei tipi di computazioni eseguibili da una macchina di Turing. Fatto sta che in fisica sono stati individuati sistemi dinamici la cui evoluzione non è Turing-computabile. Se si riuscisse a mostrare che tali sistemi sono adeguatamente simulabili

da una rete, la famosa Tesi di Church-Turing che, parlando alla buona, afferma che qualunque procedura meccanica è eseguibile da un calcolatore idealmente equivalente a una macchina di Turing, risulterebbe falsa. Sono stati proposti anche modelli ibridi, ove sono presenti, su piani diversi dell'architettura cognitiva, entrambi i tipi di computazioni. In California, da tempo il gruppo di ricerca di Feldman ha portato avanti modelli con quest'architettura doppia: ai livelli più bassi della *microcognizione* una struttura connessionista, ai livelli più alti della *macrocognizione* una sequenzialità di tipo classico.

Mentre le unità minime del cognitivista erano simboli semanticamente interpretabili, così non è per le unità del connessionista; la sintassi si colloca a un livello, la semantica a un altro, quindi non ci sono elementi che hanno simultaneamente un ruolo sintattico e semantico. Al paradigma "simbolico" classico si viene contrapponendo un paradigma "subsimbolico", in cui le unità d'elaborazione non sono localizzate: la codifica associata a un'unità di significato è distribuita nella rete. Come già accennato, ciò consente una maggiore robustezza della rete rispetto alla variabilità dei contesti d'uso dei 'simboli'. Il problema ancora irrisolto è se il *complesso* delle regole della grammatica e della logica sia riducibile a pattern d'attivazione e attrattori. Manca una descrizione matematica della struttura degli stati 'proposizionali atomici' e di come essi si combinino per dare stati 'molecolari'.

Pensiamo alla competenza visiva: com'è che emergono proprio certe *gestalt* stabili e non altre astrattamente possibili? Il fatto che una rete possa *autostrutturarsi* senza bisogno di predefiniti programmi non basta a selezionare gestalt stabili. Occorrono vincoli 'di campo' che restringano la gamma degli esiti possibili ed è ragionevole supporre che questi vincoli siano *fisiologici*. Ma se c'è una tale correlazione tra struttura materiale e funzione diventa arduo separare il software dallo hardware della mente.

Ammesso che la spiegazione sia analoga per la competenza semantica, l'integrazione tra reti neurali e meccanismi robotici dovrebbe portare a sistemi che possiedono una qualche, autonoma, semantica intrinseca. Tuttavia, per ottenere un risultato del genere non basta servirsi di reti, perché la forte non localizzazione è in contrasto con le prove a favore dell'autonomia reciproca dei diversi domini cognitivi. Se l'architettura di un sistema intelligente fosse un tutt'uno inseparabile, ogni singolo cambiamento in un nodo potrebbe generare una reazione a catena, con effetti devastanti sulla stabilità del sistema. L'adozione di un approccio parallelistico non spiega, da sé sola, come evitare un simile effetto valanga. L'impiego di modelli 'ibridi', in cui si combinano aspetti connessionisti e cognitivisti, migliora il quadro della situazione, ma è presto per dire se il miglioramento è decisivo.

Una misura dell'ottusità dei sistemi classici era l'esplosione combinatoria. Come si può evitare l'esplosione combinatoria senza pagarla col ritorno a meri riflessi condizionati? Come ottenere robustezza e possibilità di una *combinatoria*?

Si dice: un pattern d'attivazione può anche essere ciò che chiamiamo 'rappresentazione', ma, di per sé, ogni suo componente non rappresenta nulla. E si dice: un programma è fatto di simboli, giù giù fino ai simboli elementari (non importa per cosa stiano). L'emergere di un simbolo da qualcosa che *non lo è* costituisce un problema solo per il connessionismo. Ma questo problema, invece di essere un difetto, non è forse un pregio? Non dimentichiamo che né le parti di una A sono lettere dell'alfabeto, né l'attività di camminare è scomponibile in cose che camminano. Il ginocchio si flette, il bacino si solleva, la pianta del piede ruota. Le singole molecole d'acqua non sono 'liquide', gli archi a volta sono fatti di mattoncini diritti, e potremmo continuare all'infinito.

La differenza tra una rete che funziona a partire da microcaratteristiche e una che funziona in modo totalmente distribuito è una differenza di grado di interconnessione. È analoga alla differenza tra un mosaico e un ologramma. Abbiamo una teoria che ci faccia capire cosa succede via via che giriamo la manopola dall'atomismo alla globalità? No. Quindi, invece di continuare a dibattere, sarebbe più proficuo impegnarci a costruire una teoria del genere.

9. *Fuori dalla stanza cinese*

Pensieri, concetti, significati, sono entità materiali? Scommetto che direte: «No: infatti non hanno né massa né volume né posizione». Il connessionista che si preoccupa di rispondere a domande cognitive in termini, diciamo, subsimbolici è o non è alleato del materialista che vuol ridurre la cognizione umana a scariche di neuroni e flussi di ioni sodio e potassio?

Alcuni dei più accesi sostenitori di una filosofia materialista della mente hanno visto nel connessionismo un alleato. I Churchland hanno avanzato una teoria neurocomputazionale della mente che si propone di eliminare dal linguaggio scientifico tutti i fossili mentalisti legati alla psicologia del senso comune: in questo modo non si può più dire che *credere* qualcosa ha effetti causali, perché nel mondo fisico non esistono entità come le credenze.

Ma una rete neurale non è forse tanto formale quanto un diagramma di flusso? Ci sono forse differenze *di principio* tra reti artificiali e reti naturali? Non potrebbe servirsi delle reti neurali anche un funzionalista? Come potete immaginare, ci sono diverse opinioni al riguardo, suffragate con argomenti poco concludenti in un senso o nell'altro.

Alcuni, come Domenico Parisi, vedono nell'«autostimolazione interna» di una rete una condizione fondamentale perché emergano cognizioni complesse. D'altro lato, l'idea che si possa capire la mente attraverso modelli artificiali, siano quelli dell'IA classica o quelli di reti connessioniste, è stata negata da Edelman, secondo il quale c'è un solo modo per arrivare a una spiegazione della mente ed è lo studio della fisiologia dell'«oggetto più complicato di tutto l'universo conosciuto»: il cervello umano. Attraverso la formazione e la selezione di particolari popolazioni di neuroni, il cervello diventa il supporto di innumerevoli *mappe neurali,* che non solo si aggiornano di continuo ma soggiacciono a una selezione che sfocia nella capacità di 'categorizzare'. Il materialismo che ne risulta è *sui generis*, perché non elimina la mente né la identifica con le unità che la compongono. Invece che una *cosa,* fa della mente un *processo* – ovviamente, un processo speciale, reso possibile da non meno speciali sottoprocessi. Ma allora l'unica vera IA si avrà assemblando veri neuroni.

Ora, Edelman è un neurobiologo. Ci sono stati molti filosofi che hanno manifestato un diverso tipo di scetticismo nei confronti dell'IA come progetto generale, ancor prima che nei confronti di questo o quel sistema di di IA.

In un articolo del 1980, John Searle ha contestato radicalmente l'ipotesi che un qualunque sistema di IA possa raggiungere la minima competenza semantica. Con ciò, Searle non intendeva negare l'utilità di programmi per computer come strumenti ausiliari alla simulazione di questo o quel processo cognitivo; piuttosto, intendeva negare che un computer, quantunque potente e ben programmato, possa mai arrivare a *comprendere* qualcosa, e intendeva negare che le simulazioni mediante programmi possano essere prese come *spiegazioni* delle capacità cognitive così simulate. (Il famoso, o famigerato, 'Argomento della Stanza Cinese' è ormai diventato un tema di conversazione nei salotti e quindi lo darò per noto).

Con quest'argomento Searle si propone dunque di stabilire la duplice tesi negativa che ho appena enunciato (NB: benché aspirasse alla massima generalità, l'argomento aveva come bersaglio il modello di Schank per la comprensione di testi). Searle non mette in dubbio che il Test di Turing sia superabile da un computer, bensì suppone che il Test sia stato superato da un sistema che esegue esclusivamente manipolazioni sintattiche, in conformità a un qualche programma prestabilito. Se, ciononostante, il sistema non capisce il significato dei messaggi (enunciati) in ingresso e in uscita, vuol dire che la sintassi di un programma non basta ad avere una semantica. (Searle aggiunge: non c'è neanche bisogno di un algoritmo di calcolo per averla, tant'è vero che già nel caso della sintassi contesta l'idea chomskiana di una grammatica uni-

versale). La mente umana ha invece accesso al significato e, se i nostri pensieri sono causati dal funzionamento del cervello, allora la simulazione artificiale delle reti nervose mediante procedure ricorsive non può spiegare i meccanismi causali che permettono al cervello di secernere pensieri – e la stessa cosa può dirsi per le reti connessionistiche, guidate da algoritmi di (auto)apprendimento, in quanto anch'esse puramente formali e dunque indipendenti dallo hardware. Ciò che conta sono gli speciali poteri causali di quello speciale hardware che consiste nella materia grigia.

Nel 1990 Searle ha precisato le ipotesi e le conclusioni dell'Argomento della Stanza Cinese nel modo che segue.

Assioma 1.	I programmi di calcolatore sono formali (sintattici)
Assioma 2.	La mente umana ha contenuti mentali (una semantica)
Assioma 3.	La sintassi non è condizione necessaria, né sufficiente, per la determinazione della semantica.
Assioma 4.	Il cervello umano causa la mente.

Conclusione 1.	I programmi non sono condizione necessaria, né sufficiente, perché sia data una mente.
Conclusione 2.	Qualunque sistema in grado di causare una mente deve possedere poteri causali almeno equivalenti a quelli del cervello umano.
Conclusione 3.	Qualunque sistema artificiale, se svolgesse soltanto un programma, non sarebbe in grado di produrre fenomeni mentali.
Conclusione 4.	Il modo in cui il cervello umano produce fenomeni mentali non si riduce allo svolgimento di un programma.

Se l'essenza di ciò che è autenticamente 'mentale' è la capacità di *riferirsi-a*, cioè, l'intenzionalità, e questa è preclusa ai sistemi di IA in virtù dei poteri causali del cervello, allora è chiaro che la descrizione di quest'essenza dovrebbe essere alquanto diversa da tutto ciò che Brentano, prima, e Husserl, poi, hanno scritto al riguardo, perché l'individuazione delle strutture 'pure' dell'intenzionalità secondo la linea della psicologia descrittiva, prima, e della fenomenologia, poi, comporta la sospensione di qualunque legame col sostrato fisico (fisiologico) delle strutture mentali.

L'argomento di Searle, beninteso, non vuol negare che il funzionamento concreto di un programma di computer abbia effetti causali. Il

punto è che i cervelli hanno i *giusti* poteri causali, mentre ogni altra cosa sia stata finora trovata per simularne i prodotti (di pensiero) non li ha. Ma allora l'argomento rischia di voler provare troppo. Infatti la relazione d'equivalenza 'avere gli stessi poteri causali di' è usata da Searle per raffinare in maniera decisiva la relazione d'equivalenza associata al Test di Turing, cioè quella di equiestensionalità (stessi input – stessi output). Ma ciò da un lato rischia di ridurre a un singoletto la classe d'equivalenza di sistemi in grado di aver accesso a una semantica – il singoletto è ovviamente composto dalla mente umana, con predarwiniano sciovinismo. Dall'altro, sembra sottostimare l'impegno teorico richiesto per adottare come metro una tanto fine relazione, svalutando a priori tutte le ricerche rivolte a indagare sistemi, naturali o artificiali, che grazie al loro stesso hardware (pur diverso da quello specifico di homo sapiens sapiens) presentano un qualche grado di comprensione. I giusti poteri causali non è detto in alcun modo che siano limitati alle proprietà dell'hardware cerebrale, perché queste stesse proprietà si definiscono attraverso le interazioni dell'intero corpo con aspetti percettivamente salienti dell'ambiente esterno.

Inoltre potremmo anche accogliere l'Argomento della Stanza Cinese e limitarci a ricavarne, come morale, che l'idea secondo cui un computer-robot appropriatamente programmato *sia* (*indiscernibile da*) una mente (ovvero *capisce*, ha stati *realmente* cognitivi, ecc.) non è ancora riuscita a trovare la via giusta. Quale potrebbe essere questa via? Forse non una riguardante l'ingegneria dei circuiti stampati, bensì la bio-ingegneria. E non è detto che si debba imporre a un cyborg una condizione così stretta come vorrebbe Searle. Il cervello di un delfino o quello di un elefante hanno sicuramente qualche 'giusto' potere causale – semmai non hanno abbastanza ... sintassi!

Vi ho appena suggerito che i 'giusti' poteri causali non sono solo quelli del cervello ma anche quelli relativi alle interazioni fra un organismo e il suo ambiente – come inteso da chi ci vive dentro. Su questo punto, le critiche di Searle erano state precedute da quelle di Dreyfus. Nel '72, infatti, Dreyfus pubblicò un libro il cui sottotitolo suonava: *A Critique of Artificial Reason*, con richiamo evidente al titolo della celebre opera di Kant, e ne scaturì un'accesa polemica ben prima che scoppiasse quella sull'argomento della stanza cinese.

Dreyfus riproponeva la tesi husserliana secondo cui l'attività intenzionale affonda le radici nel *mondo-della-vita* e in particolare si richiamava alla nozione di *orizzonte* intenzionale, affermando che le macchine (in quanto tali, non solo quelle finora costruite) ne sono prive, perché i programmi sono privi di aspettative e di anticipazioni guidate dal contesto pratico e dagli obiettivi vitali, come invece succede con *veri* sistemi intelligenti.

Se per Searle è il cervello che fa la differenza, per Dreyfus è il *corpo* (quello umano, beninteso) ciò che consente di avere intenzionalità – un'idea già sviluppata da un fenomenologo francese, Merleau-Ponty, negli anni '40. È dunque sul terreno delle esperienze vissute che andrebbe cercato ciò che permette all'uomo di *capire* senza dover formalizzare tutto, e di comportarsi in maniera appropriata alle più diverse situazioni senza bisogno di regole precise, come invece i computer. Da un lato le miracolose peculiarità della materia grigia, dall'altro la non meno miracolosa frattura tra natura e spirito, che si ottiene svincolando l'intenzionalità da qualunque meccanismo fisico-chimico.

Gli argomenti di Dreyfus e Searle puntano il dito su importanti deficienze dell'IA, non cancellate da *fuzzy logic*, connessionismo, ecc. Perfino Winograd si è pentito, ridimensionando le iniziali speranze e orientandosi verso una più modesta concezione, centrata sul miglioramento dell'interfaccia utente. Il guaio di questa ritirata non è tanto nelle potenzialità dell'IA che trascura, quanto nel tipo di filosofia che plaude a uno iato incolmabile tra intelligenza naturale e IA: è una filosofia d'ispirazione ermeneutica, che pregiudica la stessa possibilità di spiegare l'intelligenza *naturale*. Esagero? Che altro dire di una critica dell'IA che fa delle menti coscienti un tipo di entità tanto straniere al mondo fisico quanto lo sono i programmi per computer? Conviene, piuttosto, accennare alle obiezioni mosse all'argomento di Searle, al fine di suggerire una prospettiva diversa, in cui converge una serie di spunti già presenti nelle pagine che precedono.

10. *Immergere l'artificiale nel naturale*

Searle attacca l'idea che la cognizione sia manipolazione di simboli formali e strappa il nostro applauso. Purtroppo, quest'idea è tanto controintuitiva quanto l'idea che la sede dell'intenzionalità sia un certo ammasso gelatinoso di colore grigiastro. L'omino nella stanza cinese è assimilato a un sottosistema che manipola simboli, proprio come fa la CPU di un calcolatore. (Per inciso, si potrebbe contestare che la CPU manipoli simboli – sono simboli *per noi* – e poi un sistema effettivo di IA non si riduce alla sua eventuale CPU).

Una delle repliche all'argomento di Searle è stata: «*È tutto il sistema che comprende*». Cioè, anche se non l'omino, è *la stanza* che capisce il Cinese. Morale: la ragione per cui i computer non hanno i giusti poteri causali è che non sono integrati nel giusto tipo di sistema totale; l'omino nella stanza si comporta come un cervello in un corpo che non è il suo. Ma perché fermarsi alle pareti della stanza? Perché non dire che sul pianeta Terra c'è intenzionalità e il suo soggetto è … Gaia, come sistema unitario? Vi siete mai sentiti come un neurone della Terra? E perché poi fermarsi

alla Terra? Capite bene che, lungo questo pendìo scivoloso, s'arriva alla conclusione panteistica che l'Universo è l'unico vero sistema intelligente. (I computer non sono intelligenti? Allora sono proprio come noi).

Searle si affretta a replicare: non c'è motivo per cui tutto un sistema passi dalla sintassi alla semantica quando nessun suo singolo componente ci riesce. Gli è stato obiettato che, se è così, il suo argomento prova *troppo*: neanche noi esseri umani avremmo intenzionalità, perché nessuna parte di noi (in particolare il cervello coi suoi giusti poteri causali) fa qualcosa di più che correlare ingressi e uscite. Se l'evoluzione è arrivata a noi partendo dai protozoi, perché escludere che si possa arrivare a un sistema artificiale che capisce partendo dai computer e robot attuali?

Neanche se guardassimo dentro al nostro cervello, invece che dentro ai circuiti di un computer ci troveremmo alcun *significato*. Eppure diciamo di *capire*. Una soluzione è quella che suona così: «alla macchina manca un sistema senso-motorio». Ovviamente, capire come far veleggiare una barca contro vento è diverso da capire una spiegazione verbale dell'uso delle vele. Ma questo, *mutatis mutandis*, non depone contro l'IA. E se mettessimo programmi (di stampo classico o a rete) in un robot, lasciato libero di interagire col mondo? Dopotutto, il problema non è *se una macchina può pensare*, ma se può pensare un computer fatto come quelli attuali, tant'è vero che Searle dice a chiare lettere: *noi siamo macchine (biologiche) e pensiamo*, il computer no. E che cos'hanno le macchine biologiche che i computer digitali non hanno? I giusti poteri causali? Ma abbiamo già notato che il cervello non basta e che ci vuole dell'altro.

Da un sistema davvero intelligente ci aspetteremmo libertà e creatività. Se le leggi della biologia da cui siamo programmati come macchine viventi lasciano spazio per libertà e creatività, perché non potremmo programmare un computer capace di arrivarci da sé? Anche se i giusti poteri causali fossero tali da renderci diversi da automi biologici (zombie), non basterebbero a escludere che ci troviamo in una situazione come quella descritta nel film *Matrix*.

Tolta di mezzo ogni manipolazione di simboli (elaborazione sequenziale di informazioni) dal cervello, Searle riesce a provare davvero che l'*unico* hardware possibile per un sistema che sia legittimo dire dotato di intenzionalità è qualcosa come un *cervello*? Se rispondete di sì allora potete anche dire che un rettile, un pesce e un uccello hanno una minima forma di proto-intenzionalità (hanno un sistema nervoso centrale che, per qualche aspetto, è simile al nostro) mentre una medusa, un termostato e una colonia di formiche non ce l'hanno. Se rispondete di no, allora, con McCarthy, potete affermare che anche i termostati hanno dei pensieri (credenze, stati intenzionali). Quali? I tre seguenti: qui fa troppo caldo/qui fa troppo freddo/qui va bene. Ha ragione Searle o chi accoglie i termostati tra le cose (minimamente) intelligenti?

Non che ci dispiaccia pensare che alcune proprietà degli organismi *non potranno mai* essere riprodotte da un computer, a meno che sia anch'esso un organismo. Ma non conviene assecondare troppo il nostro narcisismo. Fino al Novecento si sarebbe potuto dire che solo oggetti fatti come gli uccelli avrebbero mai volato: il volo è un fenomeno biologico. L'evidenza (ben più rumorosa) è sotto gli occhi di tutti. Pensiero, mente, intelligenza, intenzionalità, significato sono termini terribilmente carichi di vaghezza e ambiguità. Per evitarle occorre fare scelte teoriche la cui efficacia si può misurare solo con le spiegazioni che ne ricaviamo. Proprio a questo scopo potrebbero servirci (e ci servono) gli esperimenti con sistemi di IA.

L'importanza dell'hardware biologico era stata sottostimata; ora si rischia di sovrastimarla. Se per parlare di "sistema intelligente" è rilevante il *modo* in cui i componenti locali di un'architettura funzionale sono integrati in un tutto complesso, allora le differenze tra mente e computer dipendono *anche* dall'architettura; il fatto che questa sia sequenziale o parallela, digitale o analogica, localizzata o distribuita, modulare od olistica, *non* è irrilevante. La neurofisiologia ha messo in evidenza la compresenza di tutti i fattori appena elencati, nel cervello non è stata trovata alcuna CPU e si rafforza la convinzione che non è stata trovata perché non c'è. L'esistenza di ben definite gerarchie nei livelli d'organizzazione nelle cellule (per esempio, quelle della corteccia visiva) non prova il contrario: non c'è una *centrale* che gestisce il buon funzionamento di tutti i 'reparti', bensì un controllo reciproco dei reparti tra loro, un po' come un sistema democratico basato sulla divisione dei poteri (la mente come società). Eppure, quando ciascuno di noi dice 'io' si comporta *come se* una tale centrale ci fosse.

Nel cervello la divisione tra cambiamenti nel software e cambiamenti nello hardware è ben più difficile che in un computer. Lo sviluppo neurale (in particolare, la differenziazione delle aree nella prima infanzia) e lo sviluppo cognitivo vanno di pari passo. Questa correlazione può essere molto più rilevante, per la *struttura* della mente, di quanto si è creduto; e se è così, il funzionalismo si trova davvero in brutte acque. D'altra parte, un'architettura parallela è non meno formale di una sequenziale. In entrambe non c'è un quartier generale che gestisce tutti i reparti cognitivi, e allora da dove vien fuori la coscienza?

Molti filosofi si dilettano ancora a dare risposte definitive a domande per le quali i dati empirici e le teorie a disposizione sono insufficienti. Perché mai indulgere in questo diletto? Eppure, in quanto precede non si son fatta strada solo ipotesi in negativo: vi ho suggerito che l'intelligenza semantica si formi attraverso le *interazioni*, biologicamente prefigurate, del corpo con l'esterno, e che, una volta così instaurato, il significato possa aprirsi a situazioni controfattuali (che riguardano anche le sue mo-

dalità d'instaurazione!) e tutto ciò grazie al potenziale *computazionale* della mente: una specie di 'pollice del Panda' cognitivo.

Questo suggerimento rientra nell'ottica del 'naturalismo', ma un naturalismo sui generis, dato che si *auto-sospende* e quindi diverge dal realismo di tutti coloro che restano legati a una rigida teoria causale del significato.

Per contro, il serissimo appello dei fenomenologi al *conferimento di senso* non è che un caso di quella che Hofstadter ha chiamato scherzosamente la teoria 'juke-box' del significato, contro la quale svolge un ragionamento che, nella sua semplicità, orienta a cercare la risposta alle tante domande precedenti in una direzione naturalista. La teoria 'juke-box' del significato consiste nel dire che nessun messaggio contiene un significato intrinseco e che per *capire* un messaggio, come appunto la musica incisa su un disco di vinile nel juke-box, occorre e basta aggiungere l'informazione contenuta nel juke-box (sistema, teoria, struttura, ambiente, o come altro volete chiamare la totalità che fa da supporto al messaggio). Solo che, per capire che cosa significa l'informazione contenuta nel juke-box, ci vuole un altro juke-box, ben più vasto. E così via, in una serie illimitata di bambole russe.

Se una IA che intenda colmare le attuali alcune richiede un modello adeguato dell'ambiente e se tale modello non può essere computazionale perché il mondo fisico ha caratteristiche che trascendono la computabilità, il discorso non si chiude qui? Non è detto, perché se siamo esseri coscienti e abbastanza intelligenti da capire, prima o poi, le condizioni che rendono possibile la coscienza, potremo riprodurle artificialmente. Penrose ha sostenuto che la stessa matematica richiesta per capire l'universo quantistico fornisce la prova di una simile non-computabilità, ma il suo argomento dipende da una particolare *filosofia* della matematica, che può non esser condivisa.

Non dimentichiamo: noi esseri umani seguiamo tante *regole* nell'esercizio delle capacità cognitive e le seguiamo con successo, senza esserne minimamente coscienti. Impariamo a camminare prima di sapere la fisica delle leve e la chimica dei muscoli. La teoria del juke-box si scorda che la nostra intelligenza risiede in oggetti fisici relativamente autonomi: i nostri cervelli, i corpi in cui si trovano, i quali lavorano senza che si dica loro di lavorare. La capacità di riconoscere qualcosa come un messaggio fa parte della dotazione innata del cervello; il nostro hardware costitutivo, biologico, è l'ultimo juke-box semantico. Ma il punto è che questo hardware non si riduce al cervello e ai suoi giusti poteri causali. Né è detto che il *nostro* corpo sia l'unico tipo possibile di juke-box.

Così, quando Hofstadter scarica, come Searle, il peso del significato sulla giungla di neuroni corre di nuovo il rischio di sciovinismo specie-specifico; e allora l'IA si riduce a un'astuta impresa commerciale. Preferi-

sco pensare che ciò che permette di *intelligere* siano strutture presenti anhce nell'accoppiamento di altri sistemi fisici col loro ambiente. Sono le stesse leggi fisiche che, così come 'suonate' dall'universo, favoriscono l'emergere della complessità e in particolare l'emergere dell'intelligenza: la sintonia fine delle forze fondamentali alleva conoscitori di esse. Non ogni tipo di materia può evolversi in qualcosa di *significante* (banalmente, perché quest'evoluzione si verifichi, occorre che si diano certe condizioni al contorno, proprio come per ogni altro sistema dinamico) e non ogni forma di intenzionalità è legata a un particolare mammifero terrestre.

Andrew Brooks ha parlato di "intelligence without representation". La sua proposta è apprezzabile in quanto, provocatoriamente, ci ricorda che i sistemi cognitivi sono aperti all'ambiente e inseparabili da strutture della corporeità. Tuttavia nessuno dei suoi meravigliosi insetti è in grado di servirsi di una pur minima sintassi simbolica. E d'altra parte la Stanza Cinese ci ricorda che la sintassi non genera alcuna semantica, se non in senso pickwickiano. Come allora garantire il radicamento (*grounding*) dei simboli in qualcosa di corporeo che dia loro significato, evitando lo sciovinismo specie-specifico? Il progetto che è oggi sviluppato da vari gruppi di ricerca punta su un modello del radicamento in termini di cicli percezione-azione che si stabilizzano in macro-schemi cinestetici. Ne risulta un tipo di naturalismo sui generis, che ho cercato di precisare in precedenti lavori, i quali sfruttano in maniera essenziale le risorse offerte dalla teoria delle categorie: ho parlato di un "naturalismo intrecciato" e ho avanzato l'ipotesi che la sintassi stessa emerga da un processo di *lifting* di struttura da schemi legati alla corporeità – dunque una struttura ben carica di semantica. Non c'era niente in quanto ho così proposto che ne impedisse l'applicazione a sistemi di IA.

Riferimenti bibliografici

BARA B. (1990), *Scienza cognitiva: Un approccio evolutivo alla simulazione della mente*, Bollati Boringhieri, Torino.

BROOKS R. (1991), *Intelligence without representation*, in «Artificial Intelligence», 47, pp. 139-159.

BECHTEL W. (1992), *Filosofia della mente*, Il Mulino, Bologna (ed. orig. 1988).

CHOMSKY N. (1991), *Il linguaggio e il problema della conoscenza*, Il Mulino, Bologna (ed. orig. 1988).

CHURCHLAND P. (1992), *La natura della mente e la struttura della scienza: una prospettiva neurocomputazionale*, Il Mulino, Bologna (ed. orig. 1989).

DREYFUS H. (1972), *What computers can't do*, Harper & Row, New York.

FODOR J. e PYLYSHYN Z. (1988), *Connectionism and cognitive architecture: a critical analysis*, in «Cognition», 28, pp. 3-71.

FUM D. (1994), *Intelligenza artificiale,* Il Mulino, Bologna.

GARDNER H. (1990), *La nuova scienza della mente*, Feltrinelli, Milano (ed. orig. 1985).

HAUGELAND J. (1988), *Intelligenza artificiale*, Bollati Boringhieri, Torino (ed. orig. 1985).

HAUGELAND J. (1989), (a cura di), *Progettare la mente*, Il Mulino, Bologna (ed. orig. 1981).

HOFSTADTER D. (1985), *Gödel, Escher, Bach*, Adelphi, Milano (ed. orig. 1979).

HOFSTADTER D. e DENNETT D. (1985), (a cura di), *L'io della mente*, Adelphi, Milano (ed. orig. 1981).

JOHNSON-LAIRD P. (1990), *La mente e il computer*, Il Mulino, Bologna (ed. orig. 1988).

LAWVERE F.W. e SCHANUEL S. (1994), *Teoria delle categorie*, Muzzio, Padova.

LOLLI G. (1991), *Logica e intelligenza artificiale*, in «Sistemi intelligenti», 3, pp. 7-36.

MCDERMOTT D. (1991), *Una critica della ragione pura*, in «Sistemi intelligenti», 3, pp. 113-139.

MINSKY M. (1989), *La società della mente,* Adelphi, Milano (ed. orig. 1985).

PARISI D. (1989), *Intervista sulle reti neurali*, Il Mulino, Bologna.

PENROSE R. (1992), *La mente nuova dell'imperatore*, Rizzoli, Milano (ed. orig. 1989).

PERUZZI A. (1993), *Holism: the polarized spectrum*, in «Grazer Philosophische Studien», 46, pp. 231-282.

PERUZZI A. (1994), *Constraints on universals*, in Casati R. e Smith B. (a cura di) *Philosophy and the cognitive sciences*, Hölder-Pichler-Tempsky, Wien, pp. 357-370.

PERUZZI A. (1996), *Orme nel silicio, orme nella storia*, in «Paradigmi», 42, pp. 535-553.

PERUZZI A. (2000), *The geometric roots of semantics*, in Albertazzi L. (a cura di), *Meaning and cognition*, John Benjamins, Amsterdam, pp. 169-201.

PORT R. E VAN GELDER T. (1995), (a cura di), *Mind as motion: explorations in the dynamics of cognition*, Bradford Books/MIT Press, Cambridge (Mass.).

RUMELHART D. e MCCLELLAND J. (1991), *PDP*, Il Mulino, Bologna (ed. orig. 1986) trad. it. parziale.

SCHANK R. (1989), *Il computer cognitivo*, Giunti, Firenze (ed. orig. 1987).

SEARLE J. (1985), *Menti, cervelli, programmi*, in Hofstadter D. e Dennett D. (a cura di), *L'io della mente*, Adelphi, Milano (ed. orig. 1981), pp. 341-360.

SMOLENSKY P. (1992), *Il connessionismo tra simboli e neuroni*, Marietti, Casale Monferrato (ed. orig. 1988).
WALTERS R. (1991), *Categories and computer science*, Cambridge University Press, Cambridge.
WINOGRAD T. e FLORES F. (1988), *Calcolatori e conoscenza*, EST Mondadori, Milano (ed. orig. 1986).

ALBERTO PERUZZI
Università degli Studi di Firenze
Dipartimento di Filosofia
alper@unifi.it