

Parte IV

La mente come emergenza nella terza cultura

7. Il concetto di emergenza applicato alla mente

7.1 *Conoscere la mente: dalla spiegazione alla comprensione*

L'enigma della mente consiste probabilmente in questo: non abbiamo idea di quale *tipo* di “spiegazione completa della mente” potrebbe darci piena soddisfazione intellettuale [cfr. Bieri 1992, 89]. Questo significa che: (a) esistono molti *tipi* di spiegazione; (b) ci sono fenomeni (eventi e leggi) che non conosciamo, ma abbiamo un'idea di cosa considereremmo una loro spiegazione soddisfacente; (c) la mente non è uno di questi fenomeni. E' importante spendere qualche parola su queste tre asserzioni.

a) *Esistono molti tipi di spiegazione.* Questo fatto è relativamente ovvio: tutti noi facciamo continuamente esperienza di questa molteplicità di “spiegazioni”. A questo proposito Mario Bunge ha proposto di suddividere le spiegazioni in due gruppi: quelle che possono (ma non devono necessariamente) essere causali (A) e quelle decisamente non causali (B). In ciascun gruppo si trovano varie risposte possibili alla richiesta di *spiegare* un certo X (evento oppure legge). Spiegazioni di X che possono essere causali sono per esempio [cfr. Bunge 1963, 319-21]:

- A1) l'inclusione di X in una sequenza di eventi o stati;
- A2) l'inclusione di X in un processo di genesi o evoluzione;
- A3) l'associazione di X con fatti o leggi di ordine diverso;
- A4) l'analisi del funzionamento delle (eventuali) parti di X.

Ma anche le seguenti operazioni non causali, in molti casi, possono soddisfare pienamente la richiesta di *spiegare* X [cfr. Bunge 1963, 322-9]:

- B1) includere X in una classe (di eventi o leggi);
- B2) descrivere il funzionamento di X;
- B3) descrivere la struttura di X;

- B4) riferire X a qualcos'altro ("riduzione");
- B5) riferire X a una totalità ("spiegazione organicistica");
- B6) includere X in un processo stocastico ("spiegazione statistica");
- B7) descrivere il fine di X ("spiegazione teleologica");
- B8) descrivere i conflitti di X ("spiegazione dialettica").

Naturalmente, la spiegazione di un fenomeno X comporta spesso la compresenza di due o più di tali *inclusioni in e descrizioni di* "altro-da-X".

b) *Ci sono fenomeni (eventi e leggi) che non conosciamo, ma abbiamo un'idea di cosa considereremmo una loro spiegazione soddisfacente.* In presenza di un evento o una legge relativi a qualche sistema che non conosciamo, abbiamo di solito un'idea, per quanto vaga, di quale spiegazione ci darebbe completa soddisfazione [Bieri 1992, 89]. Per esempio: supponiamo che una persona di media cultura, pur sapendo usare un registratore, stia per ricevere - per la prima volta - la spiegazione del processo effettivo di trasformazione dei suoni in quel "qualcosa" che si può conservare su un nastro magnetico. Interrogata, probabilmente la persona in questione direbbe di aspettarsi che la spiegazione consista: nella descrizione del funzionamento di tutte le parti importanti del registratore, nell'*elencazione* delle principali leggi fisiche che descrivono le onde, in una descrizione strutturale del nastro (come ad esempio: «striscia di materiale plastico ricoperto da un ossido magnetizzabile...»), in una descrizione sommaria delle sue proprietà magnetiche, ecc.¹

In presenza di qualcosa che non conosciamo, sappiamo anche che cosa *non* possiamo considerare soddisfacente. Una spiegazione circolare, per esempio, non è soddisfacente:

Solo una teoria che spiega la coscienza in termini di eventi inconsci può spiegare la coscienza. Se il tuo modello della produzione del dolore da parte delle attività cerebrali presenta un riquadro con l'etichetta «dolore», non hai ancora iniziato a spiegare che cosa sia il dolore [Dennett 1991, 507].

¹L'ulteriore spiegazione delle leggi coinvolte in questa spiegazione non rientra nelle aspettative, perché tutti sappiamo - almeno intuitivamente - che «*la spiegazione non è transitiva*» [Putnam 1981a, 223], cioè che, nel nostro esempio, la spiegazione (delle leggi) delle onde non fa parte della spiegazione *della* registrazione delle onde.

Un caso particolare di spiegazione circolare, ovvio ma insidioso, è quello del “cambio di etichetta”, perfettamente esemplificato da Molière nella sua commedia *Il malato immaginario* (1673): Argante, l’ipocondriaco del titolo, spiega il fatto che l’oppio faccia dormire affermando che esso possiede una *virtus dormitiva* che fa addormentare! Numerosi esempi di “cambio di etichetta” sono forniti da Dennett: lo «schifo-serpentino» spiega il disagio che proviamo quando vediamo un serpente [Dennett 1991, 428], l’«intrinseca spaventosità» spiega il dolore [Dennett 1991, 77], la «comicità» spiega il ridere [*ibid.*], ecc.

Cosa ci può garantire *a priori* che una spiegazione non sia, in realtà, un semplice cambio di etichetta? Essenzialmente, il fatto di applicare a ciò che va spiegato categorie applicabili anche *ad altri fenomeni*. Negli anni della cibernetica, per esempio, spiegazioni soddisfacenti di molti fenomeni biologici hanno trovato fondamento nell’applicazione, agli organismi viventi e alle macchine artificiali, delle medesime categorie (organizzazione, teleonomia, ecc.).

Se invece la categoria utilizzata per “spiegare” X è coestensiva a X, la spiegazione non può risultare soddisfacente. Se «schifo-serpentino», per esempio, si applica *sempre e solo* al disagio per i serpenti, allora la spiegazione del secondo mediante il primo non può essere soddisfacente. L’*intenzionalità* (nel senso di Husserl), analogamente, non può essere una categoria esplicativa soddisfacente dell’esperienza cosciente (*Erlebnis*) se sono intenzionali *tutte e sole* le esperienze coscienti (*Erlebnisse*).

Abbiamo una “notizia buona” e una “cattiva”. La “notizia buona” è che il concetto di emergenza, applicato alla mente, non è un semplice cambio di etichetta.

La vita in quanto tale, cioè l’insieme dei fenomeni di organizzazione cellulare studiati dai biologi (fermentazione, fosforilazione, ecc.), può essere infatti classificata come fenomeno emergente (o, al limite, come insieme di fenomeni emergenti). Anche alcuni fenomeni descritti dai modelli dell’evoluzione naturale possono essere considerati emergenti, ma - nel caso della vita - abbiamo visto che tale categorizzazione è non solo lecita (cioè la definizione tecnica di «emergenza» è applicabile) ma anche *opportuna* (perché la vita “richiede” di essere descritta da una categoria caratterizzata epistemicamente). Questo fatto è molto importante, perché previene il sospetto che l’emergenza, se non coestensiva di specifici processi mentali,

possa però avere la stessa estensione della *soggettività* - cioè la categoria più ampia in cui possiamo includere i fenomeni mentali [Searle 1992, 110]. Ma la vita, in quanto tale, *non è soggettiva*: quindi la categoria di emergenza si applica ad almeno un fenomeno che non appartiene alla categoria più ampia possibile in cui i filosofi inseriscono i processi mentali.

A priori, quindi, abbiamo la garanzia che un'eventuale spiegazione completa (e completamente soddisfacente) della mente *potrebbe utilizzare la categoria di emergenza senza ridursi a un cambio di etichetta* (da «mentale» a «emergente») e *senza ridurre l'emergenza della mente a una virtus dormitiva del sistema nervoso*.

La “notizia cattiva” è l’«*ipotesi della limitazione cognitiva*» [Bieri 1992, 88]: molti pensatori ritengono che la mente non possa spiegare completamente se stessa.

L'impossibilità di una completa “auto-spiegazione” della mente è stata talvolta sostenuta chiamando in causa il Teorema di Gödel - ma utilizzando quest'ultimo, per forza, solo *metaforicamente*; il Teorema di Gödel, infatti, ha validità letterale solo in logica matematica e non asserisce nulla sui poteri auto-esplicativi della mente:

Non vedo alcuna ragione per cui il Teorema di Gödel imponga qualche limitazione alla nostra capacità di formulare e di verificare i meccanismi generali sulla base dei quali si producono i processi di pensiero nel mezzo costituito dalle cellule nervose [Hofstadter 1979, 766].

Indipendentemente dal Teorema di Gödel, però, ogni spiegazione - secondo l'ipotesi della limitazione cognitiva - deve fare i conti con certi limiti *assoluti e intrinseci*.

Questi limiti sono una diretta conseguenza della “natura” stessa della spiegazione e, a metà secolo, sono stati messi in luce, meglio che da altri, dall'economista Friedrich August von Hayek. Secondo von Hayek,

la spiegazione consiste nella formazione, all'interno del cervello, di un “modello” degli eventi complessi che si vogliono spiegare, un modello le cui parti sono definite in base alla loro posizione in un'ampia struttura di relazioni che costituisce lo schema semipermanente dal quale le rappresentazioni dei singoli eventi ricevono il loro significato [von Hayek 1952, 256].

Ora, se per sua natura il processo esplicativo comporta la costruzione di un modello degli eventi che si vogliono spiegare, allora il rapporto «*tra l'agente che effettua la spiegazione e l'oggetto spiegato deve soddisfare quelle relazioni formali che debbono*

sussistere tra un qualunque apparato di classificazione e i singoli oggetti che esso classifica» [von Hayek 1952, 263-4].

La conseguenza, importantissima, è che esiste un limite assoluto a quanto il cervello umano, inteso come apparato di classificazione, potrà mai spiegare completamente:

un qualunque apparato di classificazione deve possedere una struttura con un grado di complessità più elevato della struttura propria degli oggetti che esso classifica; e, di conseguenza, [...] la capacità di qualunque agente esplicativo è necessariamente circoscritta a oggetti dotati di una struttura con un grado di complessità inferiore rispetto al proprio. Se ciò è esatto, se ne deduce che nessun agente esplicativo può mai spiegare oggetti del suo tipo, o di pari grado di complessità, e, quindi, che il cervello umano non potrà mai spiegare completamente le sue stesse operazioni [von Hayek 1952, 264].

La conclusione di von Hayek è la stessa a cui sono pervenute, in modi e tempi diversi, varie generazioni di filosofi: poiché il carattere distintivo di un processo mentale sta nel suo essere determinato dalle relazioni con tutti gli altri processi mentali,

non soltanto la mente nel suo complesso, ma persino tutti i singoli processi mentali devono eternamente rimanere per noi fenomeni di un tipo speciale, che non riusciremo mai a spiegare completamente in termini di leggi fisiche, per quanto essi siano prodotti dagli stessi principi che sappiamo operare nel mondo fisico. Se si preferisce, si può anche esprimere questo concetto affermando che i fenomeni mentali fondamentalmente non sono “null’altro che” processi fisici; tuttavia, questo non cambia il fatto che nella trattazione dei processi mentali noi non potremo mai evitare l’impiego di termini mentali, e che dovremo sempre rimanere entro i limiti di un dualismo pratico, un dualismo che non si fonda su una qualsiasi asserzione riguardo a una differenza oggettiva tra le due classi di eventi, bensì sulle dimostrabili limitazioni dei poteri della nostra mente [von Hayek 1952, 271].

c) *La mente non appartiene a quei fenomeni che non conosciamo, ma dei quali abbiamo un’idea di cosa considereremmo una spiegazione soddisfacente. Se accettiamo la validità dell’argomentazione di von Hayek, diventa chiara la ragione di tale singolarità. Infatti, una spiegazione completamente soddisfacente, semplicemente, non può esistere. Si noti che ciò non significa che non possa esistere alcuna spiegazione: molti dei tipi di spiegazione elencati da Bunge possono essere utilizzati per la mente. Mediante la tomografia a emissione di positroni, per esempio, sappiamo associare i processi mentali a fatti e leggi che riguardano il cervello: questa è una legittima spiegazione del tipo A3; oppure consideriamo l’asserzione di von Hayek, «i fenomeni mentali fondamentalmente non sono “null’altro che” processi fisici» [von*

Hayek 1952, 271]: anche questa è una spiegazione, e precisamente del tipo B4. Gli esempi potrebbero continuare.

Ma il problema è che tutte queste spiegazioni, a causa dell'*inevitabilità* del “dualismo pratico” di cui parla von Hayek, non ci possono dare *piena soddisfazione intellettuale*. Sono spiegazioni legittime, ma non spiegazioni soddisfacenti.

E allora? Accantonare una volta per tutte il nostro desiderio (cioè la filo-) di conoscenza (-sofia) della mente? La risposta è decisamente negativa. Ci sono ancora almeno due strade, verso la conoscenza della mente, che possiamo percorrere. Una possibilità è quella della *comprensione soggettiva*, l'altra quella della *riproduzione esplicativa*.

Il concetto di comprensione, in quanto contrapposto a quello di spiegazione, risale agli ultimi decenni del XIX secolo, quando i filosofi avvertirono l'esigenza di distinguere il metodo delle “scienze della natura” da quello delle “scienze dello spirito”. Lo storico e filosofo Wilhelm Dilthey propose di sancire la differenza tra tali discipline distinguendo le due rispettive modalità di conoscenza: lo spiegare oggettivamente (*erklären*) e il comprendere soggettivamente (*verstehen*). La filosofia contemporanea non utilizza più queste categorie, i limiti delle quali sono stati segnalati - nel corso del Novecento - tanto dagli scienziati quanto dai filosofi.

Tuttavia, possiamo “rispolverare” i due concetti per indicare due forme di conoscenza della mente che desideriamo distinguere: la conoscenza completa e completamente soddisfacente e la conoscenza intuitiva e intuitivamente soddisfacente. Chiameremo *spiegazione* (o *spiegazione oggettiva*) la modalità della prima forma di conoscenza e *comprensione* (o *comprensione soggettiva*) la modalità della seconda. Esse non vanno confuse con le due forme di conoscenza (soggettiva e oggettiva) relative ai Mondi (2 e 3) di Popper.

Se von Hayek ha ragione, e come sembra dimostrare il fatto che nessuno ha un'idea precisa e condivisibile di quale tipo di “spiegazione della mente” potrebbe arrecare piena soddisfazione intellettuale, *la spiegazione oggettiva è impossibile*.

Resta allora la comprensione soggettiva, cioè la costruzione di un modo di *concepire la mente* che riduca il più possibile «*la “dissonanza intuitiva” che nasce quando si riflette sul fondamento fisico della mente*» [Churchland - Sejnowski 1992,

13]. La costruzione di un tale modo di pensare, diversamente da come procede la spiegazione, deve fare largo uso di analogie, metafore e anche “modelli” qualitativi.

Esiste infine una terza forma di conoscenza della mente, la cui modalità verrà chiamata *riproduzione* (o *riproduzione esplicativa*). Si tratta di un tipo di spiegazione non contemplato da Bunge, ma assai pertinente: creare una macchina pensante significherebbe aver risolto - in un modo o nell'altro - il “mistero”. Per dirlo con le parole del neurofisiologo ed epistemologo Humberto R. Maturana:

Solo una completa riproduzione è una spiegazione completa [Maturana 1970, 88].

Questa asserzione, liberamente decontestualizzata, indica una modalità conoscitiva del tutto diversa sia dalla spiegazione, sia dalla comprensione. E' *logicamente* possibile, infatti, che si possa riprodurre artificialmente un sistema già esistente in natura in modo tale che funzioni perfettamente, ma senza avere alcuna idea dettagliata del *perché* e del *come* ciò accada. Questo approccio non dovrebbe risultare del tutto nuovo: è stato incontrato quando abbiamo accennato alla *vita artificiale* (§5.4). Generalizzando queste considerazioni, possiamo considerare *esplicativa* la riproduzione artificiale di *qualsiasi* fenomeno mentale (risolvere un anagramma, fare un'analogia, ecc.).

Alla luce di quanto sopra, possiamo proporre la seguente chiave di lettura - quella con cui interpreteremo l'applicazione del concetto di emergenza alla mente:

- 1) Tale applicazione non è un semplice cambio di etichetta, quindi *potrebbe* essere parte integrante del tentativo di spiegare oggettivamente la mente;
- 2) tuttavia, essa *non è* parte integrante di un simile tentativo: piuttosto, fa parte del tentativo di *capire la mente per mezzo di modalità conoscitive diverse dalla spiegazione*;
- 3) tali modalità conoscitive sono: la comprensione e la riproduzione.

Dedicheremo i prossimi capitoli all'illustrazione delle principali applicazioni del concetto di emergenza alla mente, *nel contesto della terza cultura e secondo la chiave di lettura appena proposta*. In particolare, presenteremo due classi di modelli dell'attività mentale: i modelli elaborati dal “gruppo PDP” (gruppo di ricerca sul Processamento Distribuito in Parallelo) e quelli elaborati dal “FARG” (Gruppo di

Ricerca sulle Analogie Fluide). *Entrambi i gruppi utilizzano il concetto di emergenza e lo applicano sia ai loro modelli, sia ai fenomeni mentali.*

Sia i membri del gruppo PDP, sia i “FARGonauti” fanno ampio e disinvolto uso di due concetti molto delicati: quello di *livello* e quello di *simbolo*. Prima di procedere, pertanto, è importante soffermarsi su questi due concetti. I prossimi due paragrafi saranno dedicati rispettivamente alla nozione di «livello» e a quella di «simbolo»; concluderemo il capitolo con un paragrafo di carattere storico-filosofico, che costruirà il contesto culturale del gruppo PDP e del FARG.

7.2 *La nozione di «livello» e la teoria gerarchica del reale*

Secondo il *British Emergentism*, nel mondo esiste una gerarchia di livelli di complessità organizzazionale di particelle materiali («*a hierarchy of levels of organizational complexity of material particles*» [McLaughlin 1992, 50]) che include, in ordine ascendente: il livello strettamente fisico, quello chimico, quello biologico e quello psicologico [*ibid.*]. Questa è la cosiddetta “teoria gerarchica del reale”, elaborata originariamente proprio dai primi emergentisti ma successivamente adottata dalla maggioranza dei ricercatori (in particolare dagli scienziati cognitivi) - ovviamente nelle versioni più varie. Il cuore concettuale della teoria gerarchica, comune a tutte le versioni di essa, è rappresentato dalla problematica nozione di *livello*.

Il termine «livello» si trova associato, nella letteratura filosofica, a una quantità imbarazzante di significati diversi². Una rassegna di tali significati, anche parziale, oltrepasserebbe di gran lunga la portata del presente lavoro. D’altra parte, a ciascun significato di «livello» si può far corrispondere una differente teoria gerarchica del reale, cioè una diversa visione del mondo come “stratificazione di livelli”; è importante, pertanto, tentarne una precisazione, anche se minima.

²Il termine «gerarchia» non è altrettanto problematico. Benché esso significhi letteralmente «ordine delle cose sacre» e con tale significato sia stato introdotto nella filosofia occidentale (lo pseudo Dionigi l’Areopagita lo usò nel secolo V per indicare l’ordinamento delle intelligenze angeliche e il corrispondente ordinamento ecclesiastico), oggi lo si adopera in filosofia per qualsiasi disposizione entro una successione ordinata.

La prospettiva che verrà adottata per condurre questo tentativo è, almeno in parte, la stessa dell’“approccio sistemico” di von Bertalanffy. Von Bertalanffy stesso scrive:

Attualmente noi “vediamo” l’universo come se fosse una straordinaria gerarchia, dalle particelle elementari ai nuclei atomici, agli atomi, alle molecole, ai composti ad alto livello molecolare, alla ricchezza delle strutture [...] tra le molecole e le cellule [...], per giungere sino alle cellule, agli organismi e, ancor più in là, alle organizzazioni che trascendono l’individuo [von Bertalanffy 1968, 58].

A ciascun *livello* di questa gerarchia si trovano *sistemi* caratteristici di quel livello e specifici *modelli* di tali sistemi. Uno schema intuitivo e informale dei sistemi e dei loro modelli (quelli disponibili negli anni Sessanta) è il seguente [von Bertalanffy 1968, 59-60]:

LIVELLO	DESCRIZIONE ED ESEMPI	TEORIA E MODELLI
Strutture statiche	Atomi, molecole, cristalli, strutture biologiche dal livello della microscopia elettronica a quello macroscopico.	Formule strutturali della chimica; cristallografia; descrizioni anatomiche.
Meccanismi tipo orologio	Orologi, macchine convenzionali in generale, sistemi solari.	Settori convenzionali della fisica, come le leggi della meccanica (newtoniana ed einsteiniana) e altri.
Meccanismi di controllo	Termostati, servomeccanismi, meccanismo omeostatico negli organismi.	Cibernetica; teoria della retroazione e dell’informazione.
Sistemi aperti	Fiamma, cellule e organismi in generale.	a) Espansione di teorie fisiche verso sistemi che mantengono se stessi in presenza di flusso di materia (metabolismo). b) Immagazzinamento di informazione nel codice genetico (DNA). [...]
Organismi inferiori	Organismi “simili a piante”; differenziazione crescente dei sistemi [...]; distinzione di riproduzione e di individuo funzionale.	Teoria e modelli sono quasi completamente mancanti.
Animali	Crescente importanza del trasporto di informazione (evoluzione dei recettori, sistemi nervosi); apprendimento; inizio della coscienza.	Inizi della teoria degli automatismi [...], retroazione (fenomeni regolativi), comportamento autonomo [...], ecc.
Uomo	Simbolismo; passato e futuro,	Incipiente teoria del simbolismo.

	l'ego e il mondo, coscienza di sé, ecc., come conseguenze; comunicazione mediante il linguaggio, ecc.	
Sistemi socio-culturali	Popolazioni di organismi (compresi quelli umani); comunità determinate da simboli (culture) nei soli uomini.	Leggi statistiche, e forse anche dinamiche, nella dinamica delle popolazioni, in economia, possibilmente in storia. [...]
Sistemi simbolici	Linguaggio, logica, matematiche, scienze, arti, morale, ecc.	Algoritmi di simboli [...]; "regole del gioco" [...], ecc.

Che cos'è un livello, in base a questo schema? E', così sembra, una *classe di equivalenza di sistemi*. Si noti però che esistono due tipi di rapporti tra le classi di sistemi: quello tra due classi i cui sistemi, rispettivamente, *possono essere inclusi gli uni negli altri* e quello tra classi i cui rispettivi sistemi stabiliscono tra loro relazioni *diverse dall'inclusione*. Del primo tipo è per esempio il rapporto tra il livello degli animali e il livello dei sistemi socio-culturali (popolazioni di animali); del secondo tipo è invece quello tra il livello dell'uomo e il livello dei sistemi simbolici.

Per la compresenza di rapporti così diversi tra i livelli elencati da von Bertalanffy, è preferibile distinguere due accezioni di *livello* (inteso come classe di sistemi); a queste due occorre aggiungerne almeno una terza che tenga conto di *come noi descriviamo i sistemi*. Le tre accezioni che verranno considerate sono così le seguenti:

- livello come ordine di complessità;
- livello come "gradino" della scala di organizzazione;
- livello come piano di analisi.

Consideriamo la prima accezione: *il livello come ordine di complessità*. Poiché non esiste alcun criterio con cui *ordinare* l'insieme di tutti i sistemi in base alla loro complessità, ogni tassonomia basata sulla complessità va considerata come un "postulato" della teoria gerarchica. La tassonomia che qui proponiamo è la seguente:

1. Sistemi non complessi o minimamente complessi;
2. Sistemi di media complessità;
3. Sistemi complessi (e molto complessi).

I sistemi *non complessi* sono, secondo Richard Dawkins,

le cose «semplici», come le pietre, le nuvole, i fiumi, le galassie e i quark. Queste cose sono gli oggetti di cui si occupano le scienze fisiche [Dawkins 1986, 15].

Il fatto che questi sistemi siano quelli studiati dalla fisica può sembrare strano:

La fisica sembra un argomento complicato perché le idee della fisica sono per noi difficili da capire. Il nostro cervello è stato progettato per capire la caccia e la raccolta, l'accoppiamento e il compito di allevare i figli; un mondo di oggetti di media grandezza che si muovono nelle tre dimensioni a velocità moderate. Noi siamo mentalmente male equipaggiati per capire oggetti molto piccoli e molto grandi, cose la cui durata si misura in picosecondi o in giga-anni; particelle che non hanno una posizione; forze e campi che non possiamo vedere o toccare. Noi pensiamo che la fisica sia complicata perché per noi è difficile da capire, e perché i libri di fisica sono pieni di complesse formule matematiche. Ma gli oggetti studiati dai fisici sono ancora oggetti fondamentalmente semplici. Essi sono nuvole di gas o minuscole particelle, o masse di materia uniforme come i cristalli, in cui strutture atomiche si ripetono quasi senza fine. Essi non hanno parti funzionanti complesse, almeno a paragone con gli oggetti biologici. [...] Il comportamento di oggetti fisici, non biologici, è così semplice che possiamo usare, per descriverlo, il linguaggio matematico esistente ed è appunto questa la ragione per cui i libri di fisica sono pieni di matematica [Dawkins 1986, 16-7].

Sempre del primo ordine sono i sistemi *minimamente complessi*, come quelli che von Bertalanffy chiama “meccanismi tipo orologio”: orologi, pendoli e macchine convenzionali in genere. I sistemi del primo ordine possono essere caratterizzati dal fatto che nessun fenomeno emergente è associabile ad essi (abbandoniamo, con questa affermazione, la concezione di Broad, secondo cui le proprietà dei composti chimici sono ineducibili *per principio*).

I sistemi del secondo tipo, quelli di *media complessità*, sono i computer. Nel suo famoso articolo sull'auto-organizzazione, Ross Ashby scriveva a questo proposito:

“fino ad un'epoca recente non avevamo esperienza dei sistemi di media complessità; o si trattava di sistemi come l'orologio e il pendolo, e trovavamo le loro proprietà limitate ed evidenti, oppure come il cane e l'essere umano, e trovavamo le loro proprietà così ricche e notevoli che le pensavamo soprannaturali. Solo negli ultimi anni siamo stati gratificati, con i calcolatori universali, di sistemi abbastanza ricchi per essere interessanti, e tuttavia sufficientemente semplici da essere comprensibili... Il calcolatore è un dono del cielo... poiché permette di gettare un ponte sull'enorme abisso concettuale che separa il semplice e il comprensibile dal complesso e dall'interessante”.³

³Ashby W. R. [1962], “Principles of the self organizing system”, in von Foerster - Zopf (a cura di), *Principles of self organization*, Pergamon, Oxford 1962, pp.255-78; citato in Atlan [1972, 52].

I computer appartengono al livello in cui von Bertalanffy inserisce tutti i “meccanismi di controllo”: dai termostati ai meccanismi omeostatici negli organismi. Ciò che caratterizza i sistemi di media complessità, in un’ottica emergentistica, è il fatto che ad essi si possano associare fenomeni innocentemente emergenti.

I sistemi *complessi*, infine, sono gli organismi biologici. E’ importante avere ben chiaro in mente l’enorme salto di complessità che separa i primi due ordini dal terzo:

I *libri* di fisica possono essere complicati, ma essi, come le automobili e i computer, sono i prodotti di oggetti biologici: ossia di cervelli umani. Gli oggetti e i fenomeni descritti da un libro di fisica sono più semplici di una singola cellula nel corpo del suo autore. E l’autore è formato da bilioni di quelle cellule, molte delle quali diverse l’una dall’altra, organizzate con una complessa architettura e una tecnica di precisione in una macchina funzionante, capace di scrivere un libro. Il nostro cervello non è meglio equipaggiato a far fronte agli estremi di complessità che agli estremi di grandezza e agli altri difficili estremi della fisica. Nessuno ha ancora inventato la matematica necessaria per descrivere nella sua globalità la struttura e il comportamento di un oggetto così complesso come un fisico, o anche solo come una delle sue cellule [Dawkins 1986, 17].

I sistemi biologici, dal punto di vista emergentistico, sono tutti quanti caratterizzati dalla loro associazione con la *vita* - intesa come fenomeno emergente.

Il terzo ordine di complessità può essere suddiviso in due o più sotto-ordini, parzialmente corrispondenti ai tre livelli in cui von Bertalanffy inserisce rispettivamente gli organismi inferiori, gli animali e gli esseri umani. Definiamo il primo sotto-ordine per esclusione: comprende tutti i *sistemi biologici privi di cellule nervose*; dal punto di vista dell’emergentismo, ai sistemi di questo sotto-ordine possiamo associare un unico fenomeno emergente, la vita. Il sotto-ordine complementare, quello dei *sistemi neuro-biologici* andrebbe probabilmente considerato più complesso; esso corrisponde al livello in cui von Bertalanffy inserisce gli animali. Tra questi ultimi, però, alcuni emergentisti attribuiscono una rilevanza particolare solo ai sistemi neuro-biologici *e plastici*: la mente, intesa come fenomeno emergente, dovrebbe forse essere associata a tutti e soli i sistemi di *questo* sotto-ordine [Bunge 1980, 75]. Infine, volendo segnalare tassonomicamente le peculiarità cognitive degli esseri umani, l’insieme delle menti umane potrebbe essere associato a un terzo sotto-ordine di complessità: quello degli esseri umani; tuttavia, dal punto di

vista biologico, la complessità del sistema nervoso umano non è tale da meritare un ordine a sé stante.

Si noti che, all'interno di ciascun ordine (e sotto-ordine) di complessità, si hanno differenze - spesso importanti - di complessità. Ciò è analogo a quanto avviene con gli ordini di grandezza (= logaritmi in base 10 delle grandezze fisiche): oggetti di grandezze diverse appartengono allo stesso ordine. (L'analogia è migliore se consideriamo i tre gruppi in cui vengono convenzionalmente suddivisi gli ordini di grandezza: *microcosmo*, *mesocosmo*, *macrocosmo*.)

Passiamo ora alla seconda accezione di livello: *il livello come "gradino" della scala di organizzazione*. La parola «livello» viene usata in questa accezione quando si fissa l'attenzione su un unico ordine di complessità e si osserva l'organizzazione gerarchica dei sistemi di tale ordine. In questo caso, la relazione tra sistemi appartenenti a gradini diversi (dello stesso ordine) *può essere* - ma non è necessariamente - quella di *inclusione fisica*⁴. Essi, tuttavia, sono tutti considerati individualmente: non si deve immaginare, per ciascun ordine di complessità, *un unico sistema* organizzato gerarchicamente.

La scala del primo ordine di complessità comprende, "dal basso all'alto", sistemi come i seguenti: le particelle elementari, gli atomi, le molecole, tutte le macrostrutture non biologiche (fluidi inorganici, cristalli, ecc.), le stelle e i pianeti senza vita, i sistemi solari (considerati senza vita) e le galassie (considerati senza vita).

La scala del secondo ordine di complessità comprende, partendo da un computer (hardware + software) e procedendo verso l'alto: i sistemi di computer collegati in rete e le "reti di reti" (sistemi di sistemi). Naturalmente i computer sono costruiti assemblando varie componenti elettroniche e queste, a loro volta, sono costituite da parti più piccole; procedendo verso il basso si arriva di nuovo agli atomi e alle particelle elementari.

⁴L'inclusione fisica non va assolutamente confusa con l'inclusione *logica* (tra insiemi) né con la relazione "parti-tutto". Quest'ultima, in particolare, è relativa al nostro modo di descrivere *un unico sistema*, mentre l'inclusione fisica riguarda oggetti che si possono anche presentare come *sistemi autonomi* (con la parziale eccezione dei quark): una cellula epatica, per esempio, non si presenta *necessariamente* come una parte del fegato. Bunge, per sottolineare il fatto che quella tra sistemi appartenenti a "gradini" di organizzazione diversi è una relazione *sui generis*, la definisce nei termini della "funzione di composizione" (che associa ad ogni elemento di un dominio l'insieme delle parti di tale elemento) [Bunge 1977, 504].

L'organizzazione gerarchica dei sistemi biologici è più complessa. Le cellule viventi, in particolare, sono costituite da un grande numero di componenti e ciascuna componente da componenti più piccole - e così via fino alle molecole, agli atomi e ai quark. A loro volta, le cellule possono formare popolazioni di esseri unicellulari oppure, negli esseri pluricellulari, tessuti (sistemi di cellule).

Se focalizziamo l'attenzione sul sotto-ordine di complessità dei sistemi biologici *privi di cellule nervose*, vedremo una scala di organizzazione come la seguente (a partire dalle cellule): tessuti, organi, organismi (sistemi di organi), popolazioni (sistemi di organismi) e infine ecosistemi (sistemi di popolazioni). Se invece focalizziamo l'attenzione su un sistema nervoso, troveremo un'organizzazione come la seguente: neuroni organizzati in reti (microsistemi neurali) e reti organizzate in vari macrosistemi neurali (gangli o cervello, sistema nervoso gangliare o centrale); il sistema nervoso, a sua volta, occupa un "gradino" inferiore rispetto a quello occupato dal sistema neuro-endocrino. Nel caso degli esseri umani, infine, partendo dal sistema neuro-endocrino e procedendo verso l'alto, troviamo: gli organismi nella loro intelligenza (più, per usare i termini di Merlin Donald, i dispositivi di immagazzinamento simbolico esterno) e, in ultimo, i sistemi sociali (con le loro memorie non biologiche). Si presti molta attenzione a quest'ultima scala di organizzazione (quella degli esseri umani): procedendo verso l'alto e partendo dal sistema nervoso, si incontra il sistema neuro-endocrino e poi l'organismo umano nel suo complesso. *La mente non appartiene alla scala di organizzazione degli esseri umani*⁵.

In conclusione, poiché *ogni scala di organizzazione appartiene a qualche ordine di complessità*, possiamo parlare di «livelli di complessità organizzazionale» intendendo i «"gradini" di una scala di organizzazione in un dato ordine di complessità».

Prima di procedere con la terza accezione di livello, occorre affrontare la domanda cruciale che ci si può porre relativamente alla teoria gerarchica: il mondo è

⁵Questa interpretazione si contrappone a quella, relativamente famosa, del neurobiologo Gordon Shepherd, il quale include il *comportamento* tra i livelli di organizzazione del sistema nervoso [Churchland - Sejnowski 1992, 39]. Ma inserire la mente, o il comportamento, nella scala di organizzazione umana è del tutto coerente con l'inserimento in essa di cellule e cervello *solo se si considerano la mente e il comportamento entità sostanziali* (come, appunto, cellule e cervello).

“in sé” una gerarchia di livelli di complessità organizzazionale oppure sono le nostre categorie a renderlo tale? In altre parole, dobbiamo chiederci *se noi siamo in grado di capire il mondo perché è gerarchico oppure se esso appare gerarchico perché quegli aspetti di esso che non lo sono eludono la nostra comprensione e la nostra osservazione* [Simon 1962, 219].

La soluzione “classica”, fornita da Kant, è l'*indecidibilità*: è indecidibile, più in generale, se le nostre categorie siano dettate dal reale o imposte ad esso. Se però ammettiamo di poter uscire dallo spazio epistemologico chiuso in cui il problema è stato posto, possiamo utilizzare ciò che sappiamo sul reale e che *non* proviene dalla teoria gerarchica (per esempio adottando una prospettiva evoluzionistica). In questo modo, benché chiaramente il problema del confronto “nudo” tra soggetto e oggetto non venga risolto, possiamo quanto meno «*spostarne la carica di drammatizzazione*» [Stengers 1985, 69]. E' questa l'ottica con cui l'economista e scienziato cognitivo Herbert A. Simon ha affrontato il problema.

Nel celeberrimo articolo “The Architecture of Complexity” (1962), Simon ha fornito delle buone ragioni per supporre che *il mondo sia realmente gerarchico* (indipendentemente dalle nostre categorie). Si può per esempio dimostrare, con alcune semplificazioni drastiche ma ragionevoli, che un sistema con un'architettura stratificata su molti livelli di complessità organizzazionale *evolve più rapidamente* di un sistema non gerarchico con lo stesso numero di costituenti fondamentali⁶. Inoltre, il primo sistema non viene danneggiato dalle perturbazioni tanto gravemente quanto il secondo (perché nel primo caso l'effetto disorganizzante della perturbazione procede

⁶Si immagini l'evoluzione come un processo in cui gli elementi che stanno formando un certo sistema si auto-assemblano. Supponiamo che il tentativo, da parte degli elementi che si auto-assemblano, di raggiungere un *qualsiasi* stato stabile richieda sempre la stessa quantità di tempo: si potrebbe allora pensare (sbagliando) che l'evoluzione *più veloce* corrisponda al tentativo di raggiungere lo stato stabile rappresentato dal sistema finale *in un unico passo*. Non è così, perché la probabilità di farcela dipende dal numero di elementi che cercano di stabilizzarsi. Sia infatti p la probabilità che gli elementi si disperdano prima di raggiungere uno stato stabile: la probabilità che gli elementi raggiungano uno stato stabile senza disperdersi è allora $(1-p)^s$, dove s è il numero di componenti di ciascun sottosistema stabile in cui gli elementi si organizzano. Consideriamo per esempio 1000 elementi che si auto-assemblano. Se lo fanno gerarchicamente, formando sottosistemi stabili di $s=10$ componenti ciascuno, allora la probabilità di raggiungere ciascuno stato stabile (il sistema finale è il 111° stato stabile) è $(1-p)^{10}$; se invece si auto-assemblano in un unico passo, allora la probabilità di raggiungere il primo e unico stato stabile, cioè il sistema finale, è $(1-p)^{1000}$. Questa differenza è assai rilevante; con $p=0.01$, per esempio, le due probabilità diventano rispettivamente: 9 su 10 e 43 su 10⁶. E' facile dimostrare che *la maggiore probabilità di raggiungere stati stabili rende più veloce l'evoluzione del sistema che si organizza gerarchicamente* [Simon 1962, 200-5].

“dall’alto in basso” incontrando resistenza ad ogni livello, nel secondo agisce direttamente sui costituenti fondamentali, cioè, per così dire, alla base della gerarchia). Le considerazioni di Simon suggeriscono che l’architettura del mondo sia effettivamente gerarchica - ma, naturalmente, non sono una *dimostrazione* di questa tesi.

Veniamo ora alla terza e più delicata accezione di livello: *il livello come piano di analisi*. Abbiamo già visto, rielaborando (liberamente) la teoria delle emergenze di Morin, che un sistema può essere analizzato su tre piani diversi: il piano degli elementi organizzati, il piano del sistema come globalità e il piano dei fenomeni emergenti (se ci sono) associabili al sistema. Vediamo come si può tradurre questa tassonomia nel caso di due sistemi complessi: un ecosistema e un cervello in funzione.

Consideriamo per primo il caso di un ecosistema⁷. Uno dei “padri” dell’ecologia scientifica, Eugene P. Odum, afferma che ci si può avvicinare all’analisi degli ecosistemi in due modi diversi: l’olistico e il meristico. Questi due approcci si collocano rispettivamente sul piano del sistema come globalità e su quello dei componenti organizzati. E’ interessante notare che quando

i componenti sono fortemente associati, eventuali proprietà emergenti si evidenzieranno solo a livello dell’intero. Tali importanti attributi si perderebbero se fosse applicato esclusivamente l’approccio meristico [Odum 1983, 20].

Possiamo quindi distinguere un terzo piano di studio, associato al «livello dell’intero», cioè il piano relativo all’analisi delle *proprietà emergenti*. Queste ultime, secondo Odum, sono una diretta conseguenza della struttura gerarchica del mondo:

Un’importante conseguenza della organizzazione gerarchica è che appena dei componenti o dei subinsiemi si combinano per produrre un più grande insieme funzionale, ne emergono nuove proprietà che non erano presenti al livello immediatamente precedente. Concordemente, una **proprietà emergente** ad un livello ecologico non può essere prevista dallo studio delle componenti di quel livello. [...] Sebbene le conoscenze acquisite su un certo livello possano facilitare lo studio del livello successivo, mai queste riescono a spiegare i fenomeni che vi ricorrono e che devono essere invece studiati per completare il quadro conoscitivo [Odum 1983, 4-5; grassetto di Odum].

⁷Per *ecosistema* si intenda «una unità che include tutti gli organismi che vivono insieme (comunità biotica) in una data area, interagenti con l’ambiente fisico, in modo tale che un flusso d’energia porta ad una ben definita struttura biotica e ad una ciclizzazione dei materiali tra viventi e non viventi all’interno del sistema» [Odum 1983, 11].

Consideriamo ora un sistema ancora più complesso (anzi, il più complesso!): un cervello in funzione. Esso può naturalmente essere analizzato sia come “tutto” - il sistema nella sua globalità - che come insieme organizzato di cellule - le “parti” organizzate. Essendo un sistema *biologico*, inoltre, gli emergentisti considereranno un terzo piano di analisi: la vita come emergenza. Infine, un cervello funzionante produce una mente. Essa, secondo gli emergentisti, è un secondo fenomeno emergente, associato al cervello in funzione, e come tale va analizzato sul piano delle emergenze.

Utilizzando la nozione di livello nell’accezione di piano di analisi, parleremo da adesso in poi di «livello psicologico» (o «psichico» o «mentale») per l’analisi dei fenomeni mentali, di «livello sistemico» (oppure «cerebrale» o «olistico» o «globale») per l’analisi del sistema nervoso e di «livello neurale» (o «cellulare» o «neuronale» o «meristico» o «mereologico») per l’analisi dell’organizzazione cellulare.

A questo punto si impongono tre osservazioni. La prima riguarda la teoria gerarchica nella formulazione datale dal primo emergentismo: il mondo si presenta come una gerarchia di livelli di complessità organizzazionale - gerarchia che include il livello fisico, quello chimico, quello biologico e quello *psicologico*. Questa formulazione, alla luce di quanto è stato detto finora, è ambigua. *Il livello psicologico non è un livello di complessità organizzazionale: è un livello di analisi*. La teoria gerarchica del *British Emergentism* non distingue l’accezione di «livello» come «“gradino” di una scala di organizzazione in un dato ordine di complessità» dall’accezione di esso come «piano di analisi».

La seconda osservazione concerne il problema della *causalità*. Abbiamo incontrato questo problema più di una volta e in varie vesti: la causazione verso il basso, la *quausazione* mentale, il principio di eredità causale e il concetto di vincolo. Nell’ambito della teoria gerarchica, finalmente, possiamo essere più espliciti su questa questione: la causazione in senso stretto, cioè l’azione fisica e causale, si muove verso l’alto e verso il basso attraverso i livelli di complessità organizzazionale - *ma non attraverso i livelli di analisi*. Non solo: ogni sistema appartenente a un dato livello di complessità organizzazionale può agire causalmente *in quanto tale* sui sistemi appartenenti al livello inferiore. Questa azione causale è la cosiddetta *quausazione*. In altre parole, fissato un livello (prima accezione) di complessità, sistemi appartenenti a

differenti livelli (seconda accezione) di organizzazione possono interagire causalmente tra loro, possono farlo in quanto tali e possono farlo sia verso l'alto che verso il basso.

La situazione cambia se utilizziamo la terza accezione di livello. In un computer, per esempio, non c'è alcuna azione causale tra il *software* e l'*hardware*: si tratta di descrizioni diverse del medesimo sistema e *le descrizioni non interagiscono causalmente*. Nel caso di un sistema nervoso in funzione la situazione è analoga: il livello psicologico è una *modalità descrittiva* applicabile a tale sistema, così come il software e l'hardware sono modalità descrittive applicabili a un computer che esegue un programma.

Si noti che non si sta affatto sostenendo che «la mente è *una descrizione dell'attività cerebrale*»: nessun emergentista afferma una cosa simile. Si sta invece sostenendo, seguendo gli emergentisti, che:

1. Un sistema nervoso in funzione (S) può essere analizzato prendendo in considerazione: il sistema come totalità, le sue parti e *ciò che è associato ad esso*;
2. La mente non è una descrizione di S, ma un fenomeno emergente (F) *associato* a S;
3. S può essere analizzato *prendendo in considerazione solo F*: questa modalità descrittiva viene definita «livello psicologico» (o «mentale», ecc.).

Ora, *le modalità descrittive non agiscono causalmente una sull'altra*, né verso l'alto né tantomeno verso il basso. Se i livelli di analisi sono modalità descrittive, allora non possono esserci rapporti causali tra essi.

Ma allora quali rapporti intercorrono tra essi?

Concentriamoci sul sistema nervoso (in funzione). Tra il livello sistemico e quello neurale intercorre la relazione che Morin chiama “vincolo” (avente, in questo contesto, lo statuto di concetto *filosofico*). In particolare,

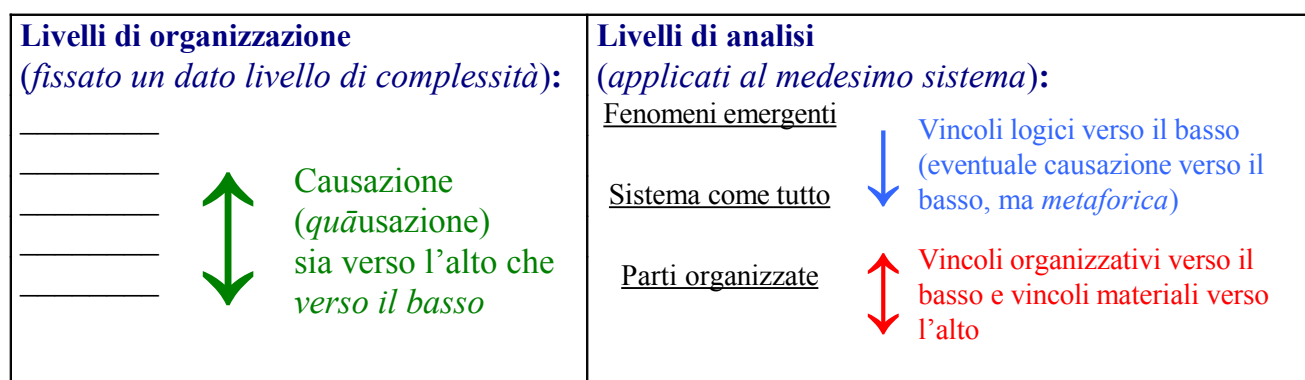
mentre i vincoli delle parti sul tutto sono anzitutto relativi ai caratteri materiali delle parti, i vincoli del tutto sulle parti sono anzitutto di carattere organizzativo [Morin 1977, 145].

Il concetto di vincolo può essere applicato anche al rapporto tra il livello mentale e quello cerebrale? Sì e no.

No, se per vincoli intendiamo vincoli organizzativi (del tutto sulle parti) o materiali (delle parti sul tutto): i processi cerebrali, come abbiamo già detto, non possono essere considerati “parti” dei fenomeni psicologici (né viceversa).

Si, se per vincolo intendiamo l'*implicazione logica*: la presenza di un fenomeno emergente implica, dal punto di vista logico, la presenza di un sistema da cui il fenomeno possa emergere. In altre parole, poiché ogni fenomeno emergente è associato *per definizione* ad un sistema, quest'ultimo è "vincolato" a funzionare in modo tale da far sì che si abbia l'emergenza. Abbiamo anche visto che, se ammettiamo la validità del principio di eredità causale di Kim (o di un principio equivalente), si possono attribuire alla mente dei poteri *metaforicamente* causali.

Concludiamo con il seguente schema che riassume quanto detto finora:



La *quausazione mentale*, nell'ambito di questa interpretazione della teoria gerarchica, viene completamente rigettata.

La terza e ultima osservazione riguarda il fatto (relativamente ovvio) che l'analisi condotta sui tre piani "organizzazione-globalità-emergenza" *non è l'unica possibile*. Consideriamo infatti un sistema di media complessità, per esempio un computer che sta eseguendo un programma. Senza abbandonare l'ottica emergentistica, possiamo analizzare questo sistema su ognuno dei seguenti piani, parzialmente corrispondenti a una *gerarchia di linguaggi*, dal "basso all'alto":

1. Il piano della "macchina fisica" (*hardware*): il linguaggio utilizzato per descrivere questa "macchina" parla di cavi, circuiti integrati, dischi magnetici, eccetera;
2. Il piano della "macchina logica": i componenti di questa macchina sono «*astrazioni logiche quali porte-or, invertitori e flip-flop (o, a un livello più alto di composizione multiplex, unità aritmetico-logiche e decodificatori di indirizzo)*» [Winograd - Flores 1986, 115];

3. Il piano del *software* (codificato in linguaggio macchina): ogni istruzione di questo linguaggio «è una semplice operazione di ricerca o di memorizzazione di un simbolo o di esecuzione di un'operazione logica o aritmetica (un confronto, un'addizione o una moltiplicazione)» [Winograd - Flores 1986, 116];
4. Il piano del *software* (scritto in linguaggio “di alto livello”): il linguaggio di alto livello è quello che è stato effettivamente usato per il programma che gira sul computer (può essere BASIC, FORTRAN, COBOL, LISP, ecc.);
5. Il piano dei fenomeni *innocentemente emergenti*: il computer, è vero, “fa solo ciò che gli si dice di fare”; ma abbiamo già visto che spesso «non si conoscono in anticipo i risultati di ciò che si dice a un calcolatore di fare» [Hofstadter 1979, 332]. Tali risultati imprevisi sono quelli che abbiamo chiamato fenomeni “innocentemente” emergenti;

Si noti che esistono anche piani di analisi *intermedi*: per esempio, tra il piano del linguaggio macchina e quello dei linguaggi di livello superiore, si colloca il piano di compilatori e interpreti⁸.

Faremo qualche altra osservazione sui livelli di analisi nelle Conclusioni.

7.3 La nozione di «simbolo» e la teoria computazionale della mente

Uno dei concetti più importanti e controversi della scienza cognitiva è quello di *simbolo*. L'interpretazione “classica” dei simboli è quella che fa riferimento ai simboli mentali, cioè alle rappresentazioni mentali postulate dalla “teoria computazionale della mente”. Avevamo introdotto la teoria computazionale quando, parlando della metafora della mente come *software*, avevamo visto che «l'analogia fra mente e calcolatore comporta [...] che si postulino simboli mentali» [Fodor 1981, 27]. Cercheremo ora di approfondire le tesi di tale teoria.

Secondo la teoria computazionale della mente, sostenuta tanto dai filosofi della mente funzionalisti quanto dai ricercatori dell'IA convenzionale:

⁸Per tutto quello che riguarda i piani di analisi di un calcolatore in funzione, si veda Hofstadter [1979, cap.X].

1. I costituenti *atomici* del pensiero sono simboli mentali, cioè rappresentazioni del mondo semanticamente interpretabili;
2. La cognizione è essenzialmente computazione, cioè *la mente è un sistema di manipolazione dei simboli mentali*.

Discende immediatamente da queste tesi che i simboli mentali hanno le seguenti tre proprietà:

- a) hanno “potere rappresentazionale”;
- b) sono “semanticamente interpretabili”;
- c) sono “manipolabili computazionalmente”.

E' opportuno soffermarsi brevemente su queste proprietà:

a) Il potere di rappresentare il mondo, da parte della mente, è stato implicitamente presente in molti dei modelli che abbiamo esaminato nel corso della tesi (ed è assolutamente centrale nei modelli dell'evoluzione della mente). Il concetto di rappresentazione è stato dunque accolto come se fosse aproblematico. In realtà è vero il contrario. Solo l'idea di base è relativamente semplice: la rappresentazione è ciò “con cui” si conosce qualcosa. A partire da quest'idea, però, si possono distinguere due accezioni di «rappresentazione», una più debole e una più forte.

Le due accezioni stanno rispettivamente “al di qua” e “al di là” del limite oltre il quale parlare di rappresentazioni comporta fare precise assunzioni ontologiche su ciò che può essere conosciuto per mezzo delle rappresentazioni stesse. Ora, mentre l'accezione debole è generalmente incontestata, quella forte è nel mirino di numerosi esponenti della terza cultura. E' il neurofisiologo Francisco J. Varela, il più autorevole epistemologo anti-rappresentazionista contemporaneo, l'autore che guida lo scontro contro l'accezione forte di rappresentazione. L'uso di quest'ultima implica, secondo Varela e colleghi, le seguenti tesi:

- 1) il mondo è prestabilito; 2) la nostra cognizione appartiene a questo mondo, anche se solo in una certa misura; 3) noi conosciamo questo mondo prestabilito rappresentando le sue caratteristiche e agendo quindi sulla base di tali rappresentazioni [Varela *et al.* 1991, 166].

Contro queste assunzioni, Varela sostiene che i sistemi cognitivi, invece di *rappresentare* un mondo prestabilito, *producono* un mondo come “dominio di distinzioni” inscindibile dalla loro stessa struttura incarnata (*embodied*) [Varela *et al.*

1991, 171]. Questa tesi, che è poi il nucleo dell'epistemologia costruttivista "radicale", deriva dall'interpretazione dei sistemi cognitivi come sistemi "operazionalmente chiusi". Si osservi però che ammettere la chiusura operativa dei sistemi cognitivi non implica *necessariamente* una posizione anti-rappresentazionista⁹.

Non è indispensabile, nel presente contesto, approfondire ulteriormente il problema della rappresentazione: ai fini della nostra discussione possiamo infatti adoperare la sua innocua accezione *debole* (cioè non fare assunzioni ontologiche).

b) Cosa significa che i simboli mentali sono "semanticamente interpretabili"? Significa che essi:

- corrispondono a concetti del linguaggio naturale (*tazza, gatto, uguaglianza, ecc.*);
- ricorrono *ogni volta* che lo stato mentale in questione è *descrivibile* utilizzando tali concetti.

L'interpretabilità semantica dei simboli mentali rende un sistema cognitivo "semanticamente trasparente". Si ha dunque una netta corrispondenza biunivoca tra gli stati interni di un sistema semanticamente trasparente e i concetti del linguaggio naturale adoperati per descrivere il comportamento del sistema [Clark 1989, 31]. Si noti che *i simboli mentali non possono dipendere dal contesto*, perché - se così fosse - gli stessi simboli non potrebbero ricorrere *ogni volta* che il sistema cognitivo si comporta in modo tale da rendere descrivibile il proprio stato mediante i concetti corrispondenti a tali simboli.

c) I simboli mentali postulati dalla teoria computazionale della mente sono "manipolabili computazionalmente". Questa proprietà è particolarmente importante perché traccia una separazione tra il concetto cognitivista classico di simbolo e tutte le concezioni alternative.

Affermare che i simboli mentali sono manipolabili computazionalmente significa:

⁹La chiusura operativa dei sistemi cognitivi viene riconosciuta anche da convinti rappresentazionisti (magari con altre parole): è un fatto ben noto, infatti, che l'attività neuronica è *prodotta da e produce solo* altra attività neuronica [Varela 1985, 146-7]. Da ciò segue naturalmente che i simboli mentali interagiscono esclusivamente in base alla loro *logica interna*. *Tuttavia*, secondo i rappresentazionisti, «*tale logica, benché operi senza mai consultare il mondo esterno, produce [...] un modello abbastanza fedele del modo in cui funziona il mondo e riesce così a mantenere i simboli più o meno "in fase" con il mondo che si ritiene essi rispecchino*» [Hofstadter 1981, 190].

- attribuire loro una sintassi e una semantica combinatoriali («*Combinatorial syntax and semantics*» [Fodor - Pylyshyn 1988, 12]);
- sostenere che i processi mentali dipendono da tale struttura combinatoriale («*Structure sensitivity of processes*» [Fodor - Pylyshyn 1988, 13]).

Attribuire ai simboli mentali una sintassi e una semantica equivale a postulare un vero e proprio *linguaggio del pensiero*. Ciò non deve stupire: per gli scienziati cognitivi, infatti, «*pensare è come parlare*» [Haugeland 1985, 85]. Perché mai? In che cosa consiste la somiglianza tra pensieri e parole? Secondo il filosofo della mente John Haugeland, la somiglianza si basa sulla *applicabilità a entrambi della distinzione tra modo e contenuto*;

Consideriamo quel che hanno in comune le seguenti asserzioni:

- [affermo:] Fido mangia gli insetti
- [chiedo:] Fido mangia gli insetti?
- [ordino:] Per favore, fa' mangiare a Fido gli insetti
- [mi scuso:] Mi dispiace che Fido mangi gli insetti

Tutte riguardano il fatto che Fido mangi gli insetti; ovvero, per dirla con i filosofi, hanno in comune il contenuto proposizionale “[che] Fido mangi[a] gli insetti”. Ma ciascuna asserzione propone questo contenuto in modo diverso, affermando che le cose stanno così, chiedendo se le cose stanno così, ordinando che le cose vadano così ecc. Queste sono differenze di modo; quindi lo stesso contenuto può essere proposto in modi diversi. [...] Sembra (questo è il punto) che un’analoga distinzione fra modo e contenuto si possa fare anche per il pensiero: posso credere, desiderare o dispiacermi (= differenze di modo) che Fido mangi gli insetti, che Fritz beva il latte, o che gli smeraldi siano costosi (= differenze di contenuto). E questo suggerisce la presenza di un collegamento più stretto fra un atto linguistico e il pensiero che esprime: cioè che debbano avere lo stesso contenuto e modi corrispondenti [Haugeland 1985, 87].

Se si postula - come fa, per esempio, Fodor - un linguaggio del pensiero, allora la distinzione modo/contenuto si traduce, molto naturalmente, in quella sintassi/semantica.

Le caratteristiche sintattiche e semantiche dei simboli mentali sono, per ipotesi, le seguenti [cfr. Fodor - Pylyshyn 1988, 12]:

- i simboli possono essere atomici o molecolari;
- i simboli molecolari sono composti da simboli che sono atomici o essi stessi molecolari;

– il contenuto dei simboli molecolari è una funzione dei contenuti dei suoi costituenti.

Tutte queste caratteristiche strutturali, secondo gli scienziati cognitivi, corrispondono a reali strutture fisiche nel cervello («*real physical structures in the brain*» [Fodor - Pylyshyn 1988, 13]); in altre parole, la struttura combinatoriale dei simboli mentali ha, per ipotesi, una controparte nelle relazioni strutturali tra le proprietà fisiche del cervello (!).

Sostenere che i processi mentali dipendono da questa struttura combinatoriale equivale ad affermare che *i processi mentali sono formali*, cioè operano su ogni simbolo che soddisfa una data descrizione strutturale e lo trasformano in un simbolo che soddisfa un'altra descrizione strutturale [Fodor - Pylyshyn 1988, 13]. Due idee sorreggono questa convinzione: nel linguaggio del pensiero, analogamente a quanto accade nei linguaggi formali della logica matematica (ma non nel linguaggio naturale), la sintassi codifica la semantica («*the syntax of a formula encodes its meaning*» [Fodor - Pylyshyn 1988, 28]); inoltre, il modo in cui i processi mentali operano sui simboli mentali è *sensibile alla sintassi* [Fodor - Pylyshyn 1988, 30].

Tenendo conto delle proprietà (b) e (c) dei simboli mentali, è corretto affermare che la teoria computazionale della mente postula una sorta di isomorfismo come il seguente:



In base a questo isomorfismo, per esempio, la relazione di “appartenenza” tra un simbolo atomico e uno molecolare (i quali corrispondono a un sotto-concetto incluso in un concetto) ha un suo analogo in qualche relazione fisica tra le sottostrutture del cervello [Fodor - Pylyshyn 1988, 13].

Questa è, in sostanza, la teoria computazionale della mente. Esposta nei termini in cui lo abbiamo fatto, essa può apparire un po’ astratta. Cerchiamo allora di individuarne le conseguenze “concrete”.

Le conseguenze sono molto semplici: se la mente è un sistema semanticamente trasparente come quello sopra descritto, cioè un elaboratore simbolico le cui operazioni sono puramente formali, allora la sua somiglianza con un computer

programmato è davvero profondissima. Così profonda che menti e computer programmati possono essere inclusi nella medesima e circoscritta classe di sistemi: i «*sistemi simbolici fisici*» [Newell - Simon 1981, 47].

L'implicazione *più importante* di tutta la teoria è che i programmi dell'IA simbolica, scritti in linguaggi di alto livello, possono essere considerati *modelli della cognizione*¹⁰. Quest'ultima affermazione non va confusa con la tesi dei *descrittivisti*, secondo i quali i programmi dell'IA convenzionale possono essere considerati solo *modelli di come noi descriviamo la cognizione* [cfr. Clark 1989, 207-11]. Facciamo un esempio.

Consideriamo la seguente attività cognitiva: la formulazione di una risposta "adeguata" a qualunque frase pronunciata da un interlocutore nell'ambito di un dominio limitato. Per rendere più chiaro l'esempio, supponiamo che l'interlocutore sia il paziente di uno psicoterapeuta rogeriano (non direttivo) e il dominio quello di una seduta terapeutica. Descrivere una simile attività mentale non è difficile. Una descrizione relativamente soddisfacente potrebbe infatti essere questa [cfr. Filippazzi 1989, 39-40]:

1. La frase dell'interlocutore (il paziente) viene analizzata per cercarvi una parola (o una locuzione) chiave. Un certo numero di parole chiave, eventualmente ordinabili per importanza, deve essere presente in memoria. Possono verificarsi due casi: la frase del paziente contiene una o più parole chiave, oppure non ne contiene nessuna;
2. Nel primo caso, alla parola chiave più importante viene associata una risposta "adeguata" parziale. E' chiaro che deve esserci in memoria un insieme di risposte adeguate parziali per ogni parola chiave;
3. Una parte della frase del paziente viene coniugata opportunamente e aggiunta alla risposta adeguata parziale: si ottiene così la risposta adeguata completa da dare al paziente.

¹⁰Questa è l'implicazione *più importante*, almeno nell'ambito del discorso che stiamo conducendo. Non è, tuttavia, l'implicazione *più forte*. Quest'ultima è sicuramente l'ipotesi secondo cui tutti (!) i sistemi simbolici fisici sono «*in possesso degli strumenti necessari e sufficienti per l'azione intelligente generale*» [Newell - Simon 1981, 49], ovvero l'ipotesi secondo cui un computer programmato potrebbe essere *veramente* intelligente [cfr. Clark 1989, 23].

4. Nel caso in cui non vi siano parole chiave nella frase del paziente, la risposta adeguata viene estratta da un repertorio di risposte interlocutorie (anch'esso deve essere presente in memoria): «Si spieghi meglio», «Davvero?», ecc.

E' molto facile tradurre questa procedura in uno dei linguaggi di programmazione di alto livello, anzi: è stato fatto. Nel 1966 l'informatico Joseph Weizenbaum ha presentato un programma, chiamato ELIZA, che si basava proprio su tale algoritmo [Filippazzi 1989].

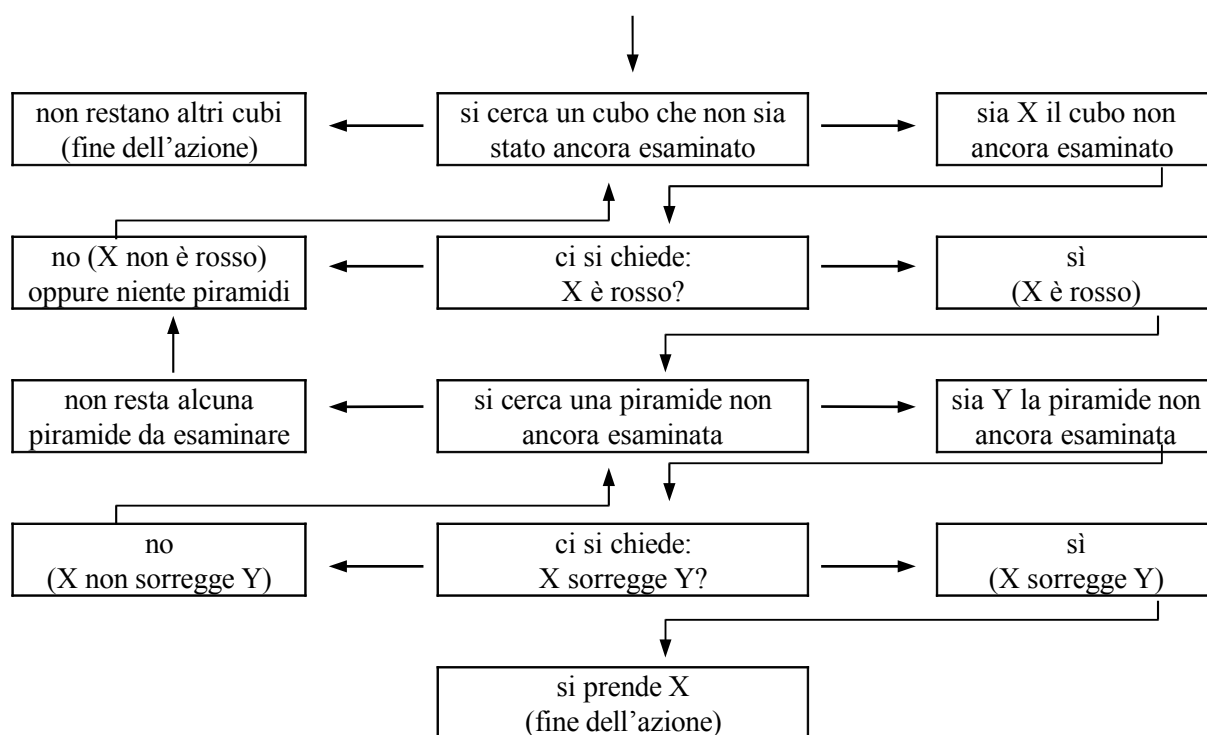
Per quanto sopra, il programma ELIZA è certamente un modello *della nostra descrizione* (non l'unica possibile) dell'attività cognitiva corrispondente alla formulazione di una risposta "adeguata". Ma ELIZA è anche un modello *dell'attività cognitiva stessa?*

Se definiamo "di alto livello" ogni descrizione *traducibile in un linguaggio di programmazione di alto livello*, possiamo anche riformulare la questione e dire: certamente ELIZA è un modello dell'attività mentale analizzata ad alto livello; ma questo livello di analisi è il livello "più giusto"? Se la mente è davvero un sistema semanticamente trasparente di manipolazione simbolica, allora la risposta deve essere affermativa¹¹.

Prima di sollevare delle obiezioni, facciamo un secondo esempio. Consideriamo la seguente attività mentale: la pianificazione di un'azione in un dominio limitato. Supponiamo che il dominio limitato sia il cosiddetto "mondo dei blocchi", cioè uno spazio virtuale costituito esclusivamente da un piano (di dimensioni limitate) sul quale si trovano alcuni solidi aventi geometrie semplici: cubi, parallelepipedi e piramidi.

¹¹Si noti che, nonostante la semplicità dell'algoritmo, le prestazioni di ELIZA sono sorprendenti. Quando il programma venne fatto circolare nelle università statunitensi, le reazioni furono così incredibili che Weizenbaum le raccolse in un libro, *Computer Power and Human Reason* (1976). «Egli rimase colpito da come persone non sprovvedute si lasciassero coinvolgere dalle conversazioni con ELIZA fino al punto da scordare la sua natura e attribuirle una vera e propria personalità. Weizenbaum racconta, ad esempio, come la sua segretaria, pur sapendo che si trattava solo di un programma, cominciò ad avere con esso un rapporto emotivo, al punto da chiedere di essere lasciata sola nella stanza del computer per parlare con ELIZA di argomenti strettamente privati. Weizenbaum scoprì, inoltre, che ELIZA veniva consultata nottetempo da molte persone usando i terminali distribuiti nei campus. [...] Un altro fatto che lo lasciò sconcertato fu che alcuni docenti di psichiatria attribuirono al programma un effettivo valore terapeutico, al punto da proporre che una versione estesa e perfezionata fosse messa a disposizione dei pazienti nelle cliniche per malattie mentali» [Filippazzi 1989, 42].

La pianificazione di un'azione, per esempio “prendere un cubo rosso che sorregge una piramide”, potrebbe essere descritta per mezzo di un diagramma di flusso come il seguente [Waltz 1982, 68]:



Anche questa volta, il diagramma può essere facilmente tradotto in un linguaggio di programmazione di alto livello e, anche in questo caso, qualcuno lo ha fatto: si tratta dell'informatico ed epistemologo Terry A. Winograd e il programma è il famoso e impronunciabile SHRDLU.

Analogamente a prima, domandiamoci: SHRDLU è un modello della pianificazione o solo della descrizione (non l'unica possibile) che possiamo dare di essa? O equivalentemente: il livello di analisi di SHRDLU è quello “più giusto”? Di nuovo, i sostenitori della teoria computazionale della mente devono rispondere affermativamente.

Questa volta, però, dovrebbero immediatamente sorgere dei seri dubbi circa questa risposta. Quando la mente (immersa nel mondo dei blocchi) cerca il cubo o la piramide, come fa a individuarli? Com'è possibile che l'operazione di riconoscimento, cioè la *percezione*, non compaia nel diagramma? SHRDLU è programmato in modo da registrare internamente le proprietà e la posizione dei blocchi, quindi “sa” sempre dove si trovano gli oggetti del suo mondo; ma la mente, a differenza di SHRDLU, deve

percepirli. Cosa succede quando percepiamo qualcosa o qualcuno? Non è facile rispondere.

Ecco il punto: non è semplice descrivere le operazioni mentali “automatiche”, cioè quelle che, come la percezione, richiedono tempi dell’ordine del decimo di secondo o meno. Ora, se i programmi dell’IA convenzionale sono modelli *di come noi descriviamo la cognizione*, allora la “linea dei 100 millisecondi” non costituisce un problema: è ovvio che le operazioni mentali automatiche non vengano scomposte e analizzate a loro volta.

Ma, se i programmi dell’IA convenzionale sono modelli *della cognizione* (questa è l’implicazione più importante della teoria computazionale), allora la mente - coerentemente con tali modelli - non compie alcuna operazione *cognitivamente rilevante* al di sotto della “linea dei 100 millisecondi”. Per usare le parole di Herbert Simon (1980):

Ogni cosa interessante nella cognizione avviene al di sopra del livello dei 100 millisecondi - il tempo che occorre per riconoscere la propria madre.¹²

7.4 Il risveglio dal “sogno booleano”

Che cosa c’è al di sotto della linea dei 100 millisecondi? *Hic sunt leones*. Gli scienziati cognitivi “classici” non si sono mai avventurati al di sotto di tale linea. Si può dire che i ricercatori della terza cultura siano stati i primi a superare tale confine e che, in parte, la loro avventura intellettuale consista proprio nell’esplorazione di quegli aspetti della mente che, quando si sommano e superano la linea dei 100 millisecondi, subiscono una completa trasfigurazione e, trasfigurati, appaiono più facilmente studiabili.

Cerchiamo prima di tutto di identificare i processi automatici della mente che, da adesso in poi, chiameremo “subcognitivi”¹³. Abbiamo già detto che il

¹²«Everything of interest in cognition happens above the 100-millisecond level - the time it takes you to recognize your mother» (citazione a memoria di Hofstadter [1982b, 632]).

¹³Potremmo forse chiamarli subconsci, ma occorrerebbe ribadire continuamente che non si tratta di una classificazione psicoanalitica.

riconoscimento percettivo è uno di essi. Quali altre attività possono essere considerate subcognitive?

Essenzialmente, tutte quelle “facili”: le analogie, i *lapses*, le intuizioni. Esse si contrappongono alle attività “difficili” - tipiche del pensiero analitico, logico-formale e matematico - che chiameremo attività “cognitive”. La differenza tra le due è molto netta: le prime sono quelle che non creano problemi a un infante umano ma che sono impossibili per le macchine (quelle tradizionali); al contrario, per le seconde, la mente umana necessita di un lungo apprendistato e ciò nonostante non riesce mai a raggiungere le prestazioni di un calcolatore programmato [Varela *et al.* 1991, 112-3].

Per dirlo con altre parole: le attività subcognitive sono presenti tanto nelle menti più complesse quanto in quelle più semplici, ma la cognizione richiede quelle più complesse. Più semplici qui significa: menti come quelle dei neonati, oppure come quelle dei primi ominidi. Più complesse qui significa: menti come quelle degli adulti, oppure come quella dell'*Homo sapiens* attuale. Ora, se i processi subcognitivi (ma non quelli cognitivi) caratterizzano anche le menti più semplici, allora la natura della mente umana - che, come abbiamo già detto, non dipende dalla complessità - avrà certamente a che fare più con l'attività subcognitiva che con quella cognitiva.

L'idea di un legame indissolubile tra natura della mente e processi subcognitivi ha cominciato a circolare alla fine degli anni Settanta. Si è rivelato un *meme* estremamente contagioso, capace di imporsi all'attenzione dell'intera comunità degli scienziati cognitivi. All'improvviso la teoria computazionale della mente è apparsa forzata e innaturale, la metafora della mente come software ha perso la sua efficacia, molti giganti del pensiero cognitivista sono stati buttati giù dai loro piedistalli.

Nell'IA, forse più che in qualunque altra disciplina, la trasformazione - tuttora in corso - si è avvicinata a una vera e propria “rivoluzione scientifica” (in senso kuhniano). Spezzato il legame con la linguistica generativa, l'IA ha rivolto lo sguardo verso discipline come la fisiologia e la neurobiologia¹⁴.

¹⁴L'attacco dell'informatico Roger Schank contro Noam Chomsky riassume con efficacia l'enfasi con cui questa piccola rivoluzione si sta compiendo: «Chomsky è disposto a studiare la mente solo se si comporta secondo principi ben formulati. Questo proprio non lo capisco: a me interessa studiare la mente qualunque cosa essa sia; fosse anche fango, mi andrebbe bene lo stesso. Molti scienziati vorrebbero che la mente fosse in tutto e per tutto un dispositivo razionale. Se non lo è, se le sue strutture non sono chiare e matematiche, non ne vogliono nemmeno sentir parlare. Chomsky ha sempre adottato la filosofia della scienza dei fisici, secondo cui bisogna procedere per ipotesi e verifiche. L'epistemologia dell'Intelligenza Artificiale, invece, è più vicina alle scienze biologiche. I

In questo paragrafo cercheremo di mettere in evidenza alcuni aspetti di questo cambiamento nell'IA.

Parte del processo è, in realtà, un *ritorno* a idee sorte quando i tempi non erano del tutto maturi. Prima che il cognitivismo facesse “piazza pulita” di tutte le concezioni a esso alternative, infatti, molti pensatori stavano costruendo teorie sul cervello e sulla mente vicinissime a quelle in àuge nella terza cultura. In particolare, negli anni Quaranta, era convinzione diffusa che, prima di poter studiare la mente, occorresse costruire un modello del cervello. Molti ricercatori, impegnati in tal senso, avevano conseguito notevoli risultati.

Nel 1943, per esempio, Warren McCulloch e Walter Pitts avevano già ideato le reti di neuroni logico-formali che oggi chiamiamo *reti neurali*. L'idea di base delle reti neurali è quella secondo cui si può produrre un modello del cervello anche ignorando tutte le complicazioni fisiologiche e chimiche dei neuroni reali:

Un neurone formale sarà indicato da un cerchio, che simboleggia il corpo del neurone, e da una linea che ne esce, che è l'assone. Una freccia serve a indicare che l'assone di un neurone è incidente al corpo di un altro neurone; ogni neurone formale ha due stati, di eccitazione e di non eccitazione, non ulteriormente specificati, che trasmette agli altri a cui è collegato. Ognuno riceve da altri da cui dipende eccitazioni positive o negative con pesi, e a sua volta trasmette a quelli a cui è collegato eccitazioni positive o negative con pesi. I neuroni si attivano solo se l'eccitazione totale calcolata con una funzione aritmetica delle eccitazioni in arrivo supera una certa soglia. Il funzionamento di queste reti è sincrono, a intervalli regolari discreti, i neuroni servono da interruttori logici; una volta aggiunta la memoria si ottiene l'equivalente della macchina di Turing universale [Lolli 1995, 252].

Le reti neurali destarono l'interesse di quel gigante del pensiero scientifico che è stato Johann (John) von Neumann, inventore dei computer digitali e di numerose discipline di matematica pura e applicata (tra cui la teoria dei giochi)¹⁵. Egli fu colpito specialmente dall'equivalenza tra le reti neurali formali e la macchina di Turing universale, anche se si rese conto che il problema della memoria era un vero e proprio

biologi partono dal presupposto che gli organismi sono così come appaiono e vanno studiati per quello che sono. Se un modello non va d'accordo con la realtà, va aggiustato e corretto. L'intelligenza artificiale condivide con la biologia uno sguardo empirico sul mondo e l'amore per l'esplorazione e la scoperta» [Schank 1995, 158].

¹⁵Peccato non poter dire che egli fu un gigante del pensiero *tout court*. Oggi si sa che von Neumann fu «una figura chiave nella progettazione e nella costruzione della bomba atomica, e nel successivo sviluppo della bomba all'idrogeno e della politica nucleare degli USA. Nella sua qualità di membro della Commissione per l'energia atomica - poco prima della sua morte, nel 1955 - raccomandò a Eisenhower un attacco nucleare preventivo contro l'Unione Sovietica» [Varela 1985, 142].

“tallone d’Achille” del modello di McCulloch e Pitts (d’altra parte, come ricordava von Neumann, la memoria è un mistero *anche nel cervello*).

Dall’equivalenza tra reti neurali e macchina di Turing, comunque,

von Neumann trae la seguente lezione: qualcuno sostiene che le attività e le funzioni del sistema nervoso umano sono così complicate che nessun meccanismo le potrebbe svolgere, e si è anche tentato di indicare specifiche funzioni la cui natura presenterebbe tale limitazione; «si è tentato di mostrare che tali specifiche funzioni, logicamente e completamente descritte, sono per sé incapaci di realizzazione meccanica, neurale. Il risultato di McCulloch e Pitts mette fine a questi tentativi. Esso prova che ogni cosa che può essere descritta in modo esaustivo e privo di ambiguità, ogni cosa che può essere messa in parole in modo completo e non ambiguo, è *ipso facto* realizzabile da una opportuna rete neurale finita» [Lolli 1995, 255].

A partire dal 1946, McCulloch e von Neumann - insieme al “padre” della cibernetica, Norbert Wiener, e a quello dell’informatica, Claude Shannon - organizzarono la serie di seminari più celebre del secolo: le Macy Conferences.

Nel corso delle Macy Conferences, a cui parteciparono i più importanti pensatori del tempo (Margaret Mead, Gregory Bateson, Heinz von Foerster, oltre a quelli citati sopra), si continuò ad affrontare il problema della mente *parlando del cervello*:

ci fu un’approfondita discussione sul fatto che nel cervello reale sembra che non ci siano regole, né un’unità centrale di elaborazione logica, né infine un sistema di archiviazione preciso dell’informazione [Varela *et al.* 1991, 111].

Insomma, ancora a metà secolo il cervello era al centro di ogni dibattito sui fenomeni mentali; la situazione mutò radicalmente negli anni dell’“ortodossia cognitivista”, durante i quali i computer di von Neumann - *e non il cervello* - furono la fonte di ogni ispirazione per la filosofia della mente.

La storia delle reti neurali continua con la nascita del *perceptrone*, una rete neurale con uno strato di connessioni, ideata da Frank Rosenblatt nel 1958. Il perceptrone possiede un algoritmo di “apprendimento” (la regola delta) che modifica le connessioni tra i neuroni logici, in modo tale da rendere la rete sensibile a certe configurazioni: il perceptrone, in questo modo, può “imparare” per esempio a risolvere il problema logico della disgiunzione (*vel*). Nonostante questa e altre proprietà interessanti, il futuro del perceptrone non fu roseo: nel 1969, uno dei fondatori dell’IA,

il matematico e informatico Marvin Minsky, insieme al suo collega Seymour Papert, esaminò accuratamente i perceptron e ne mise in luce una serie di limiti fondamentali. Inoltre, Minsky e Papert si dissero pessimisti sulla *versione a più strati delle reti neurali*:

Il *perceptron* [...] ha molte proprietà attraenti: la sua linearità, il suo affascinante teorema d'apprendimento, la sua limpida ed esemplare semplicità come modello di computazione parallela. Non c'è ragione di supporre che nessuno di questi pregi sia ereditato dalla versione a più strati. Nondimeno, consideriamo un importante programma di ricerca la giustificazione (o la confutazione) del nostro giudizio intuitivo sulla sterilità di questa estensione.¹⁶

La ricerca sulle reti neurali subì così una battuta d'arresto. Nel frattempo, la scienza cognitiva stava nascendo e rapidamente andava consolidandosi: in pochissimi anni l'approccio cognitivista divenne il protagonista indiscusso dell'indagine sulla mente¹⁷.

Solo alla fine degli anni Settanta la filosofia funzionalista della mente e l'IA simbolica, accomunate dalla fiducia nella teoria computazionale della mente, cominciarono a subire critiche più o meno sistematiche. Le critiche si basavano

su due lacune, ampiamente riconosciute, del cognitivismo. La prima è che l'elaborazione dell'informazione simbolica è basata su regole *sequenziali*, che vengono applicate una alla volta. [...] Una seconda limitazione importante sta nel carattere localizzato dell'elaborazione simbolica: la perdita o il cattivo funzionamento di una porzione qualunque dei simboli o delle regole del sistema produce una grave disfunzione [Varela *et al.* 1991, 112].

Queste considerazioni spingevano nella direzione di sistemi computazionali basati, come già le reti di McCulloch e Pitts, sull'effettiva struttura del cervello e sulle modalità del suo funzionamento reale. Il sistema nervoso, infatti, non si basa su regole sequenziali, né il suo funzionamento dipende in maniera cruciale dalle singole cellule. Fu così che i ricercatori tornarono a occuparsi delle reti neurali, le quali - in ciò essendo del tutto simili al cervello - non presentano nessuno dei due limiti dell'elaborazione simbolica.

¹⁶Minsky M. - Papert S. [1969], *Perceptrons*, MIT Press, Cambridge 1969, pp.231-2; citati in Rumelhart *et al.* [1986a, 252].

¹⁷Naturalmente, questa storia è semplificata. In realtà alcuni ricercatori isolati continuarono a studiare le reti neurali: tra loro vanno menzionati almeno John Holland, Teuvo Kohonen, Stephen Grossberg e James Anderson.

All'inizio degli anni Ottanta, varie piste intellettuali, che si allontanavano sempre più dal cognitivismo classico, stavano infine per convergere:

- un eterogeneo gruppo di pensatori (tra cui Mario Bunge ed Edgar Morin), come abbiamo visto, aveva coltivato il concetto di emergenza - allontanandolo dall'interpretazione "quasi-sostanziale" che ne aveva dato Popper - e lo stava già applicando alla mente;
- alcuni ricercatori, particolarmente brillanti, stavano studiando il modo di applicare alle reti neurali gli strumenti e le tecniche matematiche utilizzate dai fisici nello studio di alcuni sistemi dinamici complessi con proprietà anomale (i cosiddetti "vetri di *spin*");
- la nascente scienza della complessità, come sappiamo, stava cominciando a simulare i sistemi dinamici complessi e ai fenomeni simulati applicava sia il concetto di auto-organizzazione (mutuato dalla neocibernetica) che quello di emergenza;
- molti ricercatori dell'IA convenzionale, infine, stavano prendendo coscienza dei limiti della teoria computazionale della mente ed alcuni cominciavano anche a parlare della mente come di un fenomeno emergente.

Tutte queste piste si incontrarono. Tutti i protagonisti delle suddette ricerche furono influenzati e stimolati dal lavoro di tutti gli altri. E tutti, alla fine, *applicarono alla mente il concetto di emergenza*.

Nel 1982 John J. Hopfield pubblicò un articolo nel quale mostrava come applicare a certi tipi di reti neurali (quelle che oggi chiamiamo "reti di Hopfield") le leggi fisiche e le equazioni matematiche che i fisici utilizzano per trattare i "vetri di *spin*" - sostanze amorfe rese interessanti da certe proprietà caratteristiche. Lo *spin*, come forse è noto, è uno dei parametri fisici che determinano lo stato microscopico di una particella (è il quarto numero quantico) e può assumere solo due valori. Ora, in condizioni standard, una sostanza amorfa

è tale che gli *spin* dei suoi elementi costitutivi sono distribuiti nel più completo disordine. Si sa che un certo tipo di ordine può tuttavia essere introdotto a forza nel sistema immergendo quest'ultimo in un campo magnetico. Ci si aspetta, in tal caso, che gli *spin* si riorganizzino in modo tale da far prevalere uno dei loro due valori possibili. Nelle sostanze che si chiamano vetri di *spin*, invece, questa riorganizzazione viola alcune aspettative [Bellone 1992, 115].

John Hopfield chiamò “emergenti” le proprietà dei vetri di spin e del proprio modello di reti neurali. Il suo articolo, intitolato significativamente “Neural Networks and Physical Systems with Emergent Collective Computational Abilities”, gettò un ponte tra la teoria delle reti neurali e il concetto di emergenza. I *connessionisti*, come cominciarono a chiamarsi - proprio a partire dal 1982 - gli studiosi delle reti neurali, in seguito hanno continuato ad applicare il concetto di emergenza ai fenomeni mentali simulati dai loro modelli.

Nel corso di quello stesso anno, lo scienziato cognitivo Douglas R. Hofstadter, già Premio Pulitzer per *Gödel, Escher, Bach* (1979), esaminò accuratamente tutte le differenze tra i programmi di alto livello dell’IA simbolica e le fluide modalità con cui opera la mente umana. Un suo saggio del 1982, ripubblicato nel 1985 con il titolo “Waking Up from the Boolean Dream”, può essere considerato il “manifesto” dell’idea secondo cui la natura della mente umana non è tanto caratterizzata dai processi cognitivi, quanto da quelli subcognitivi. Cercare di catturare la dinamica dei pensieri con l’elegante formalismo dei linguaggi di programmazione di alto livello è, secondo Hofstadter, un “sogno” simile a quello che condusse George Boole, nel 1854, a chiamare «leggi del pensiero» le regole della nascente logica matematica (che egli stesso stava creando). La cognizione, cioè l’attività mentale di alto livello, è generata dalla subcognizione, cioè l’attività mentale “di basso livello” (al di sotto della linea dei 100 millisecondi) - questa è l’essenza di ciò che Hofstadter chiama «*statistically emergent mentality*» [Hofstadter 1982b, 654]: la mente come emergenza statistica.

Nel 1984, nel New Mexico, Murray Gell-Mann fondò il Santa Fe Institute. Nel giro di pochissimi anni, il Santa Fe divenne l’istituto di riferimento di tutti gli scienziati della complessità. E, come abbiamo visto esaminando l’opera di Stuart Kauffman, gli scienziati del Santa Fe fanno ampio uso del concetto di emergenza e del concetto di auto-organizzazione. Scienza della complessità ed emergenza si erano indissolubilmente legate. In quello stesso anno, a Milano, una serie di pensatori (per lo più non statunitensi) si riunì intorno a Edgar Morin. Egli lanciò la sua “sfida della complessità” (questo il titolo degli incontri), rivolta sia a filosofi che a scienziati, parlando approfonditamente di organizzazione, auto-organizzazione, emergenza. Anche filosofia della complessità ed emergenza si erano ormai legate.

Nello stesso periodo in cui a Milano veniva lanciata la “sfida della complessità”, cioè nell’autunno del 1984, Hofstadter si trasferì dal MIT (Massachusetts Institute of Technology) all’Università del Michigan. Fu lì che egli fondò il FARG (Gruppo di Ricerca sulle Analogie Fluide) e cominciò a raccogliere intorno a sé tutti gli scienziati cognitivi che, come lui, giudicavano l’IA convenzionale un “sogno booleano” dal quale risvegliarsi al più presto.

Nel frattempo, un gruppo di connessionisti dell’Università della California (a San Diego), coordinato da due psicologi, David E. Rumelhart e James L. McClelland, e soprannominatosi “gruppo PDP” (gruppo di ricerca sul Processamento Distribuito in Parallelo), aveva riscoperto un potente algoritmo di “apprendimento” per le *reti neurali estese a più strati di connessioni*, inventato nel 1974 da Paul J. Werbos ma passato inosservato [Hinton 1992, 118]. Rumelhart e colleghi si resero conto che l’algoritmo di Werbos, la regola delta generalizzata, confutava il giudizio intuitivo di Minsky e Papert sulla sterilità dell’estensione delle reti. Nel 1986 il gruppo PDP pubblicò una raccolta di saggi su vari aspetti della ricerca sulle reti neurali, tra cui una descrizione della regola delta generalizzata. I tempi, questa volta, erano maturi: della pubblicazione del gruppo PDP furono vendute 6000 copie *il giorno stesso in cui venne messa sul mercato* [Rose 1992, 108]. Come già Hopfield, anche il gruppo PDP considerava “emergenti” le proprietà delle reti neurali.

Nei dieci anni successivi agli ultimi eventi qui riportati, lo studio della complessità e dell’IA connessionista hanno conosciuto un *boom* incredibile. Un esame esaustivo dei risultati conseguiti dai ricercatori che hanno operato in tali campi è oltremodo al di là della nostra portata. Pertanto, in questa sede ci limiteremo a presentare solo quei modelli con caratteristiche *realmente* interessanti per la filosofia della mente.

Tali modelli, diversamente da come si potrebbe immaginare, non sono numerosi. Questo è dovuto essenzialmente alle seguenti ragioni:

- a) le reti di Kauffman e gli altri dispositivi studiati al Santa Fe vengono oggi usati per simulare l’evoluzione di sistemi complessi di importanza basilare per l’evoluzionismo, ma senza alcuna rilevanza psicologica immediata;
- b) le reti neurali più recenti, a differenza di quelle elaborate dal gruppo PDP (che era guidato da due *psicologi*), vengono concepite o come modelli neurobiologici veri e

propri oppure come “sistemi esperti”: l’indagine sulle loro (eventuali) capacità mentali è praticamente ferma;

- c) nel campo dell’IA, l’approccio del FARG allo studio dei processi cognitivi tramite quelli subcognitivi non è ancora diventato epidemico: in verità, il gruppo di Hofstadter è per adesso l’unico ad aver scritto programmi ad “architettura emergente”.