

MODELLI DI RETI NERVOSE E FUNZIONAMENTO CEREBRALE

Renato Nobili

*Dipartimento di Fisica 'G.Galilei'
via Marzolo 8 — 35131 PADOVA*

1. Il sogno di Boole.

Per quasi tutto il quarantennio successivo all'apparizione dell'articolo di S.W.McCulloch e W.Pitts del 1943 *A Logical Calculus of Ideas Immanent in Nervous Activity*, il neurone è stato considerato come un *decisore a soglia*, ossia un dispositivo che genera in uscita un segnale a gradino quando una somma di segnali positivi o negativi, applicati in ingresso, supera un certo valore. Durante quel periodo, principalmente a causa dell'esiguità della conoscenza sperimentale, la complessità dei processi nervosi era generalmente sottostimata e il ruolo funzionale del neurone poteva apparire tanto elementare quanto quello di una valvola termoionica o, più tardi, di un transistor. Ignorando il tessuto di sostegno e alimentazione della rete nervosa (la glia), i neuroni si presentano come gli unici componenti del sistema nervoso. Era perciò conseguente ritenere che i processi mentali siano essenzialmente riconducibili all'interazione di quelle semplici unità funzionali (M.Arbib, 1965). L'analogia coi circuiti elettronici pareva tanto più stretta in quanto si riteneva che i segnali nervosi si propagassero elettricamente attraverso i contatti sinaptici, intesi come semplici punti di saldatura tra cellule nervose. La localizzazione della memoria appariva invece problematica: si è ritenuto che l'informazione nervosa fosse memorizzata sotto forma di valori delle soglie neuronali o codificata come il DNA in strutture molecolari; l'ipotesi che oggi è ritenuta più valida è che la memoria consista nella formazione o alterazione dei contatti sinaptici (D.O.Hebb,1949). Per farsi un'idea della distanza intercorrente tra la visuale di allora e quella attuale, basti considerare che la complessità funzionale di un singolo neurone è recentemente paragonata a quella di un microprocessore contenente migliaia di transistors.

L'aspetto più suggestivo di quell'ottimistica visione era fornito dalla risposta neuronale di tipo *tutto o niente*. Questo semplice principio di funzionamento trovò ben presto legittimazione teorica nella rappresentazione in cifre binarie (*bit*), ossia mediante stringhe di 0 e 1, dell'informazione prodotta da una sequenza di *decisioni dicotomiche* (C.Shannon, 1949). Tutto ciò corrispondeva pienamente a quella che poteva sembrare la profetica eredità culturale di George Boole (1853), fondatore del calcolo logico: l'idea che le regole di composizione del linguaggio logico (negazione, congiunzione, disgiunzione e loro combinazioni) concepite come operazioni algebriche su variabili proposizionali a valori digitali (*tutto* = 1 = *vero* e *nulla* = 0 = *falso*), non siano altro che le stesse operazioni fondamentali del pensiero. In effetti, assumendo che i costituenti elementari delle reti nervose siano

decisori a soglia, diviene naturale ipotizzare che nel cervello avvengano processi simili al calcolo booleano. Si può capire con quanto divertimento McCulloch e Pitts, e in modo più completo e rigoroso S.C.Kleene (1956), siano riusciti a dimostrare che tutte le operazioni del calcolo di Boole possono essere effettuate da reti costituite soltanto di decisori a soglia ritardati (v. A. de Luca & L.M.Ricciardi, 1981). Viceversa, poiché il funzionamento di un decisore a soglia è approssimabile quanto si vuole da un dispositivo costituito di unità elementari capaci di implementare le operazioni della logica booleana, ogni elaborazione d'informazione effettuabile da una rete formata da elementi del primo tipo risulta approssimabile quanto si vuole da una rete sufficientemente complessa del secondo tipo. Insomma, le reti neurali di McCulloch e Pitts e i calcolatori digitali sarebbero due generi funzionalmente equivalenti. L'impetuoso sviluppo delle tecnologie informatiche degli anni '60 contribuì ad alimentare ulteriormente questa illusione di semplicità.

Mediante applicazioni ricorsive di processi booleani su stringhe digitali fu possibile implementare il calcolo aritmetico e, attraverso questo, calcoli matematici di qualsiasi genere e complessità. La questione del rapporto tra logica e aritmetica si collegò in modo naturale alla stupenda problematica della ricorsività: l'analisi di Gödel, le macchine di Turing, la teoria dei linguaggi formali, ecc. Sembrò così potersi effettuare anche la quadratura del cervello: nulla accadere nella mente che non sia simulabile da un calcolatore sufficientemente complesso. Si poteva pretendere di più? Il circolo sembrava chiudersi così nel modo più soddisfacente. Quali dubbi avrebbero potuto sorgere circa la validità di questo approccio se matematici come J.von Neumann, C.Shannon, E.P.Moore, M.Arbib, ecc. contribuirono a legittimarlo elevandolo al rango di una teoria sistematica e completa? Tale concezione ha informato fino a oggi la ricerca sull'intelligenza artificiale (IA), a proposito della quale merita ricordare i nomi di M.Minsky, A.Newell, H.Simon. Purtroppo in questo campo i risultati sono stati tanto deludenti quanto le prospettive entusiasmanti. Banalmente, il software IA non ha ancora incontrato il favore di mercato.

2. Gödel, Turing e von Neumann.

Alcuni teorici dell'IA sembrano nutrire da vari anni, con prudenza e riserbo, un ambizioso progetto: creare un calcolatore — e un corrispondente processo di elaborazione dell'informazione — che possa definirsi capace di autocomprensione, che riproduca, cioè, la capacità del pensiero umano di pensare se stesso; in altri termini l'*autocoscienza*. L'idea è stata divulgata in forma spettacolare da uno di essi, Douglas Hofstadter, nel libro *Gödel, Escher, Bach* (1979). L'autore trae ispirazione dalla teoria dell'*autoreferenzialità* dell'aritmetica scoperta da Gödel (1931) e a quella, del tutto affine, dell'*autoriproducibilità* degli automi, delineata da John von Neumann in una serie di conferenze e appunti dal 1949 al 1956. Nel pensiero di questi due sommi teorici, le possibilità, rispettivamente, dell'autoreferenza e quella dell'autoriproduzione, dipendono dalla condizione che i sistemi considerati — rispettivamente l'algoritmo aritmetico e l'automa costruttore — siano sufficientemente complessi. La questione è così importante da meritare una breve riflessione.

Un *algoritmo* è un insieme di regole procedurali, applicabili ricorsivamente ai dati di un certo insieme, mediante le quali è generalmente possibile, se il processo non continua all'infinito, produrre risultati collocabili nell'insieme stesso. Nella pratica matematica gli algoritmi sono procedure operazionali che permettono di produrre idealmente gli

enti matematici. Anche le regole di costruzione geometrica mediante riga e compasso costituiscono un algoritmo. Diciamo che un algoritmo ne *interpreta* un altro se, operando su un proprio sottoinsieme di dati, corrispondenti all'insieme di dati dell'altro, è capace di produrre risultati esattamente corrispondenti a quelli dell'altro. Brevemente, se è in grado di simulare i comportamenti dell'altro. Un algoritmo che non sia abbastanza complesso (che esegua, per esempio, solo operazioni di addizione) riesce, al più, a interpretare algoritmi simili (per esempio il calcolo delle moltiplicazioni) o più semplici. Ma se la sua complessità supera un certo livello critico, esso può divenire *universale*, cioè capace di interpretare qualsiasi altro algoritmo. Si può comprendere quanta importanza abbia avuto la scoperta che *l'aritmetica è un algoritmo universale*. È questa la ragione per cui tutte le procedure matematiche (per esempio anche quelle puramente geometriche), purché opportunamente interpretate, sono riconducibili a calcoli aritmetici.

D'altronde, ogni algoritmo possiede una struttura che può essere presentata in forma assiomatica ed elaborata con procedimenti logici; precisamente, come un sistema puramente formale costituito di *nozioni primitive*, *assiomi* e *regole d'inferenza*. Componendo le nozioni primitive secondo i principi della logica si ottengono espressioni più o meno complesse. Tutte le espressioni ottenibili per combinazione logica di altre, indipendentemente dalla verità di tali altre, si dicono *formule ben formate*. Formule complesse d'uso frequente possono essere opportunamente rimpiazzate, per definizione, da formule (o nozioni) più semplici. Gli assiomi sono formule ben formate assunte *vere* per definizione. Le formule che si possono ottenere componendo, elaborando e riducendo logicamente le formule costruibili con gli assiomi e le regole d'inferenza del sistema assiomatico considerato si chiamano *teoremi*.

In pratica l'elaborazione logica si presenta come la formalizzazione del ragionamento matematico. Essa consiste in una successione di passaggi alcuni dei quali hanno l'effetto di aumentare la lunghezza delle espressioni, altri di diminuirla, fino a produrre, come risultato finale, un enunciato interessante. La diminuzione di lunghezza può avvenire sia applicando le regole d'inferenza del sistema assiomatico, le quali in pratica stabiliscono che certe formule possono essere rimpiazzate da altre meno complesse, sia applicando quelle della logica (ad esempio il *modus ponens* che permette di rimpiazzare la formula 'se A allora B ' semplicemente con B). L'assiomatizzazione dell'algoritmo aritmetico, effettuata da R.Dedekind (1888) e G.Peano (1901), si basa sulle nozioni primitive di *numero* e *successore* e sulla enunciazione, nei termini di queste, degli assiomi e delle regole inferenziali che istituiscono la possibilità di eseguire quante si vogliono addizioni e moltiplicazioni e inoltre di enunciare una verità generale da una concatenazione ricorsiva infinita di verità particolari (*assioma d'induzione*). La formalizzazione logica dell'algoritmo aritmetico, esteso ai numeri reali, non è altro che l'algebra elementare.

Dunque, un algoritmo può essere considerato da due punti di vista diversi: uno concreto, operativo, comportamentale, diacronico; l'altro astratto, rappresentazionale, strutturale, sincronico. Dal primo esso si presenta come un sistema di procedure atte a trasformare concretamente un insieme finito di dati particolari in un altro insieme dello stesso genere; dal secondo come una struttura formale che definisce, in astratto, un campo infinito di possibilità esistenziali. Questa relazione tra *algoritmica* e *logica*, che caratterizza in modo profondamente dualistico l'intera matematica, ha creato storicamente una

tensione che si è manifestata prima nel tentativo monistico, perseguito da Gottlob Frege (1879, 1903), di ridurre l'aritmetica alla logica, e poi nell'opera di Kurt Gödel (1930,1931) volta alla ricerca della possibilità di rappresentare e interpretare aritmeticamente le stesse dimostrazioni logiche. Questi approcci, che possono ritenersi riusciti solo parzialmente e sotto certe condizioni, hanno avuto l'effetto di ampliare oltre i limiti delle loro formulazioni originarie, prima la logica (attraverso i lavori di A.N.Whithead & B.Russel, E.Zermelo, A.Fraenkel, D.Hilbert, A.Tarsky ecc.) e poi l'algoritmica (con i contributi di A.Church, A.Turing, S.C.Kleene, E.P.Moore, ecc.).

La possibilità di tradurre le procedure logiche in termini aritmetici, scoperta da Gödel, si basa sul fatto che le elaborazioni logiche concrete, ad esempio le dimostrazioni dei teoremi di un sistema assiomatico, possono interpretarsi come produzioni algoritmiche di dati a partire da altri dati; precisamente come operazioni di calcolo booleano applicate a collezioni di valori di verità/falsità (in pratica stringhe di 0 e 1). Così, ad esempio, il teorema $3^2+4^2 = 5^2$, — da interpretarsi logicamente come l'enunciato di verità aritmetica: “ è vero che $3^2+4^2 = 5^2$ ” — sarà rappresentabile, in primo luogo, come una collezione di valori logici che stabilisce, nell'ambito delle proposizioni logicamente ammissibili, la verità dell'enunciato in questione e, in secondo luogo, risulterà producibile in questa medesima forma applicando una opportuna sequenza di operazioni alle stringhe binarie che rappresentano gli assiomi e le regole d'inferenza del sistema aritmetico.

Ora, l'aritmetica, in quanto algoritmo universale, può rappresentare numericamente le stringhe dei valori logici e interpretare in forma aritmetica le operazioni logiche applicate alle dimostrazioni dei teoremi di un qualsiasi sistema assiomatico, pertanto anche le dimostrazioni dei suoi propri teoremi! Perciò, ad esempio, in primo luogo si potrà assegnare un numero al teorema $3^2 + 4^2 = 5^2$ e altri numeri agli assiomi e alle regole inferenziali dai quali il teorema stesso è deducibile, e in secondo luogo esisteranno certi calcoli aritmetici che, applicati ai numeri degli assiomi e delle regole inferenziali, produrranno come risultato il numero del teorema. È questa l'idea che sta alla base dell'analisi gödeliana. Con tutta generalità, l'*interpretazione autoreferenziale del sistema assiomatico di un algoritmo* (gödelizzazione) si ottiene associando, in primo luogo, certi dati agli assiomi e alle regole d'inferenza del sistema e rappresentando, in secondo luogo, le elaborazioni logiche del sistema assiomatico mediante operazioni dell'algoritmo medesimo applicate a quei dati (generalmente in modo ricorsivo). L'esistenza di teoremi indimostrabili corrisponde all'occorrenza di procedure ricorsive illimitate. Si noti che, affinché l'autointerpretazione funzioni, è essenziale che la struttura logico-formale dell'algoritmo sia codificata in dati utilizzabili dall'algoritmo stesso. Si noti ancora che l'autointerpretabilità è possibile perché i dati, ai quali è lecito applicare con tutta generalità le operazioni dell'algoritmo (nel caso dell'aritmetica, numeri interi e collezioni di numeri interi), costituiscono un *insieme infinito di possibilità*. Solo un insieme infinito, infatti, ammette corrispondenze biunivoche di tutto l'insieme entro sue parti proprie.

Originariamente l'analisi di Gödel riguardava il rapporto tra struttura logica e potenza algoritmica della matematica. Ciò rifletteva in modo naturale l'interesse maturato nel dibattito scientifico del primo '900. I risultati di Gödel furono reinterpretati da Alan M.Turing (1936) come problema del rapporto tra comportamento e struttura di macchine calcolatrici. Una delle calcolatrici più semplici a immaginarsi è la *macchina di Turing*. Essa

consiste di una testina che può spostarsi di un passo avanti o uno indietro su un nastro infinitamente lungo diviso in caselle uguali, oppure leggere o scrivere in queste caselle una cifra, o un carattere di un alfabeto finito; il tutto secondo regole definibili in una tabella di programmazione. *La macchina di Turing è universale*: opportunamente programmata, essa può interpretare qualunque macchina calcolatrice; essa può fare, cioè, tutto quello che può essere fatto dai calcolatori più complessi. In tal modo l'intera teoria dei processi di calcolo si riconduce allo studio delle possibilità di calcolo di questo elementare congegno.

La teoria degli automi di von Neumann ricalca lo stesso schema concettuale e, per certi aspetti, rappresenta una naturale estensione del punto di vista di Turing. Un automa è una macchina capace di usare oggetti reperibili nell'ambiente per produrre altri oggetti da riporre nel medesimo ambiente. A tale scopo esso deve disporre di un elenco di istruzioni e possedere un repertorio sufficientemente ricco di sensori e strumenti di lavoro per eseguire tutte le necessarie operazioni di cernita e assemblaggio sulla base di tali istruzioni. Perciò, in generale, deve anche essere capace di elaborare informazione. Un automa di complessità strutturale insufficiente non riuscirà a produrre altro che oggetti di complessità inferiore alla sua propria. Ma se la complessità supera un certo livello critico, l'automa diviene *universale*, cioè capace, se opportunamente programmato, di produrre qualsiasi oggetto; anche automi di complessità uguale o superiore alla propria, dunque anche una copia di se stesso. Evidentemente, *la cellula vivente è un automa universale!* Von Neumann — richiamandosi alla teoria di Gödel — rileva che l'automa non potrà 'autocopiarsi' semplicemente esaminando la propria struttura interna mediante sensori o altri ipotetici apparati d'introspezione, poiché nessuna automisurazione o autoosservazione può essere completa. Affinché l'autoriproduzione abbia luogo è essenziale che esso disponga di una descrizione della sua struttura interna e di un apparato di decodificazione che gli permetta di dedurre le procedure costruttive o, equivalentemente, dell'elenco completo delle istruzioni riguardanti le modalità di costruzione e assemblaggio delle sue parti. Tale è appunto il codice genetico degli organismi viventi. Si ha in ciò l'analogo della codificazione di un algoritmo universale in sistema assiomatico, senza la quale l'interpretazione autoreferenziale non potrebbe effettuarsi. È notevole il fatto che von Neumann giunse ad affermare tutto questo circa tre anni prima della scoperta del DNA (J.D.Watson & F.Crick, 1953), suggerendo, tra l'altro, sulla base di argomenti di funzionalità ottimale, che l'informazione genetica sia codificata in sequenze lineari di dati (come sui nastri delle macchine di Turing). Il sommo matematico ha fornito l'esempio di un semplice automa planare capace di autoriprodursi, costituito di cellette quadrate uguali, dotate di un certo numero di stati interni, ognuna delle quali può agire trasformando gli stati di quelle contigue, dimostrando che l'autoriproduzione di un assembramento di cellette, caratterizzato da una certa distribuzione di stati, è possibile solo con cellette suscettibili di assumere almeno 29 stati.

Il punto cruciale di questa profonda visione, che deve ritenersi l'istanza fondante della *biologia teorica*, è il ruolo svolto dalla nozione di *complessità critica* nell'analisi del rapporto tra struttura e comportamento di un sistema formale o naturale. Al di sotto di un certo grado di complessità strutturale la varietà dei comportamenti di un algoritmo o di un automa è logicamente deducibile in modo completo dalla conoscenza della struttura. In queste circostanze, la potenza interpretativa dell'algoritmo, o la capacità produttiva dell'automa, resta limitata. Corrispondentemente, l'autoreferenziale, o l'autoriproduzione,

risulta impossibile. Ma oltre la complessità critica accade un fatto sconvolgente: la varietà dei comportamenti possibili esplode a infinità cantoriane e si danno infiniti comportamenti che *non sono logicamente deducibili dalla conoscenza della struttura*, per quanto questa possa essere perfetta. L'unico modo di conoscerli è quello di *osservarli* nel loro effettivo svolgimento. In queste stesse circostanze, la potenza interpretativa dell'algoritmo, o quella produttiva dell'automa, diviene universale. Da un punto di vista filosofico ed epistemologico è importante rilevare come questo fenomeno spiazzi in modo definitivo la polemica tra riduzionismo e olismo, imponendo con la forza della ragione matematica una concezione dualistico-complementarista delle scienze naturali.

L'idea di Hofstadter ricalca ancora, ma in modo piuttosto oscuro, lo stesso schema. Dipanandola da metafore, doppi sensi, giochi di parole e dalla festosa allegria da *cartoons* che anima lo spirito del libro, essa può essere brevemente descritta nel seguente modo: un cervello pensante è la sede di processi di produzione e autoriproduzione di certi stati di eccitazione della rete nervosa che l'autore definisce *simboli attivi*. Questi sono così definiti perchè, a differenza dei simboli passivi che intervengono nella matematica, non ricevono i loro significati da relazioni di corrispondenza poste da un soggetto esterno, ma sono essi stessi capaci di 'attivare' relazioni con oggetti esterni. Per mantenere l'analogia coi casi precedenti bisogna assumere che nella macchina mentale abbiano luogo processi di trasformazione di simboli attivi in altri simboli attivi, secondo procedure memorizzate in qualche luogo e modo. Naturalmente, non deve esistere alcun apparato supervisore, nessun *homunculus* nascosto, che legga e interpreti un elenco di istruzioni al fine di far eseguire al cervello tali operazioni: l'attività di produzione e riproduzione simbolica di cui consiste ciò che chiamiamo *mente* deve generarsi e sostenersi autonomamente per interazione causale dei simboli attivi che si accendono e spengono nella struttura cerebrale. In analogia coi casi precedenti, aggiungeremo il seguente concetto: la mente *A* è capace di *comprendere* la mente *B* se è possibile stabilire una corrispondenza tra ogni produzione simbolica di *B* e una di *A*. Affinché ciò possa aver luogo, *A* avrà bisogno di mezzi atti ad acquisire e decodificare i simboli e le produzioni simboliche di *B*; dovrà pertanto possedere una capacità di rappresentazione interna coordinata ad apparati d'interazione col mondo esterno. Una mente che abbia una complessità strutturale sottocritica non potrà comprendere altro che produzioni simboliche di complessità uguale o inferiore a quella delle sue; ma se è sovracritica essa diverrà *universale* e quindi, purché sia opportunamente istruita, potrà comprendere ogni altro genere di produzione simbolica: quella di una macchina calcolatrice, di un animale, di altre menti dotate di qualsiasi grado di complessità. È evidente che, in questo senso, *la mente umana è universale!* Hofstadter non spiega come in un cervello animale possa prodursi una mente universale, poichè, nel libro citato, o in un altri del medesimo autore, la questione dell'universalità risulta implicitamente riposta nell'idea della corrispondenza coi teoremi di Gödel. A parte tutte le considerazioni di carattere filogenetico, si può ipotizzare che tale capacità si organizzi fin dalle primissime fasi dell'età evolutiva attraverso la comunicazione emozionale, l'apprendimento emulativo e l'assimilazione dei codici di comportamento.

Nemmeno riguardo la genesi dell'autocoscienza, lo studioso americano fornisce uno sviluppo concettuale adeguato all'idea. Possiamo tentare di completare in questa sede l'argomentazione, spingendo fino in fondo l'analogia con l'autoreferenzialità Gödeliana.

La questione sembra riguardare il rapporto tra due fondamentali facoltà della mente umana: l'*immaginazione* (identificabile con l'attività di produzione simbolica descritta sopra) e il *linguaggio naturale*. Sembra infatti ragionevole farle rispettivamente corrispondere ai termini del dualismo matematico: l'attività algoritmica e il linguaggio logico. La gödelizzazione del rapporto tra immaginazione e linguaggio naturale potrebbe allora procedere nel seguente modo.

Il linguaggio naturale, che dal punto di vista strutturale si presenta come sistema di relazioni tra possibilità semantiche astratte e universali, dal punto di vista comportamentale si presenta come generazione grammaticale di significanti concreti e particolari. In quest'ultima forma, un'elaborazione linguistica sintatticamente corretta, indipendentemente dal suo significato originario, diviene simile a una produzione simbolica. Come tale, essa è suscettibile di essere interpretata e compresa da una mente sufficientemente complessa tramite la corrispondenza con una produzione di simboli attivi concreti e particolari; in breve, esemplificata da una sequenza di immagini prodotte dalla fantasia. Una mente universale potrà interpretare e comprendere, ossia esemplificare immaginativamente, anche le espressioni linguistiche attinenti alle sue stesse capacità e modalità di produzione simbolica. L'autocoscienza, nella sua essenza, dovrebbe consistere proprio di questa attività di comprensione linguistico-simbolica autoreferenziale.

Perseverando nell'analogia, dovremmo ipotizzare che una mente universale non possa giungere all'autocoscienza semplicemente 'percependo' o 'guardando entro se stessa', poiché tale pretesa 'autopercezione diretta' dovrebbe ritenersi tanto impossibile quanto l'autointerpretazione dell'aritmetica senza la codificazione numerica del suo sistema assiomatico inferenziale, o l'autoriproduzione cellulare senza la codificazione genetica. Dovremmo perciò ipotizzare l'esistenza di uno speciale sistema simbolico atto a esemplificare immaginativamente le libere espressioni grammaticali, sintatticamente corrette, riferibili agli stessi processi di produzione simbolica mentale: un codice o un piano di costruzione dell'organizzazione mentale che, generalizzando la nozione di 'simbolo del sè' introdotta da Hofstadter, chiameremo il *sistema simbolico del sè*. L'autore non spiega dove risieda e come si generi questo codice mentale. Ma se il parallelo coi casi precedenti deve valere fino in fondo, dobbiamo assumere che nel processo di formazione dell'autocoscienza, assieme alle capacità di produzione simbolica e di comunicazione con altri esseri pensanti, si stabiliscano anche relazioni di corrispondenza proiettiva e correlazione simbolica tra i modi di comportamento e comunicazione propri e quelli altrui. Secondo questa teoria, l'autocoscienza si innesca a cominciare dall'istante in cui una mente, capace di generazione linguistica e produzione simbolica, raggiunta la complessità critica, diviene un interprete prima universale e poi autoreferenziale e quindi si arricchisce, attraverso l'esperienza, di nuove capacità espressive ed imaginative.

Il grande sogno dell'IA, la creazione dell'autocoscienza artificiale, sembra dunque condizionato dalla possibilità di costruire macchine elaboratrici di informazione capaci di comunicare linguisticamente con gli esseri umani, interpretarne il pensiero, introiettarne lo schema strutturale; fino al punto di costruire un proprio sistema simbolico del sè e divenire quindi autoreferenziali. I calcolatori odierni sono ancora estremamente troppo semplici e lenti per poter giungere a tanto.

3. Processi seriali e paralleli.

Serialità e finitezza di funzionamento, ossia ordinamento temporale e limitatezza numerica delle operazioni eseguibili nell'unità di tempo, sono proprietà caratteristiche dei calcolatori ordinari e della macchina di Turing che li rappresenta paradigmaticamente tutti quanti. Ciò significa che: 1) il tempo impiegato da queste macchine per effettuare un certo numero di calcoli indipendenti non può essere minore della somma dei tempi necessari per effettuarli singolarmente; 2) in un tempo finito sono possibili solo elaborazioni applicate a un numero finito di dati iniziali. Poiché i tempi impiegati dai processi reali hanno un limite naturale nella velocità della luce, nessun progresso tecnologico potrà mai ridurre a quantità arbitrariamente piccole i tempi di calcolo di queste macchine. In effetti, quando si cerca di finalizzare un processo di calcolo al governo o alla regolazione di un sistema di una certa complessità, per esempio a coordinazioni con menti umane in tempo reale, i tempi di calcolo delle macchine seriali risultano irrimediabilmente lunghi. Si presenta pertanto in modo naturale il problema di determinare quale struttura debba avere una macchina calcolatrice per elaborare informazione nel modo più veloce possibile. E' evidente che per avviare a soluzione questo problema dobbiamo rinunciare al paradigma della serialità a favore di quello della sincronicità, ossia della spazialità. Dobbiamo cioè spostare l'attenzione dagli aspetti diacronici e comportamentali a quelli sincronici e strutturali; dunque anteporre gli aspetti logici a quelli algoritmici.

Come è stato rilevato in precedenza, ogni algoritmo ammette una descrizione logico-formale che ne definisce e contempla tutte la possibilità. Perciò la potenza di calcolo di un algoritmo, in particolare la capacità operativa di una macchina di Turing, può ritenersi interamente contenuta nei teoremi logicamente implicati dal suo sistema assiomatico. Ora si può dimostrare (G.Birkhoff, 1933; M.H.Stone, 1934) che, facendo corrispondere all'implicazione logica 'se A allora B ' l'inclusione insiemistica $A \in B$, un reticolo di proposizioni logiche (cioè un insieme di proposizioni organizzato come sistema di tutte le possibili relazioni di implicazione logica) diviene isomorfo a un reticolo di sottoinsiemi di un insieme, di cui si può idealmente fornire una rappresentazione sincronica. La possibilità di questa corrispondenza fa comprendere come il campo semantico della logica booleana possa essere ampliato con l'adozione di regole tipicamente insiemistiche: la distinzione tra l'appartenenza come membro di una classe e l'essere sottoclasse di una classe; l'esistenza di relazioni; l'enunciazione di una proprietà per tutti gli elementi di una classe ecc. L'ampliamento diviene enorme quando, facendo riferimento alle infinite insiemistiche di Cantor, si assume inoltre la validità dell'*assioma della scelta* (E.Zermelo, 1904). Cioè la possibilità di scegliere *simultaneamente* un elemento campione da ogni sottoinsieme di un insieme (v. A. Mazurkiewicz, 1977). In sostanza, questo assioma asserisce l'ammissibilità logica di ciò che nessuna procedura algoritmica sarebbe in grado di effettuare, e sembrerebbe dunque indicare una sorta di superiorità strategica della logica così ampliata (*logica del primo ordine*) rispetto all'algoritmica. Sembrerebbe così che l'universo delle implicazioni logiche di un sistema assiomatico, e in esso l'iperuranio di tutte le sue verità, potesse eludere i limiti della serialità. Sembrerebbe, in pratica, che potesse esistere, almeno idealmente, un supercalcolatore capace di operare simultaneamente su insiemi infiniti di dati e dedurre in tempi arbitrariamente brevi la verità o falsità di ogni possibile proposizione. Tutto questo senza ricorrere a procedure seriali, ma solo attraverso l'interazione

sincronica delle parti. Esso sarebbe infinitamente più potente di qualsiasi calcolatore seriale. Si potrebbe persino ipotizzare un calcolatore tanto potente da verificare o falsificare simultaneamente tutte le proposizioni di un sistema assiomatico. Questa macchina ideale, come il demone di Laplace, sarebbe in grado di indovinare o predire istantaneamente qualsiasi aspetto passato o futuro di un sistema deterministico e sarebbe pertanto in grado di esercitare la miglior attività di controllo possibile su ogni processo naturale di cui arrivasse a possedere un'adeguata interpretazione logico-formale.

Si potrebbe pensare di implemetare simili procedure sincroniche in processi fisici caratterizzati dall'interazione simultanea di un'infinità di parti, ad esempio in processi ottici. Del resto è noto che una semplice lente, che raccolga un fascio parassiale di luce coerente filtrato attraverso un fotogramma, produce sul piano focale la trasformata di Fourier del fotogramma. Questo semplice dispositivo fisico è dunque in grado di produrre sincronicamente, e in un brevissimo istante, quello che può ottenersi anche serialmente, ma solo eseguendo una lunghissima sequenza di moltiplicazioni e addizioni. Quale fantastica potenza di calcolo si potrebbe ottenere se, con un dispositivo simile, fosse possibile filtrare il 'fotogramma' di una proposizione di un qualsiasi sistema assiomatico in modo da avere proiettata su un piano focale la risposta se essa è o no un teorema! Purtroppo la possibilità che esista una macchina di estensione finita capace di operare su insiemi infiniti di dati è immaginabile solo nel quadro della fisica classica. La discretizzazione quantistica delle strutture e dei processi materiali ne vieta l'esistenza fisica reale. Se tale macchina esistesse avrebbe un'estensione infinita.

Anche da un punto di vista puramente teorico, questo ideale logicista, che possiamo definire 'pansinottico', incontra difficoltà di principio che furono già intuite da H.Poincaré prima che la critica logico-intuizionista di L.E.J. Brouwer e l'analisi di Gödel ne facessero emergere le profonde ragioni. La pretesa di fondare l'intero edificio della matematica sulle regole della logica booleana, perseguita ad esempio da G.Frege, è inconsistente, non tanto a causa dell'antinomia di B.Russell, che può essere aggirata con la teoria dei tipi logici e superata con l'istituzione della logica del primo ordine, ma perché nel sistema assiomatico dell'aritmetica si contempla un assioma che non appartiene alla logica del primo ordine in quanto, rispetto a questa, si pone a un livello metalinguistico. Si tratta di quell'assioma fondamentale della matematica, basato sul concetto tipicamente aritmetico di *numero naturale*, che va sotto il nome *principio d'induzione*. Questo afferma che, se P_0, P_1, P_2, \dots è un sistema ordinato di *proprietà*, e se inoltre P_n implica P_{n+1} e P_0 è vera, allora tutte le P_n sono vere. Nella pratica matematica, la dimostrazione induttiva di una verità generale è possibile solo se la dimostrazione del passaggio $P_n \Rightarrow P_{n+1}$ ha una complessità che non cresce all'infinito all'aumentare di n . Se questa condizione non è soddisfatta, la dimostrazione equivale a un processo algoritmico divergente. In questa circostanza, la pretesa verità del caso generale potrebbe non sussistere o addirittura essere indimostrabile. In effetti, potrebbe darsi che una procedura ricorsiva esibisse comportamenti che al passo $n + 1$ dipendono in modo essenziale dall'ordinamento temporale dei risultati ottenuti fino al passo n e manifestasse aspetti che nessun sistema sincronico di implicazioni logiche riuscirebbe a predeterminare. Di questo genere potrebbe essere, ad esempio, la produzione algoritmica dell'insieme di Mandelbrot (1982) che, a dispetto dell'estrema semplicità dell'algoritmo generatore, evidenzia una ricchezza di forme che non

cessa di manifestare nuovi indescrivibili aspetti quando il singolo dettaglio dell'immagine viene ingrandito a volontà. Così, globalmente, mentre da un lato l'algoritmica appare impotente a 'raccontare' l'infinita potenzialità sincronica delle implicazioni logiche, la logica, dall'altro, appare incapace di 'prevedere' le infinite possibilità della ricorsione algoritmica. Dal punto di vista booleano-zermeliano e gödeliano, nell'iperuranio matematico si dà sia l'esistenza di entità incalcolabili che la producibilità di risultati imprevedibili.

Sulla base delle argomentazioni finora presentate, si può comprendere perché anche il più piccolo calcolatore programmabile acquistato in cartoleria sarebbe in grado, purché dotato di memoria esterna di illimitata capacità, di fare le stesse cose del più grosso calcolatore esistente sulla Terra. Ciò che rende impraticabile questa possibilità è la rilevante differenza tra i tempi di calcolo del primo rispetto al secondo: secoli invece di secondi. Il fatto è che la potenza algoritmica di una macchina calcolatrice dipende assai meno dall'organizzazione spaziale che da quella temporale. Ogni insieme finito di dati può sempre essere trasferito da una memoria a un'altra tramite la conversione in una sequenza temporale di segnali; d'altronde ogni trasformazione di dati eseguibile con procedure sincroniche può sempre essere effettuata operando ricorsivamente su sequenze temporali. Tuttavia, la varietà dei comportamenti esibiti dalle sequenze ricorsive, originate da diversi dati iniziali, può risultare tanto complicata da non poter essere determinata in modo più economico mediante un insieme di condizioni simultanee sui dati. Con opportune procedure di scansione, un insieme finito di possibilità sincroniche può facilmente essere ordinato temporalmente. Ma, in generale, risulta assai difficile, se non addirittura impossibile, organizzare spazialmente un insieme molto grande di relazioni ricorsive in modo da rendere sincroniche le operazioni che compongono un processo seriale. La traduzione di una serie temporale di operazioni in un unico sistema di operazioni simultanee è possibile in casi particolari, ad esempio, quando la mancata esecuzione di un'operazione della serie non abbia conseguenze sulla corretta esecuzione delle altre, cioè nel caso che le operazioni siano *serialmente indipendenti*, ma può risultare difficile o praticamente impossibile se la serie è una catena di operazioni non commutative. La trasformata di Fourier è otticamente implementabile con una semplice procedura sincronica perché è una somma di un grandissimo numero di prodotti serialmente indipendenti. Tuttavia, l'indipendenza seriale non è una condizione necessaria per l'ammissibilità di procedure sincroniche. Infatti, sono implementabili in processi fisici simultanei anche processi costituiti di sottoprocessi *commutativamente interdipendenti*.

4. Criteri di parallelizzazione.

Parallelizzare al massimo un processo di calcolo — ciò che in generale potrà farsi in diverse maniere — significa organizzarlo in modo che siano eseguiti simultaneamente quanti più calcoli sono possibili, riducendo al minimo indispensabile i procedimenti che necessitano l'ordinamento temporale. Questo richiede che il processo sia decomposto in una sequenza di fasi in ognuna delle quali siano processati in parallelo quanti più sottoprocessi serialmente indipendenti e/o commutativamente interdipendenti è possibile. Perciò, alla necessaria decomposizione temporale del processo in una successione di fasi, dovrà corrispondere una decomposizione spaziale in una successione di stadi, ciascuno costituito di molte unità di calcolo affiancate. La parallelizzazione appare tanto più conveniente se si considera che, ad

esempio, l'acquisizione di dati dal mondo esterno, le letture e le scritture da e su memorie interne, il filtraggio e la selezione di componenti informative particolari, si presentano immediatamente come flussi d'informazione organizzati in parallelo. Si può comprendere come le funzioni e l'organizzazione ottimale delle unità di calcolo di un medesimo stadio possano dipendere in modo essenziale dalle proprietà generali dei flussi d'informazione entranti, in particolare dal grado di omogeneità delle sezioni del flusso, dalle loro forme di ridondanza e dall'organizzazione della loro complessità, nonché dalle funzioni a cui sono destinati i flussi uscenti, dalla molteplicità delle destinazioni e dagli ordinamenti temporali richiesti.

Il trattamento dei flussi paralleli richiede quasi sempre che le informazioni trasferite dalle varie componenti entrino in relazione tra loro a qualche stadio ulteriore del processo. A tale fine, tutte le unità locali di un medesimo stadio iniziale devono trovarsi nella condizione di far convergere i risultati del loro funzionamento su una o più unità di uno stadio avanzato. Se tra le componenti parallele del flusso iniziale devono stabilirsi relazioni complesse, allora, in generale, ogni coppia di componenti deve convergere a più unità di uno stadio avanzato; cosicché le unità degli stadi più avanzati dovranno ricevere simultaneamente dati da più unità di uno stadio precedente.

Un aspetto caratteristico della tecnica di processamento parallelo, che non ha l'analogo nei processi seriali, è la possibilità di organizzare e controllare a ogni stadio le relazioni di fase tra le componenti dei flussi d'informazione. Se un flusso d'informazione parallela viene processato sincronicamente da un insieme di unità di calcolo indipendenti o commutativamente interdipendenti, le relazioni di fase vengono conservate. Si può intuire come questa proprietà possa giocare un ruolo essenziale nel trattamento delle caratteristiche temporali dei flussi. L'elaborazione dell'informazione espressa sotto forma di ordinamento temporale dei flussi, richiede sia l'effettuazione di ritardi locali sistematici e controllati che l'applicazione di procedure ricorsive. A questo fine, le unità locali devono poter produrre risposte ritardate in modi controllati e le informazioni uscenti dalle unità di uno stadio devono poter retroagire su quelle dello stadio precedente. Inoltre, da ogni stadio dovranno potersi derivare più flussi di informazione, secondo varie esigenze di utilizzazione sincronica terminale.

La macchina calcolatrice parallela si delinea dunque come un sistema di unità locali organizzato secondo configurazioni a cascata o reticolari, percorse da flussi d'informazione in entrambi i sensi. Queste proprietà si ritrovano tutte nell'organizzazione nervosa della corteccia cerebrale! Non dobbiamo affatto meravigliarcene. La lentezza della propagazione dell'informazione lungo le fibre nervose (velocità inferiori a quella del suono) e la limitatezza dello spettro delle frequenze ammissibili per i segnali nervosi (circa 1000 hertz), a fronte del vantaggio evolutivo degli organismi più veloci, hanno imposto che i cervelli animali si conformassero al *principio di massima parallelizzazione possibile a parità di funzioni*.

5. Vecchi e nuovi perceptrons.

Fin dai primi tentativi di modellizzazione delle reti nervose risultò chiaro (H.B.Barlow, 1959) che la diretta equiparazione del cervello all'*automa universale di Turing*, sebbene affascinante e ricca di implicazioni generali, è errata e fuorviante. In quell'epoca, presumibilmente a causa delle scarse conoscenze sul cervello animale, non fu facile nè conveniente

sviluppare una teoria del calcolo parallelo altrettanto ricca e significativa di quella del calcolo seriale. Gli approcci furono deboli e caratterizzati più da spirito ingegneristico che logico-matematico, e continuarono ad esserlo per molti anni; tanto che, ancora oggi, manca una teoria sistematica generale dei processi paralleli. Il ritardo è in buona parte dovuto al fatto che le possibilità di parallelizzare i processi di calcolo nei dispositivi artificiali, a livelli paragonabili a quelli dei sistemi nervosi animali, ha incontrato difficoltà pratiche che la tecnologia sta tentando di superare solo da pochi anni. Il primo esempio di processo parallelo fu proposto da F.Rosenblatt (1959) con il modello del *perceptron*.

Il perceptron è un rete di McCulloch e Pitts, costituita da uno o più strati di neuroni connessi a cascata, in modo che i neuroni di ogni strato, se ve n'è più d'uno, agiscono su quelli dello strato successivo ma non su quelli del medesimo strato o degli eventuali strati precedenti. La rete è atta a trasformare un insieme di segnali applicati in parallelo all'ingresso (stimolo) in un insieme di segnali prodotti in parallelo all'uscita (risposta). Opportunamente 'addestrata', deve dimostrarsi capace di apprendere a far corrispondere a certe classi di stimoli certe risposte. Si assume che le procedure di apprendimento del perceptron avvengono per correzione dei coefficienti di trasmissione sinaptica (pesi) tra i neuroni della cascata. Se questi pesi hanno valori casuali, un qualsiasi stimolo sarà tradotto dal perceptron in una risposta apparentemente casuale. I pesi dovranno essere modificati dal processo di apprendimento in modo che certi stimoli (per esempio parole scritte) siano tradotti in corrispondenti risposte volute (per esempio parole parlate). Ora, per i perceptrons a strato singolo sono possibili relazioni stimolo-risposta piuttosto semplici e di scarso interesse applicativo. Invece, per quelli costituiti da più strati, pur di aumentarne abbastanza il numero, sono possibili, in linea di principio, relazioni stimolo-risposta di qualsivoglia complessità. C'è da rilevare un difetto generale di questo modello: esso non prevede l'esistenza di una rete accessoria per il calcolo e la correzione effettiva dei pesi; cosicché l'intervento correttivo risulta semplicemente rappresentato come possibilità matematica. Inoltre, mentre per i perceptrons a strato singolo le regole per la correzione dei pesi risultarono facili da determinare, per quelli multistratificati non si riuscirono a trovare regole analoghe, che fossero concretamente implementabili, per i neuroni degli strati intermedi (*unità nascoste*). Molti anni dopo si sarebbe capito che la difficoltà era dovuta proprio all'infinita ripidità delle risposte a gradino assunte per i neuroni.

Ricerche ulteriori, svolte per oltre un decennio, principalmente negli Stati Uniti da B.Widrow e i suoi allievi, ma anche in Italia (E.R.Caianiello e collaboratori), hanno contribuito in vario modo all'analisi di tali processi paralleli, e altri di carattere più speciale, e portato ulteriori perfezionamenti a quell'idea. Purtroppo il successo fu alquanto scarso: in parte per le difficoltà dei calcolatori di allora a effettuare simulazioni di comportamento, e in parte perché a quell'epoca nessuno era in grado di formulare ipotesi attendibili circa la struttura e il funzionamento delle reti nervose reali. È giusto comunque ricordare le prime modellizzazioni dei fenomeni di autocorrelazione dell'informazione sensoriale dovuti a W.Reichardt (1956), dell'inibizione laterale (D.Varju, 1962), della temporizzazione dei segnali della rete nervosa cerebellare (V.Breitenberg, 1962). Questo filone di ricerca non ebbe allora molto seguito e fu abbandonato per oltre un decennio dopo che M.L.Minsky & S.Papert (1969) misero in evidenza i suaccennati limiti e difetti indicando implicitamente nell'approccio informatico all'IA una via più meritevole di finanziamenti.

Negli anni recenti la teoria del perceptron, sotto il nuovo nome di teoria delle *reti a retropropagazione*, ha avuto un'improvvisa ripresa. La storia è cominciata quando D.E.Rumelhart, G.E.Hinton e R.J.Williams (1986), riesaminando la vecchia teoria, si accorsero che, rimpiazzando i neuroni con risposta a gradino con neuroni a risposta sigmoideale, si potevano ottenere perceptrons multistratificati che ammettono processi di apprendimento rapidamente convergenti nonché straordinariamente efficienti. Fu allora chiaro che proprio le risposte a gradino avevano bloccato la teoria dei perceptrons per più di vent'anni!

Il procedimento di apprendimento dei nuovi perceptrons multistratificati è basato su una ipotetica propagazione a ritroso, attraverso i vari strati, di segnali originati (in uscita) come differenze comparative tra le risposte effettive e quelle volute, e sulle correzioni dei pesi in modi proporzionali a certi effetti esercitati localmente da questi segnali. Affinché il procedimento funzioni, le correzioni devono essere almeno approssimativamente proporzionali alle derivate delle risposte dei neuroni delle unità nascoste (ciò spiega l'insuccesso dei neuroni con risposte a gradino). Una difficoltà, già segnalata da F.Crick (1989), sta nel fatto che anche per questa nuova versione del perceptron non si danno indicazioni concrete di come dovrebbe materialmente funzionare la retropropagazione per produrre le giuste variazioni dei pesi. Ciononostante è possibile simulare il funzionamento di queste reti facendo funzionare per molte ore grossi calcolatori.

Le simulazioni al calcolatore dei nuovi perceptrons hanno prodotto risultati spettacolari. Reti all'inizio balbettanti imparano a parlare o a riconoscere figure, altre a compiere operazioni complesse (ad esempio governare in retromarcia un camion con rimorchio) (B.Widrow & M.A.Lehr, 1990), altre riescono a catturare la legge sottostante a un insieme di stimoli (ad esempio estraggono risposte invarianti rispetto a gruppi di trasformazioni operanti sugli stimoli), altre a comprimere l'informazione eliminando la ridondanza di messaggi applicati come stimoli all'ingresso ecc.

La spiegazione di tali comportamenti è molto semplice: la procedura di apprendimento per retropropagazione abilita la rete a stabilire relazioni di tipo voluto tra l'insieme degli stimoli possibili e quello delle risposte possibili. Se nella fase di utilizzazione della rete le connessioni tra i neuroni funzionano tutte in avanti, cioè se non vi sono connessioni tra neuroni di uno stesso strato, e inoltre la retropropagazione viene soppressa dopo il processo di apprendimento, la risposta di ogni neurone è funzione univoca e continua delle risposte fornite da quelli degli strati precedenti. Pertanto le anche le relazioni stimolo-risposta risultano mappe univoche e continue e non si manifestano fenomeni critici (multistabilità, isteresi, oscillazioni, comportamenti caotici ecc.). Ora l'insieme degli stimoli possibili si caratterizza matematicamente come *spazio* se è possibile definire una 'distanza' naturale tra coppie di stimoli possibili. Ciò si verifica, ad esempio, se tra stimoli diversi possono stabilirsi relazioni variabili con continuità (criteri di dissimilarità, reciproca trasformabilità per operazioni di tipo geometrico o per deformazioni, classificabilità secondo determinate leggi ecc.). Queste relazioni vengono così proiettate nello spazio delle risposte dove, in generale, risultano organizzate in modi diversi a seconda dei valori assunti dai pesi durante le procedure di addestramento. Così zone convesse dello spazio degli stimoli possono risultare mappate in zone non convesse; zone di uguali dimensioni in zone di dimensioni diverse; più zone non intersecantesi in zone intersecantesi ecc.. Si è potuto dimostrare

che una rete con un solo strato, abbastanza numeroso, di unità nascoste, è sufficiente per approssimare mappe di qualsivoglia complessità (G.Cybenko, 1989; G.Girosi & T.Poggio, 1990). In particolare può accadere che la rete riesca a far corrispondere tutti gli stimoli di una stessa classe a risposte molto simili, così da avere, con buona approssimazione, una corrispondenza del tipo *classe* \rightarrow *punto*. Ora, la proprietà veramente importante mostrata dalle reti è la seguente: affinché la rete apprenda a mappare uno spazio di stimoli in uno di risposte è sufficiente che essa venga sottoposta a un campionario sufficientemente rappresentativo di corrispondenze stimolo–risposta. Sotto queste condizioni, la mappa si deforma e si ripiega in modo da soddisfare a quelle relazioni e non ad altre. Questo procedimento vincola alla medesima relazione tutte le altre possibili corrispondenze. In altri termini la rete ha catturato la ‘legge’ sottostante alle corrispondenze.

Una notevole proprietà delle reti a retropropagazione, destinata ad avere importanti applicazioni pratiche, riguarda la possibilità di ottenere la *compressione dell’informazione*. La questione è in stretta relazione con un teorema di C.Shannon (1949). La codificazione ottimale di un messaggio è quella che ne elimina tutta la ridondanza. Questa è tanto maggiore quanto maggiore è il numero di relazioni tra le parti del messaggio (ripetizioni, simmetrie, derivabilità del testo da un programma di lunghezza ridotta ecc.). Per ridurre la ridondanza, ossia comprimere l’informazione, si possono applicare i seguenti criteri: codificare in forma abbreviata i messaggi, o le parti dei messaggi, che ricorrono più frequentemente, riservando le forme prolisse a quelle più improbabili; codificare separatamente l’elenco delle parti prive di relazioni e le informazioni sulle relazioni tra le parti; codificare un programma generatore del messaggio invece del messaggio. L’importanza della compressione dell’informazione nella teoria della comunicazione deriva dal fatto che i canali di trasmissione dell’informazione non possono avere capacità illimitate; per intensi flussi d’informazione, essi si comportano come colli di bottiglie. Naturalmente, per reintegrare il messaggio originale si richiede un processo di codificazione inversa dell’informazione giunta a destinazione (*decompressione*). Tali procedimenti di compressione e decompressione possono risultare più o meno facilmente calcolabili quando si abbia una buona conoscenza delle forme di ridondanza caratteristiche dei messaggi da trasmettere; ma i calcoli possono risultare proibitivi se le forme di ridondanza sono complesse e imprevedibili. Appare perciò interessante che i segnali uscenti delle unità nascoste di una rete a retropropagazione ben addestrata costituiscano una sorta di codificazione compressa realizzata in modo automatico. Nel caso più semplice questo effetto si ottiene con una rete a tre stadi: uno d’ingresso, uno d’uscita (entrambi con lo stesso numero N di terminali) e uno stadio intermedio con un numero K di unità nascoste sensibilmente inferiore a quello dei terminali. Il sistema viene addestrato a fornire in uscita gli stessi segnali applicati in ingresso. Si è trovato che per comprimere N treni di segnali digitali applicati in parallelo si può ridurre il numero di linee intermedie fino a circa $K = \log_2 N$. Si prevedono già importanti applicazioni di questo procedimento alla telefonia e alla televisione ad alta definizione.

Recenti ricerche di simulazione hanno dimostrato che, a parità di ingressi e uscite, un processo di apprendimento funziona meglio se viene prima decomposto, ove sia possibile, in sottoprocessi indipendenti e quindi implementato in una rete costituita di molte sottoreti a moderata connettività, piuttosto che in un’unica rete a grande connettività (F.Fogelmann, 1990). Il fattore limitante decisivo, riguardo l’impiego delle reti a retro-

propagazione, è il tempo di convergenza dei processi di apprendimento, che cresce esponenzialmente coll'aumentare della connettività, e che si riduce notevolmente col frazionamento di questa. Se inoltre si consideri che le simulazioni delle reti nervose mediante gli ordinari calcolatori seriali richiedono tempi di calcolo dell'ordine delle ore o dei giorni, si può capire quale impulso riceva, in vista delle potenziali applicazioni, la tecnologia del calcolo parallelo (reti nervose artificiali).

6. *La retropropagazione nella corteccia cerebrale.*

Uno degli aspetti più interessanti delle reti a retropropagazione sta in una certa loro somiglianza con l'organizzazione nervosa della corteccia cerebrale, sebbene nelle prime manchi l'analogo delle connessioni tra unità dello stesso stadio che sono tipiche della struttura corticale. Si sa infatti che ogni distinta area corticale, caratterizzata da una specifica omogeneità citoarchitettonica — ve ne sono circa trenta per la sola corteccia visiva umana — proietta le fibre dei propri neuroni efferenti in modo approssimativamente topografico su alcune altre aree, e che queste proiezioni sono invariabilmente reciprocate da connessioni decorrenti in senso inverso. Così il flusso di informazione nervosa che nasce dagli organi sensoriali, si smista a cascata attraverso un reticolo di aree e areole corticali variamente specializzate, talvolta connesse trasversalmente, fino a sfociare nelle aree associative e quindi investire l'attività premotoria. Questo flusso è simultaneamente accompagnato da un controflusso che ripercorre a ritroso tutti gli stadi della cascata. Il modello a retropropagazione suggerisce che una delle funzioni di questo controflusso sia di indurre modificazioni sinaptiche tali da determinare l'apprendimento adattativo di precise corrispondenze funzionali tra stimoli sensoriali e risposte associative e premotorie. Si spiegherebbe così, nei suoi tratti essenziali, la formazione (o memorizzazione a lungo termine) delle facoltà percettive.

Ma il controflusso in questione potrebbe avere anche altre funzioni, non interpretabili dai modelli finora studiati; in particolare, essere operante non solo per processi formativi irreversibili del genere appena considerato, ma anche per processi volatili, che richiedono, cioè, solo eccitazioni transitorie e reversibili della rete neurale, come, ad esempio, quelli proposti da S.Grossberg (1987) per spiegare l'*attenzione selettiva*. Grossberg ha ipotizzato che questo fenomeno, assai studiato dagli psicologi sperimentali, sia generato da segnali in controflusso e provenienti dagli stadi terminali del sistema percettivo e retroagenti su tutti gli stadi intermedi di questo processo. Tale retroazione interverrebbe a modulare le capacità di filtraggio ed elaborazione di ciascuna area corticale, favorendo, stadio per stadio, la promozione delle componenti informazionali più significative e l'eliminazione di quelle irrilevanti (rispetto alle funzioni finali del processo percettivo). In effetti, molti esperimenti psicometrici dimostrano che ogni tipo di attenzione selettiva dipende fortemente dalla particolare preparazione o aspettativa motoria del soggetto (G.Berlucchi, 1989).

Un'altra interessante idea riguardante la retropropagazione è stata recentemente proposta da K.Okajima (1990). Questo autore attribuisce alle mappe topografiche corticali funzioni di trasduzione e trasformazione integrale simili a trasformate di Fourier locali. Questa idea, che ha il principale referente sperimentale nei risultati di D.H.Hubel e T.N.Wiesel (1962–1977) sulle risposte a stimoli ottici dei neuroni delle aree visive, risale ai lavori di E.L.Schwartz (1976) il quale ha dimostrato come attraverso una mappa logaritmico-polare del campo visivo, abbastanza simile a quella che la retina proietta

effettivamente sull'area visiva primaria, e una successiva trasformata di Fourier, ogni immagine visiva possa tradursi in una risposta caratteristica invariante per rotazione e dilatazione e, sotto certe condizioni, anche per traslazione prospettica. Se, come ha ipotizzato B.Cavanagh (1978), la mappa logaritmico-polare fosse preceduta da un'altra trasformazione di Fourier, si potrebbe ottenere un risultato invariante anche per traslazione piana. Questa informazione terminale sarebbe infine utilizzata nel riconoscimento degli stimoli visivi. Naturalmente è difficile pensare che le reti nervose animali possano effettuare le trasformate di Fourier; si può invece immaginare che lo facciano in modi approssimativi o approssimativamente equivalenti, magari attraverso 'trucchi' ingegnosi non ancora scoperti. Un indizio di questo è il fatto che i neuroni 'complessi' della corteccia visiva sembrano rispondere agli stimoli sensoriali in modi che richiamano vagamente proprietà simili a trasformate di Fourier locali. È certo che la percezione visiva umana risulta effettivamente invariante per ampie rototraslazioni, dilatazioni e trasformazioni prospettiche dell'immagine.

Okajima ipotizza che la retropropagazione fornisca, stadio per stadio, la trasformazione inversa rispetto a quella che ha luogo attraverso la propagazione diretta. Ma poiché — riferendoci al modello matematico — le trasformazioni dirette rimuovono parte dell'informazione sensoriale, la completa rigenerazione, mediante trasformazioni inverse, di stati di eccitazione corticale simili a quelli formati durante la percezione originale, sarebbe possibile solo se l'informazione rimossa venisse reintegrata stadio per stadio.

Si può tuttavia notare che questa ipotesi, o una simile migliorata, sembra accordarsi con quei recenti risultati di fisiologia e psicologia della percezione che dimostrano come l'evocazione di ricordi visivi sia accompagnata da un'eccitazione corticale che procede a ritroso dalle aree associative temporali e parietali fino alle aree visive secondarie (M.Farah, 1989). Questo avviene come se l'evocazione mnemonica visiva comportasse la ricostruzione a ritroso di stati di eccitazione corticale simili a quelli occorsi durante la percezione diretta. Si potrebbe dunque ipotizzare che la memoria percettiva (cognitiva, descrittiva) sia distribuita in tutte le aree del reticolo areolare corticale e non, come alcuni autori hanno proposto, in aree associative terminali, specializzate per tale funzione, o in centri specifici del cervello.

7. I modelli olografici.

La questione dei modelli olografici merita un breve cenno. Negli aspetti concernenti la neurofisiologia essa si può far risalire ai risultati della trentennale ricerca condotta da K.S.Lashley sulla localizzazione dei ricordi nei cervelli animali. I suoi esperimenti sembravano indicare in modo inequivocabile che ogni ricordo è disseminato in modo piuttosto uniforme in tutta la corteccia cerebrale! Dopo la scoperta dell'olografia, prima teorica (D.Gabor, 1959) e poi pratica (E.N.Leith & J.Upatnieks, 1962), diversi studiosi notarono una certa analogia tra le proprietà messe in evidenza da Lashley e quella degli ologrammi. In questi, infatti, l'informazione relativa a ogni dettaglio di un'immagine è uniformemente sparpagliata su tutta la lastra fotografica nella forma della traccia lasciata dalla figura di interferenza tra l'onda luminosa emessa da una sorgente-oggetto e quella emessa da una sorgente-chiave. È noto che investendo la lastra fotografica con l'onda emessa dalla sola sorgente-chiave si produce, per diffrazione attraverso la traccia, un'onda identica a quella

originariamente emessa dalla sorgente–oggetto. La produzione di una intensa immagine olografica è dovuta all’interferenza costruttiva tra componenti ondulatorie in coincidenza di fase. Se l’onda evocatrice ha relazioni di fase genericamente diverse da quelle dell’onda–chiave appropriata, al posto dell’immagine olografica si ha un’emissione luminosa di intensità debole e statisticamente uniforme. L’evocazione delle immagini olografiche può prodursi anche in modo associativo, usando come sorgente–chiave la radiazione filtrata attraverso un altro ologramma. Sembrò così che le stesse ben note proprietà associative della memoria trovassero nel modello olografico la più naturale interpretazione.

Il primo modello olografico di memoria associativa, basato sull’ipotesi che nella corteccia cerebrale abbiano luogo propagazioni di tipo ondulatorio, fu proposto da P.Van Heerden nel 1962, e riproposto in sede neurofisiologica da K.H.Pribram (1969) e T.W.Barret (1969). Una versione temporale del modello olografico, basata sull’idea di un’integrazione temporale dei segnali emessi da una miriade di oscillatori e riferita all’apparente attività oscillatoria di origine subcorticale, fu proposta da H.C.Longuet–Higgins nel 1968. A questa fece immediatamente seguito un articolo di D.Gabor (1968) che propose un’interpretazione analogica del modello olografico–temporale e dimostrò come le operazioni tipiche dei processi di registrazione e riproduzione degli ologrammi (le trasformate di Fourier) possono trasportarsi nell’ambito di una teoria dell’integrazione di segnali stocastici o di aspetto stocastico (noise–like). Una migliore formulazione matematica di questo approccio, basata sulle operazioni coniugate di convoluzione e correlazione, venne fornita da A.Borsellino & T.Poggio (1974) ed esemplificata con simulazioni al calcolatore da S.Bottini (1977). Lo stesso autore del presente scritto ha cercato di fornire una base bioelettrica all’ipotesi (non confermata) che l’attività elettroencefalografica sia qualcosa di simile a un processo di propagazione ondulatoria (R.Nobili, 1985, 1987). Tuttavia, sebbene recentemente nella corteccia cerebrale si siano rilevati fenomeni di eccitazione simili a propagazioni ondulatorie, la maggioranza dei neurofisiologi è concorde nel ritenere che l’attività elettrica macroscopica della corteccia dipende dalla sincronizzazione di segnali ritmici di origine subcorticale (R.Elul, 1972). Alcuni autori ritengono che questi fenomeni abbiano un ruolo essenziale nella sintesi corticale di capacità di risposta dipendenti da proprietà globali della stimolazione sensoriale (R.Eckhorn & H.J. Reitboeck, 1989).

8. Alcuni aspetti dell’organizzazione corticale

Rivolgiamo ora l’attenzione ad alcuni aspetti del funzionamento delle reti nervose cerebrali che sono stati trascurati nei precedenti paragrafi e che riguardano il problema del processamento parallelo dell’informazione nervosa e il carattere distribuito e per alcuni aspetti associativo della memoria.

Tornando a considerare le proprietà delle reti a retropropagazione multistratificate illustrate nel paragrafo 6, possiamo notare che ogni neurone è il punto di convergenza di segnali provenienti dai neuroni di uno strato precedente e contemporaneamente di emissione a ventaglio di segnali diretti ai neuroni di uno stadio seguente. Sono invece assenti le interconnessioni tra neuroni di uno stesso stadio. Anche le connessioni decorrenti in senso inverso risultano di fatto ignorate per quanto la teoria dell’apprendimento per retropropagazione le richieda per la formazione della memoria. La ragione di quest’ultima circostanza è piuttosto semplice: nessuno ha ancora trovato un modo di implementare in

processi concreti le procedure di correzione dei pesi, sebbene queste siano matematicamente ben definite. Perciò anche questi modelli, come i vecchi perceptrons, devono ritenersi insufficienti e incompleti. Ora è opportuno osservare che non solo i neuroni con risposte a gradino, tipici dei perceptrons, ma anche quelli con risposta sigmoideale, caratteristici delle reti a retropropagazione, appaiono in ultima analisi come semplici decisori a soglia. La gradualità delle risposte dei secondi è importante ai fini della realizzazione del processo di apprendimento, ma è scarsamente rilevante per quanto riguarda il buon funzionamento della rete una volta che i pesi sinaptici siano stati aggiustati a valori adatti. La decisione dicotomica è l'operazione più elementare che possa effettuarsi localmente in un processo parallelo suddiviso in stadi, mentre invece si deve ritenere che le unità colonnari della corteccia, considerate come sistemi locali di convergenza e divergenza di alcune migliaia di segnali nervosi simultanei, effettuino processi paralleli di tipo più complesso. La condizione che un processo locale o globale abbia un carattere più parallelo che seriale comporta che abbiano un ruolo dominante l'organizzazione spaziale del sistema e i fenomeni d'interazione sincronica tra le parti, piuttosto che quelli di ordinamento temporale e relazione causale tra le varie fasi del processo. Ora, da un punto di vista fisico-funzionale, si può osservare che nell'interazione tra le parti di un sistema, a causa degli inevitabili fattori di ritardo nella propagazione dei segnali, la sincronicità degli effetti è soddisfatta tanto meglio quanto più la struttura è simmetrica rispetto alla permutazione delle parti.

Per studiare meglio il problema e cogliere alcune importanti differenze tra lo stato delle cose reali e le proprietà generalmente assunte nei modelli, è opportuno fare una breve digressione in ambito neurofisiologico e analizzare con maggior dettaglio certe proprietà della corteccia cerebrale. Come si è già rilevato, la corteccia cerebrale, spesso circa due millimetri, è suddivisa in varie aree (un centinaio per il cervello umano) distinguibili per le diverse tessiture neurali. Ognuna di queste può essere analizzata sia in senso verticale che orizzontale. Verticalmente, ogni area appare suddivisa in strati neurali, tipicamente sei, differenziati per taglia e organizzazione funzionale. Orizzontalmente, invece, appaiono omogenee. Tuttavia, a un esame più attento, l'organizzazione orizzontale rivela una struttura fine; precisamente la rete nervosa appare suddivisa in strutture colonnari a loro volta connesse secondo schemi di varia forma e disposizione: bande, raggruppamenti più o meno regolari, ecc. Ogni struttura colonnare contiene migliaia di neuroni di vari tipi e dimensioni. Interessa qui rilevare che le strutture colonnari, del diametro di alcune centinaia di micrometri, hanno proprietà simili a quelle che si dovrebbero richiedere per funzionare come unità locali di processamento parallelo. In particolare neuroni omologhi, cioè disposti allo stesso modo entro queste unità, sembrano connessi in modi approssimativamente simmetrici.

Per alleviare un poco la penosa impressione che la grande complessità delle strutture nervose corticali può suscitare, mettiamo in evidenza alcune proprietà semplificative generali. Nel sistema nervoso centrale non si trovano neuroni che siano simultaneamente eccitatori e inibitori. I neuroni piramidali, gli unici dotati di afferenze ed efferenze extra-corticali, sono eccitatori e costituiscono la struttura di riferimento della corteccia; gli altri possono essere chiamati interneuroni, poichè svolgono funzioni ausiliarie tra e per i piramidali. I piramidali hanno una struttura abbastanza tipica, caratterizzata dalla presenza di una formazione dendritica elevantesi ad albero dal corpo cellulare, fornita di rigogliose ram-

ificazioni alla radice (dendrita basale) e all'estremità superiore (dendrita apicale). Il tronco dendritico riceve principalmente segnali eccitatori di origine esterna: fibre talamiche che portano informazione sensoriale di tipo specifico (fibre specifiche) o fibre di piramidali dello stadio precedente della cascata corticale, sia direttamente o attraverso interneuroni eccitatori (cellule stellate-spinose). Queste ultime raccolgono segnali provenienti dall'esterno convogliandoli sulle spine degli alberi dendritici di alcuni piramidali limitrofi mediante terminazioni assoniche ascendenti, a forma di coda di cavallo. Il dendrita basale riceve principalmente segnali eccitatori da alcune migliaia di piramidali omologhi o giacenti a livelli inferiori. Il dendrita apicale riceve principalmente segnali eccitatori da fibre talamiche aspecifiche e dai piramidali di altre aree che inviano i segnali di retropropagazione.

La maggior parte degli interneuroni agiscono inibitoriamente sui piramidali; in particolare le cosiddette *cellule a canestri* si caratterizzano per esercitare intense azioni inibitorie sui piramidali limitrofi; ognuna di esse avvolge coi suoi terminali assonici i corpi piramidiformi di alcune di queste cellule con innervazioni cestiformi. Esse ricevono segnali eccitatori principalmente dai piramidali della stessa unità colonnare e in misura minore da quelle adiacenti; cosicché al sinergismo dell'interazione eccitatoria piramidale si contrappone, per il loro tramite, un'azione inibitoria antagonista, in parte autogena e in parte di origine laterale. Un altro tipo di interneuroni inibitori sono le *cellule a candelabro*. Con le loro tipiche diramazioni assoniche rivolte all'insù, esse agiscono direttamente sugli assoni di alcuni piramidali limitrofi come interruttori di uscita bloccando l'emissione di segnali verso l'esterno (C.Asanuma & F.Crick, 1987).

Considerando che la sola interazione eccitatoria tra i piramidali, omologhi e non, di una medesima unità colonnare porterebbe rapidamente il sistema al parossismo, si comprende quanto siano essenziali gli interneuroni inibitori. Ma affinché l'attività inibitoria abbia l'effetto di limitare quella eccitatoria, senza tuttavia impedirla del tutto, bisogna che la dipendenza dell'azione inibitoria sui piramidali cresca con legge di potenza maggiore di quella eccitatoria. Poiché le cellule a canestri sono eccitate dagli stessi piramidali, a tal fine è sufficiente che l'azione inibitoria si trasmetta con legge di potenza maggiore di uno. A questo proposito è interessante considerare il fatto che, a causa di un meccanismo d'attivazione a doppio sito, l'azione esercitata sui recettori presenti nei corpi dei piramidali dei segnali inibitori prodotti dalle cellule a canestri dipenda quadraticamente dalla concentrazione del neurotrasmettitore inibitorio, l'acido γ -amminobutirrico (GABA) (J.Borman & D.E.Clapham, 1985).

Circa le possibili modalità di funzionamento delle unità colonnari e le interazioni tra unità diverse, non si è ancora raggiunta una comprensione soddisfacente. Le ricerche di Huebel e Wiesel sulle risposte dei neuroni della corteccia visiva primaria a stimoli visivi indicano che la complessità delle interazioni locali aumentano dal quarto strato (strato di arrivo dei segnali talamici) verso gli strati superiori e inferiori, cioè secondo una struttura approssimativa a doppio cono. È lecito presumere che le proprietà più importanti dipendano in modo essenziale dal sincronismo dei processi d'interazione locale; ma anche i fenomeni di ordinamento temporale potrebbero avere grande importanza. Sarebbe perciò poco prudente ritenere che il gioco di sinergismi e antagonismi entro le strutture colonnari produca risultati più rilevanti dal punto di vista dei processi paralleli e meno da quello dei processi seriali. Si sa, ad esempio, che l'azione inibitoria locale segue quella eccitatoria con

un certo ritardo temporale; se le modalità di generazione e propagazione di queste azioni sono asimmetriche — come tutto porta a ritenere — possono derivarne importanti ruoli funzionali. Anche la sollecitazione eccitatoria ritmica, esercitata sui neuroni corticali dalle strutture sottocorticali, deve certamente avere grande importanza funzionale.

9. *Stabilità, multimodalità e isteresi.*

Rivolgiamo ora l'attenzione al problema della modellizzazione delle strutture corticali locali, privilegiando gli aspetti che possono riguardare maggiormente le funzionalità e i comportamenti tipici dei processi paralleli.

I modelli di reti a retropropagazione discussi in precedenza inducono facilmente a considerare le aree corticali come filtri analizzatori dei flussi d'informazione nervosa (sensoriale, motoria, centrale ecc.) e codificatori delle loro componenti significative. Questi sistemi sarebbero capaci di 'mappare' con continuità, nei modi più convenienti, lo 'spazio degli stimoli possibili' in 'spazi di risposte corrispondenti'. Ciò che viene ignorato in questa visuale è la possibilità di avere risposte *stabili, multimodali e isteresiche*. Per stabilità delle risposte intendiamo la loro sostanziale invariabilità rispetto a piccole e arbitrarie variazioni degli stimoli. Per multimodalità la possibilità che allo stesso stimolo possano corrispondere più risposte diverse. Per isteresi la tendenza di ogni risposta a permanere a dispetto anche di ampie variazioni dello stimolo e a cambiare bruscamente quando lo stimolo viene variato oltre certi limiti. La multimodalità è comunemente osservata negli esperimenti sulla percezione (figure ambigue, interpretazioni dipendenti dal contesto, ecc.). Se le risposte sono multimodali, l'isteresi si manifesta nel seguente modo: per transitare da una risposta stabile a un'altra bisogna variare gli stimoli oltre certi limiti critici; e quando si tenta di riottenere la risposta precedente, i limiti critici della transizione inversa risultano dislocati altrove rispetto ai precedenti. Da un punto di vista matematico l'isteresi è una proprietà generale dei sistemi *strutturalmente stabili*, cioè tali che un'arbitraria perturbazione non ne distrugga le proprietà (comportamenti, varietà delle risposte, condizioni di funzionamento ecc.). Questa non è solo una condizione matematica cui un modello deve soddisfare per avere senso, ma anche una condizione fisicamente necessaria per il buon funzionamento di una rete nervosa reale. Essa infatti richiede che la rete possa sopportare perturbazioni e persino danni di una certa entità senza gravi conseguenze funzionali.

Stabilità, multimodalità e isteresi sono proprietà che emergono con evidenza da tutti gli esperimenti di psicologia: dall'analisi delle capacità di discriminazione percettiva di dettagli elementari alla percezione gestaltica delle forme complesse, dall'evocazione mnemonica alla cognizione concettuale. Esse si presentano come regole dominanti e straordinariamente importanti di tutti i processi mentali. Ora è interessante considerare che queste proprietà sono anche caratteristiche dei sistemi non lineari costituiti di parti uguali interagenti in modi almeno approssimativamente simmetrici e sottoposte all'azione di parametri di controllo. Siamo così portati in modo naturale a studiare modelli di reti nervose capaci di funzionare come sistemi multistabili a controllo multiparametrico. A questo proposito conviene osservare che la ragione essenziale per cui la stabilità, la multimodalità e l'isteresi non si manifestano nei perceptrons e nei modelli a retropropagazione, considerati in precedenza, è sostanzialmente dovuta sia al carattere elementare e unimodale delle unità di processamento parallelo locale (decisori a soglia) sia all'assenza di retroazioni positive

tra stadi successivi in condizioni di funzionamento effettivo. Dovremmo dunque attenderci progressi decisivi rimpiazzando i decisori semplici delle unità nascoste con *decisori complessi* capaci di comportamenti multimodali.

Se i fenomeni di multimodalità e isteresi osservati nei sistemi nervosi reali non dipendano solo da interazioni neurali aventi luogo localmente nei singoli strati (multistabilità locale), ma anche da fenomeni di retroazione di ciascun stadio sui precedenti (multistabilità globale) dovremmo aspettarci che il gioco delle interdipendenze funzionali tra livelli di multistabilità locali e globali sia capace di generare processi complessivi di straordinaria ricchezza. Abbiamo già potuto considerare come le reti a retropropagazione siano capaci di memorizzare semplici corrispondenze dirette tra certi stimoli e certe risposte. Abbiamo anche intravvisto la possibilità che modelli più appropriati siano capaci di spiegare funzioni di memoria assai più complesse: l'attenzione selettiva; la ricostruzione retropropagativa dello stimolo originale; la dipendenza dei ricordi dal contesto (che potrebbe interpretarsi come un effetto isteresico); l'organizzazione automatica delle contestualizzazioni dei ricordi secondo criteri di similarità e peculiarità distintive (capacità di categorizzazione); la produzione di pseudoricordi (evocazioni oniriche o fantastiche dotate di significati non banali); ecc. Cerchiamo dunque di descrivere con una certa generalità alcuni ragionevoli modelli di unità colonnari.

10. Un modello generale di decisore complesso.

Da un punto di vista matematico un insieme di n neuroni eccitatori (piramidali), controllati da afferenze esterne, interagenti direttamente tra essi e anche indirettamente per il tramite di interneuroni inibitori, costituisce un sistema multistabile che può essere matematicamente descritto da un sistema di n equazioni non lineari dipendente da parametri di controllo. In tale rappresentazione lo stato di eccitazione (frequenza di scarica) del generico piramidale i a un certo istante t è rappresentato da una variabile reale positiva $y_i(t)$. Essa può ritenersi una funzione istantanea $F[v_i(t)]$ della depolarizzazione postsinaptica $v_i(t)$ complessivamente esercitata da tutte le afferenze, eccitatorie e inibitorie, nel piramidale i . Poiché la frequenza di scarica di un piramidale si annulla per depolarizzazioni negative e inoltre non può superare un certo valore massimo, determinato dal periodo di refrattarietà tra due *spikes* successivi, $F(v)$ avrà una forma sigmoideale. Potremo convenzionalmente assumere che i suoi valori minimo e massimo siano rispettivamente 0 e 1. Un'espressione tipica è $F(v) = 1/[1 + \exp(-4\lambda v)]$, dove $\lambda > 0$ è la pendenza centrale della sigmoide. La forma precisa di questa funzione non è rilevante, poiché il funzionamento dei sistemi multistabili interessa assai più per i loro aspetti qualitativi piuttosto che per quelli quantitativi. La depolarizzazione v_i dipende a sua volta dagli stati di eccitazione di tutti i neuroni afferenti negli istanti immediatamente precedenti e dall'eventuale stimolo esterno x_i direttamente applicato al piramidale i (parametro di controllo). Assumiamo per semplicità che v_i dipenda a sua volta dalle stesse attività dei piramidali (direttamente oppure tramite gli interneuroni) con ritardi temporali multipli di un certo intervallo elementare τ (ritardo sinaptico); potremo pertanto esprimere le depolarizzazioni

$$v_i(t + \tau) = g_i[y(t), y(t - \tau), y(t - 2\tau), \dots] + x_i ;$$

dove $g_i(y)$ sono opportune funzioni dipendenti dagli accoppiamenti sinaptici in gioco. Si otterrà così un sistema di equazioni che descrivono la dipendenza dello stato di eccitazione

di tutti i piramidali delle rete, semplicemente dello *stato della rete*, all'istante $t + \tau$, che indicheremo semplicemente con $y(t + \tau)$, dagli stati di tutti i neuroni (piramidali e non) agli istanti $t, t - \tau, t - 2\tau, \dots$:

$$y_i(t + \tau) = F\{g_i[y(t), y(t - \tau), y(t - 2\tau), \dots] + x_i\} \quad ; \quad i = 1, 2, \dots, n .$$

Le soluzioni di questo sistema descrivono l'evoluzione temporale dello stato della rete per certi valori dei parametri di controllo. In generale, in corrispondenza a fissati valori dei parametri di controllo, tale sistema ammetterà una moltitudine di soluzioni generalmente corrispondenti a stati stabili o instabili, o a comportamenti oscillatori e persino caotici di vario genere. Tuttavia, se sono soddisfatte certe condizioni di simmetria almeno approssimativa, per arbitrari valori dei parametri di controllo il sistema ammetterà almeno un stato stabile. In tal caso, agendo su questi parametri, si potranno determinare brusche transizioni tra stati stabili. In questo caso si può dimostrare con tutta generalità (R.Nobili, 1991) che esiste una specie di funzione $U[y; x]$, dipendente dallo stato della rete e dai parametri di controllo, (funzione di Ljapunov) che ha i minimi relativi in corrispondenza degli stati stabili (sebbene in generale il comportamento dinamico del sistema non sia rappresentato dalle linee di massima pendenza!). Essa rappresenta una sorta di *buca di potenziale*. Agendo sui parametri di controllo la buca può deformarsi in modo tale da determinare il brusco 'versamento' di uno stato da una configurazione stabile a un'altra.

11. Il modello di Hopfield.

Il più semplice modello di rete neurale del tipo ora descritto è quello di J.J.Hopfield, sia nella prima versione dei neuroni con risposte a gradino (1982), sia in quella successiva dei neuroni a risposta sigmoidale (1984). Nel primo caso la funzione sigmoidale $F(v)$, descritta nel paragrafo precedente, viene assunta al limite per $\lambda \rightarrow +\infty$ in modo da ottenere la funzione a gradino di Heaviside: $F(v) = 0$ per $v < 0$; $F(0) = 1/2$; $F(v) = 1$ per $v > 0$. Nel secondo caso λ viene assunta finita; anzi, senza perdita di generalità, uguale a 1. In entrambi i casi si assume la dipendenza lineare delle $g_i(y)$ dallo stato della rete: $g_i(y) = \sum_j W_{ij}y_j$, dove le quantità W_{ij} (pesi sinaptici) possono assumere entrambi i segni e soddisfano la condizione di simmetria $W_{ij} = W_{ji}$ e di non autointerazione $W_{ii} = 0$. Le equazioni della rete diventano allora $y_i(t + \tau) = F[\sum_j W_{ij}y_j(t) + x_i]$. Per $x_i \gg 1$ o $x_j \ll 0$ lo stato di eccitazione del neurone i tende rispettivamente a un valore u_i uguale a 1 o a 0. Imponendo simili valori estremi ai parametri di controllo si può forzare la rete a memorizzare uno stato $u = (u_1, u_2, \dots, u_n)$, $u_i = 0, 1$, cioè uno stato costituito da un sottoinsieme di neuroni massimamente attivi e dall'insieme complementare di neuroni silenti. Si può dimostrare che gli stati stabili della rete sono i minimi della funzione di Ljapunov $U[y; x] = -\frac{1}{2} \sum_{ij} W_{ij}y_iy_j - \sum_i y_i x_i$.

L'idea di Hopfield fu di attribuire alla multistabilità funzioni di memoria. Ogni stato stabile della rete costituisce un attrattore per numerosissimi stati iniziali prossimi a esso. Inizializzando la rete in uno di questi, lo stato evolverà rapidamente verso l'attrattore. Pertanto, se si riesce a fare in modo che certi stati di eccitazione della rete siano *appresi* come stati stabili, ciascuno di questi potrà ritenersi un ricordo evocabile da porzioni approssimative del suo contenuto (*content-addressable memory*). In questo modo la rete potrà funzionare come una memoria autoassociativa. In effetti ciò può ottenersi in modo

molto semplice assumendo una regola di variazione dei pesi sinaptici simile a quella proposta da Hebb nel 1949: le connessioni sinaptiche tra due neuroni devono rinforzarsi se i neuroni fanno la stessa cosa (se sono entrambi accesi o entrambi spenti), indebolirsi se fanno cose opposte (se uno è acceso mentre l'altro è spento). Possiamo esprimere matematicamente questa condizione dicendo che le variazioni dei pesi ΔW_{ij} sono legate agli stati forzati u_i, u_j dei neuroni i e j , corrispondenti a un voluto ricordo u , dalla relazione $\Delta W_{ij} = \alpha(u_i - \frac{1}{2})(u_j - \frac{1}{2})$, dove α è una quantità positiva dipendente dal tempo di stimolazione. Queste regole sembrano in buon accordo coi fenomeni di *long term potentiation* (LTP) (T.V.Bliss & T.Lomo, 1973; W.B.Levy & O.Steward, 1979) e *long term depression* (LTD) (P.k.Stanton & T.J.Sejnowski, 1989) rilevati studiando gli effetti di stimolazioni combinate sui piramidali. Il fenomeno della LTP si manifesta come rinforzo (temporaneo) del contatto sinaptico quando una scarica sinaptica coglie il neurone bersaglio in uno stato di depolarizzazione di ampiezza maggiore di circa 20 mV, mentre quello della LTD consiste nell'indebolimento del contatto quando la scarica lo coglie in uno stato di depolarizzazione inferiore a detto valore. Tenendo conto che la depolarizzazione di un neurone è sufficientemente alta solo se il neurone è attivo, si deduce che la regola di Hebb–Hopfield sarebbe completamente confermata, almeno per le sinapsi eccitatorie dei piramidali, se, oltre alle proprietà rilevate, avvenisse anche il rinforzo sinaptico di neuroni inattivi.

All'epoca in cui Hopfield propose questo modello erano già noti gli importanti risultati teorici ottenuti da Giorgio Parisi (1979) circa le proprietà di un modello matematicamente equivalente, che andava sotto il nome di *vetro di spin* (S.Kirkpatrick & D.Sherrington, 1978). Più tardi fu possibile mettere in evidenza importanti proprietà di auto-organizzazione degli stati stabili del sistema, in particolare dei ricordi impressi con criteri hebbiani. Precisamente la loro tendenza a 'organizzarsi' secondo uno schema *ultrametrico*, ossia gerarchico o ad albero, conformemente al criterio di ascendenza verso similarità crescenti (G.Parisi, 1983; M.Mèzard, G.Parisi, N.Sourlas, G.Toulouse e M.Virasoro, 1983; H.Gutfreund, 1988). Questo comportamento trova spiegazione nel fatto che la facilità di transizione da un ricordo all'altro è tanto maggiore quanto più questi ricordi possiedono una porzione comune, ossia quanto maggiore è il numero dei neuroni eccitati in comune nei ricordi evocati. Si noti che, per modelli di reti di tipo più generale, non accade necessariamente che un insieme di ricordi, o stati stabili della rete, risultino collegate da relazioni di reciproca accessibilità secondo un'organizzazione gerarchica. Ad esempio le strutture topologiche delle relazioni di similarità riscontrate in certe analisi statistiche assomigliano a distanze tra punti di uno spazio multidimensionale. La proprietà mostrata dalle reti di Hopfield dipende essenzialmente dal fatto che, avendo forzato la rete ad assumere certi ricordi, non si può impedire che si generino, proprio a causa del carattere combinatorio-lineare delle azioni postsinaptiche, stati stabili di similarità intermedie (stati spuri) (J.Hopfield, D.I.Feinstein & R.G.Palmer, 1983; D.J.Amit, H.Gutfreund & H.Sompolinsky, 1985, 1987). Ed è proprio per intercalazione di questi stati spuri tra i ricordi autentici che si forma un sistema complessivo di stati stabili organizzati secondo una topologia ultrametrica. A prima vista, la formazione degli stati spuri potrebbe essere interpretata come una specie di capacità di generare spontaneamente contenuti nuovi; ma a un'analisi più attenta ci si rende conto che tali presunte creazioni sono in realtà chimere illogiche prive di sensate relazioni con i ricordi autentici.

Purtroppo il modello di Hopfield è risultato poco convincente ai neurofisiologi e a tutti coloro che nel frattempo hanno cercato modelli più realistici. I difetti strutturali più spesso denunciati, a parte quelli comportamentali sopra rilevati, sono il *connessionismo massimale*, cioè la condizione che ogni neurone sia connesso con tutti gli altri, e la precisa *simmetria degli accoppiamenti*. Tuttavia queste critiche possono essere facilmente respinte in primo luogo proponendolo solo come modello di unità colonnari corticali (nelle quali, infatti, è ravvisabile sia la connessione massimale che una simmetria approssimativa degli accoppiamenti tra neuroni omologhi) e in secondo luogo facendo valere la dimostrazione matematica che il modello di Hopfield è strutturalmente stabile per piccole violazioni della simmetria dei coefficienti di accoppiamento (R.Nobili, 1991).

A parere dello scrivente, i difetti più rilevanti del modello sono invece i seguenti: 1) la simmetria tra azioni inibitorie ed eccitatorie; 2) l'eccessiva semplicità della rappresentazione dei ritardi temporali; 3) l'improponibilità della regola di Hebb per le variazioni dei coefficienti sinaptici ad effetto inibitorio.

Riguardo al primo punto, c'è da osservare che i potenziali postsinaptici del modello di Hopfield sono combinazioni lineari delle attività dei neuroni afferenti, con coefficienti positivi per le azioni eccitatorie e negativi per quelle inibitorie; mentre, a causa della molteplicità dei siti di attivazione dei canali sinaptici (da due a tre, a seconda del neurotrasmettitore), si deve assumere che gli effetti postsinaptici siano funzioni non lineari delle ampiezze dei segnali afferenti. Di conseguenza, poiché gli interneuroni intervengono moltiplicativamente nella trasmissione dei segnali, sarebbe naturale attendersi una sostanziale asimmetria tra gli effetti delle afferenze eccitatorie dirette (piramidale \rightarrow piramidale) e quelli delle afferenze inibitorie indirette (piramidale \rightarrow cellula a canestri \rightarrow piramidale). Come abbiamo avuto occasione di rilevare in un precedente paragrafo, questa asimmetria avrebbe un ruolo importante nell'impedire che, a causa del carattere sinergico dell'interazione eccitatoria, le attività dei neuroni raggiungano rapidamente livelli di saturazione (mentre invece, nel modello di Hopfield, è proprio la saturazione che limita l'attività neurale). Ciò sarebbe in accordo col fatto che le frequenze di scarica dei piramidali osservate *in vivo* superano di rado i 100 *spikes* al secondo, mentre dovrebbero arrivare a saturazione a oltre un migliaio di *spikes/sec*.

Riguardo al secondo punto, l'assunzione di un comune tempuscolo di ritardo sinaptico τ per tutti i neuroni contrasta chiaramente col fatto che gli interneuroni inibitori introducono ritardi suppletivi rispetto alle interazioni eccitatorie dirette. D'altronde è noto che l'improvvisa stimolazione di una popolazione di neuroni corticali genera alla prima una breve e intensa fase eccitatoria a cui fa seguito una fase inibitoria di durata maggiore, al termine della quale il sistema si spegne.

Per quanto riguarda il terzo punto, bisogna considerare che le variazioni dei pesi sinaptici secondo la regola di Hebb non è applicabile se il bersaglio sinaptico di un piramidale è un interneurone che può trovarsi in uno stato di depolarizzazione generalmente indipendente da quello dei piramidali a cui esso afferisce.

Dunque, se si ritiene che l'asimmetria e lo sfasamento temporale tra le componenti eccitatorie e inibitorie, unitamente al carattere non hebbiano della depotenziazione inibitoria dei piramidali, abbiano importanti ruoli funzionali nelle reti nervose reali, allora è particolarmente significativo rilevare che questi ruoli funzionali non possono essere adeguatamente

interpretati dal modello di Hopfield.

12. Generi prossimi e differenze specifiche.

Nonostante i difetti elencati, il modello di Hopfield ha suscitato molto interesse tra i fisici, i quali in pochi anni hanno saputo esibire una profusione di risultati teorici su varianti e aspetti particolari (A.Crisanti & H.Sompolinsky, 1987; K.E.Kürten & J.W.Clark, 1987; B.Lautrup, 1988 ecc.). In particolare è stata studiata la possibilità di farlo funzionare come sistema di memoria auto-organizzante. Si è così trovato che il procedimento più adeguato per sfruttare in questo senso le risorse del modello è quello cosiddetto della *ricottura (annealing)*. Si tratta, in breve, di simulare una specie di ‘riscaldamento’ della rete stimolando i neuroni con segnali rapidamente fluttuanti in modi casuali e indipendenti attorno a certi valori medi complessivamente assunti come stimolo iniziatore di un ricordo. In queste condizioni lo stato della rete transita erraticamente entro un insieme di stati, generalmente instabili, costituenti una specie di *intorno* topologico comune a uno o più stati stabili. Aumentando l’ampiezza delle fluttuazioni, ossia la ‘temperatura’, l’intorno si espanderà abbracciando nuovi stati stabili fino a invadere l’intero spazio degli stati oltre una certa temperatura critica. Diminuendo la temperatura esso si restringerà in un sottointorno dal quale alcuni stati stabili potranno restare esclusi dal precedente intorno. Infine, alla temperatura nulla, il sottointorno si sarà ridotto a un singolo stato stabile. Evidentemente, la probabilità che al diminuire della temperatura il sistema resti confinato in uno piuttosto che un altro sottointorno, dipenderà dai valori medi dei segnali stimolatori e saranno favorite le restrizioni a sottointorni di stati stabili maggiormente prossimi a quei valori medi. Pertanto, a ogni definita temperatura, lo spazio degli stati si presenta suddiviso in un insieme di intorni non comunicanti. Questa suddivisione si *ramifica* in suddivisioni via via più fini in corrispondenza di temperature via via minori, fino a ridursi, a temperatura nulla, alla semplice collezione degli stati stabili. Invece, sopra la temperatura critica, tutti i rami convergono al tronco. L’organizzazione gerarchica degli stati stabili, secondo relazioni di maggiore o minore similitudine, consiste proprio in questa diramazione degli intorni durante i processi di ‘ricottura’. Il procedimento di recupero di un ricordo può effettuarsi nel seguente modo: prima ‘riscaldando’ la rete, sotto l’influenza di uno stimolo medio evocatore (frammento di ricordo), a una temperatura tale che l’intorno possa ‘catturare’ lo stato stabile corrispondente al ricordo completo; poi ‘raffreddandola’ fino a ottenere il confinamento del ricordo completo entro un’intorno abbastanza piccolo. Si può ravvisare in ciò qualcosa di simile al procedimento di *categorizzazione per generi prossimi e differenze specifiche* (Aristotele, 306–367 a.C.); precisamente interpretando la risalita verso il tronco, determinata dal riscaldamento, come la fase di individuazione del genere prossimo e la discesa verso un ramo terminale, determinata dal raffreddamento, come quella di recupero di una differenza specifica. È evidente che un ‘riscaldamento’ moderato, sotto l’effetto di uno stimolo medio iniziatore, applicato a una rete che già si trova in un certo stato avrà l’effetto di farla accedere con maggiore probabilità a un genere proprio di questo stato. Si potrebbe vedere in ciò una specie di fenomeno di *permanenza del contesto*. Purtroppo questo interessante comportamento delle reti di Hopfield ha un serio difetto: anche se i ricordi sono ‘ortogonali’, cioè privi di parti comuni, la probabilità di evocare chimere insensate, invece che ricordi autentici, rimane non trascurabile; così la

pretesa capacità categorizzante risultà in realtà piuttosto illusoria.

13. *Alla ricerca della memoria categorizzante.*

Ponendo a confronto il comportamento del modello di Hopfield con quello dei sistemi nervosi reali, si possono osservare alcune analogie ma anche differenze piuttosto rilevanti. Che la memoria animale, e umana in particolare, non sia un semplice archivio di dati indirizzati, ma possieda una specie di capacità di organizzazione automatica dei ricordi, secondo schemi di categorizzazione abbastanza indipendenti dalle modalità di acquisizione, traspare con grande evidenza da tutta la letteratura scientifica sul linguaggio, dalle indagini della psicologia gestaltica e cognitivista e dalle ricerche sull'intelligenza artificiale (v. C.Cornoldi, 1978; H.Gardner, 1988). Se la percezione del genere prossimo è una funzione mentale essenziale nel riconoscimento delle forme, altrettanto lo è il discernimento delle differenze specifiche. La prontezza con cui la mente umana è in grado di 'percepire', anche i generi più astratti e le più sottili differenze, non ha bisogno di essere documentata. Altrettanto evidente è la tendenza più o meno pronunciata alla permanenza del contesto. Ma il comportamento della memoria animale diverge in vari punti da quello esibito dal modello in questione, e rivela persino aspetti apparentemente contraddittori. In primo luogo il processo di 'ricottura' delle reti di Hopfield deve effettuarsi in modo *adiabatico*, cioè in tempi molto lunghi rispetto alla durata media di una fluttuazione, mentre per la memoria animale reale i tempi di evocazione dei ricordi hanno in genere valori confrontabili con le durate delle fasi di eccitazione della rete nervosa ($\approx 1/10$ di sec.). Inoltre, nel modello, l'individuazione del genere prossimo è ottenibile solo effettuando una media temporale sugli stati erraticamente fluttuanti e il recupero di una differenza specifica può essere ottenuto sottraendo tale media dallo stato finale della rete 'ricotta' e raffreddata. Ciò significa che in pratica, per l'effettivo funzionamento della memoria, si richiedono apparati di calcolo ausiliari. Un problema simile si presenta in relazione alla registrazione dei ricordi. Alcuni autori ritengono che l'organizzazione dei ricordi appartenenti a un genere comune cominci dalla memorizzazione di un prototipo, o comunque avvenga in relazione alla memorizzazione di uno stereotipo. Verrebbe da pensare che il cervello registri solo differenze specifiche tra gli stimoli nuovi e quelli già memorizzati. Ciò troverebbe una conferma nel fatto che, nel riconoscimento delle forme, il processo evocativo sembra funzionare in modo tale da evidenziare le novità neutralizzando gli aspetti costanti o rimuovendo la componente relativa al genere, ad esempio inibendo per assuefazione la percezione del contesto. Questo tipo di funzionamento, che nei modelli finora considerati potrebbe ottenersi in modi alquanto artificiosi, è stato formalizzato da T.Kohonen & E.Oja (1986, 1987), i quali hanno assunto l'esistenza di procedure ricorsive di *ortogonalizzazione* preliminare degli stimoli da ricordare rispetto a quelli già registrati.

Un'altra rilevante discrepanza è rinvenibile nel fatto che nelle reti nervose reali gli stimoli agiscono in modi immediati e transitori, in contrasto con la necessità della loro permanenza nei processi di ricottura artificiali. Inoltre le categorizzazioni della memoria reale non sono necessariamente *ad albero*, ad esempio quelle della memoria linguistica possono avere strutture *a rizoma*, *a reticolo*, ecc.: vale a dire che una stessa differenza specifica può fare capo a più generi diversi, ossia che uno stesso sottointorno dovrebbe poter appartenere a topologie diverse. In altri termini per le relazioni di transitività tra

intorni di stati stabili di una rete reale devono essere possibili topologie intermedie tra quelle ultrametriche e quelle metriche, con la possibilità di controllo multiparametrico degli accessi ai generi prossimi e la riduzione alle differenze specifiche.

Poiché tutti questi difetti di comportamento sono concomitanti ai difetti strutturali messi in evidenza alla fine del paragrafo precedente, è naturale cercare di capire se reti neurali meno difettive dal punto di vista strutturale ammettano comportamenti più soddisfacenti. Cercando di correlare le proprietà strutturali che dovrebbe avere un modello più evoluto di quello di Hopfield (asimmetria tra gli effetti postsinaptici eccitatori e quelli inibitori, anticipazione dei primi sui secondi, regole di variazione dei coefficienti sinaptici per la componente inibitoria neurofisiologicamente ammissibili, ecc.) con alcune, almeno, di quelle comportamentali, attendibili per le reti reali (transitorietà della stimolazione, rapidità del processo di categorizzazione, selezione delle pure differenze specifiche, ecc.) si giungerebbe a prospettare, invece del procedimento di ‘ricottura’ descritto sopra, il seguente principio di funzionamento: *la stimolazione della rete ha un effetto rapido e transitorio; essa innesca l’accesso immediato al genere prossimo nella fase eccitatoria, cui segue la rapida selezione della differenza specifica nella fase inibitoria* (senza bisogno di simulare processi di ricottura, calcolare valori medi ed effettuare sottrazioni). Ciò significa che nella fase eccitatoria dovrebbero attivarsi spontaneamente funzioni sinergiche estensive (tipo unione insiemistica) e in quella inibitoria funzioni anergiche intensive (tipo intersezione insiemistica). Un modello di rete multistabile, ad effetti eccitatori lineari e inibitori quadratici ritardati, capace di funzionare secondo questo principio, è stato studiato dallo scrivente (1990). È stato possibile mettere in evidenza le seguenti caratteristiche: 1) sostanziale riduzione degli stati spuri; 2) organizzazione dei ricordi in modi dipendenti dalla struttura dei confini di criticità nello spazio degli stimoli possibili; in tal modo si trova che la teoria dei punti critici (v. V.I. Arnold A. Varchenko e S. Goussein-Zadè, 1986), comunemente nota come teoria delle catastrofi, diviene lo strumento matematico più adeguato allo studio delle forme di categorizzazione; 3) formazione preliminare dei generi prossimi come attrattori dello spazio degli stati; 4) successiva trasformazione degli attrattori in repulsori e formazione, attorno a questi, di un nuovo sistema di attrattori, corrispondenti alle differenze specifiche, in reciproca competizione isteresica.

14. Osservazioni conclusive. La macchina della mente.

Nei paragrafi precedenti, utilizzando concetti informativi suggeriti da alcuni modelli di reti nervose, abbiamo rivolto l’attenzione a taluni aspetti particolari del funzionamento cerebrale; principalmente a quelli che riguardano l’organizzazione della percezione, l’attenzione selettiva, il funzionamento della memoria, le sue proprietà categorizzanti, i processi di codificazione e trasduzione parallela dei flussi di informazione. Tutti aspetti, si noti che potrebbero farsi rientrare nel genere delle relazioni stimolo-risposta. Si tratta di processi nervosi a esito statico: un complesso di segnali applicati simultaneamente e transitoriamente in ingresso determina un processo più o meno rapido di attivazione e disattivazione della cascata reticolare della rete nervosa in modo che in uscita risulta prodotto un flusso d’informazione organizzato in modo intelligente, altamente funzionale alla varietà degli usi possibili e condizionato dall’esperienza percettiva accumulata dalla rete. Ma in questo modo la nostra indagine ha potuto toccare essenzialmente soltanto gli aspetti cog-

nitivi del processo mentale, rimanendo escluso quelli *comportamentali*. In altri termini, abbiamo posto attenzione al versante sensoriale del flusso nervoso e ignorato tutto ciò che può riguardare l'attività motoria e quella centrale autonoma; abbiamo analizzato questioni che possono avere importanza in relazione alla formazione degli *stati mentali* ma non quelle che possono riguardare le loro modalità di trasformazione; abbiamo considerato il problema della formazione delle reti nervose negli aspetti che possono riguardare la *memoria dichiarativa* e ignorato o trascurato quelli riguardanti la *memoria procedurale*.

Il fatto che esistano almeno due generi diversi di memoria è un'acquisizione piuttosto recente (v. R.G.M.Morris, E.R.Kandel & L.R.Squire, 1988). Il ricordo di un fatto, di un volto, di un nome ecc., sono caratteristici del primo tipo di memoria, mentre il saper guidare l'automobile, il reagire in un certo modo in presenza di una certa persona, il saper organizzare un certo discorso ecc., sono tutte manifestazioni della memoria del secondo tipo. Non si tratta di una semplice distinzione formale, poiché le lesioni del circuito di Papez (corteccia → ippocampo → amigdala → corteccia) negli umani e nei mammiferi in genere, distrugge la possibilità che si formino nuovi ricordi nella memoria dichiarativa, ma non compromette affatto la capacità di acquisire nuove abilità procedurali. In effetti, tutte le forme di apprendimento che abbiamo considerato nei precedenti paragrafi hanno un carattere cognitivo e riguardano la memoria dichiarativa. Ma ve ne sono altre che non possono rientrare nei modelli da noi trattati (ad esempio l'apprendimento per prove ed errori, che coinvolge inequivocabilmente la memoria procedurale); e altre ancora, come l'apprendimento emulativo, che dipendono dalla combinazione di processi cognitivi e comportamentali e che, presumibilmente, richiedono la coordinazione dei due generi di memoria. Quest'ultima facoltà, già menzionata in relazione alla decentrazione proiettiva e simbolo del sè, deve ritenersi tipica del funzionamento cerebrale degli animali in genere.

In effetti, in questa breve escursione è stato trascurato il problema della *macchina della mente*, ossia dell'esistenza e del funzionamento di un apparato cerebrale capace di trasformare ricorsivamente gli stati mentali: infatti nessuno dei processi nervosi modellizzati nei precedenti paragrafi è in grado di fornire un'interpretazione della capacità di un animale di agire spontaneamente o rivelare una qualche sorta di attività nervosa permanente. Se il cervello consistesse di una semplice cascata areolare frapposta tra l'apparato sensoriale e quello effettore, una volta soppresso il flusso sensoriale l'attività nervosa si spegnerebbe immediatamente. È invece evidente che i cervelli reali sono capaci di promuovere attività motorie anche in assenza di stimolazioni esogene e inoltre è noto che il cervello esibisce un'attività permanente anche in condizioni di deprivazione sensoriale e motoria. Come si producono queste autonome capacità di funzionamento? L'esistenza di un processo ricorsivo centrale è presumibilmente riconoscibile nella funzione del circuito extrapiramidale (corteccia → gangli basali → talamo → corteccia), la cui attività è sostenuta e regolata a livello talamico da quella sorgente permanente di segnali attivatori che è la formazione reticolare ascendente. In questa sede non possiamo che accennare di sfuggita a tale questione. Limitandoci ad osservare che una 'macchina della mente' come questa appare indispensabile affinché, nel cervello umano, abbia luogo il processo autoreferenziale discusso nel paragrafo 2.

Padova 21.12.90

BIBLIOGRAFIA

- D.J.Amit, H.Gutfreund & H.Sompolinsky, *Phys. Rev. A*, **32**, 1007–18 (1985); **35**, 2293–2303 (1987); *Ann. of Phys.*, **173**, 30–67 (1987).
- M.Arbib, *La mente, le macchine e la matematica.*, Ed. Boringhieri (Bologna, 1968).
- Aristotele, *Categorie*, in *Opere*, vol.I, Ed.Laterza, Bari (1973).
- V.Arnold, A.Varchenko & S.Goussein–Zadè, *Singularité des application différentiables* — voll.1,2, Ed MIR (Mosca, 1986).
- C.Asanuma & F.Crick, in *Parallel Distributed Processing*, J.L. McClelland & D.E.Rumelhart Eds., Vol. 2, 333–371, (1986).
- H.B.Barlow, in *Cybernetics* C.R.Evans & A.D.J.Robertson Eds., p. 183–207, Butterworths (London, 1968).
- G.Berlucchi, *Acta Psychologica*, 1990; in *Brain and Reading*, C.von Euler Ed., McMillan London (1990).
- G.Birkhoff, *Proc. Camb. Phil. Soc.* **29** 441–64 (1933).
- T.V.Bliss & T.Lomo, *J. Physiol. Lond.* **232**, 331–356 (1973).
- G.Boole, *Indagine sulle leggi del pensiero*. Einaudi Ed. (Torino, 1976).
- J.Borman & D.E.Clapham, *Proc. Natl. Acad. Sci.* **82**, 2168–72 (1985).
- A.Borsellino & T.Poggio, *Kibernetik*, **13**, 10 (1972).
- S.Bottini, *Biol. Cybern.*, **36**, 211 (1980).
- B.Cavanagh, *Perception*, **7**, 167–177, (1978).
- G.Cibenko, *Math. Control System Signals*, in stampa (1990).
- C.Cornoldi, *Modelli della memoria*. Ed. Giunti Barbera (Firenze, 1978).
- A.Crisanti & H.Sompolinsky, *Phys. Rev. A*, **36**, 4922–39 (1987).
- R.Eckorn & H.J. Reitboeck, 99–111; C.M.Gray, P.König, A.K.Engel and W.Singer, 82–98; in *Synergetics of Cognition*, H.Haken and M.Stadler Eds., Springer–Verlag (Berlin, 1989).

- M.Farah, *TINS*, **12**, 395–399, (1989).
- F.Fogelman, *Int. Neur. Net. Conf.* Tutorial paper, 1990.
- H.Gardner, *La nuova scienza della mente*. Ed. Feltrinelli, Milano (1988).
- D.Gabor, *Nature*, **217a**, 548 (1968).
- G.Girosi & T.Poggio, *Science*, **247**, 978–82, (1990); *Biol. Cybern.*, **63**, 169–176 (1990).
- S.Grossberg, *Neural Networks and Natural Intelligence*. The MIT Press Ed. (1988).
- R.Elul, *Int. Rev. Neurobiol.*, 15 (1972).
- D.O.Hebb, *The Organization of Behavior*. John Wiley Ed. (1949); trad. ital. *L'organizzazione del comportamento*, Ed. F.Angeli (1975).
- P.van Heerden, *Appl. Opt.*, **2**, 387 (1973).
- D.R.Hofstadter, *Gödel, Escher, Bach: un'Eterna Ghirlanda Brillante*. Adelphi Ed. (Milano, 1985).
- J.J.Hopfield, *Proc. Natl. Acad. Sci. USA*, **79**, 2554–58 (1982) e **81**, 3088–92 (1982).
- J.Hopfield, D.I.Feinstein & R.G.Palmer, *Nature*, **304**, 158–59 (1983).
- D.H.Hubel, *Occhio, cervello e visione*. Ed. Zanichelli, (Bologna, 1989).
- S.Kirkpatrick & D.Sherrington, *Phys. Rev. B*, **17**, 4384 (1978).
- T.Kohonen & E.Oja, *Biol. Cybernetics*, **21**, 85–95, (1976).
- K.E.Kürten & j.W.Clark, *Phys. Lett*, **114 A**, 413–418 (1986).
- W.B.Levy & O.Steward, *Neurosciences* **8**, 791–797 (1983).
- H.C.Longuet-Higgins, *Nature*, **217a**, (1968).
- A.de Luca e L.M.Ricciardi, *Introduzione alla cibernetica.*, F.Angeli Ed. (Milano, 1981).
- B.B.Mandelbrot, *The Fractal Geometry of Nature*, W.H.Freeman & C. Ed., New York (1983).

- A.Mazurkiewicz, voce *Algoritmo* in *Enciclopedia*, Einaudi Ed., (Torino, 1977).
- S.W.McCulloch & W.Pitts, *Bull. Math. Biophysics*, **5**, 115–133 (1943).
- M.Mèzard, G.Parisi, N.Sourlas, G.Toulouse & M.Virasoro, *Phys. Rev. Lett.*, **52**, 1156–1159 (1984).
- R.G.M.Morris, E.R.Kandel & L.R.Squire, *TINS, special issue* **11**, 125–127, (1988).
- J.von Neumann, *Theory of Self-Reproducing Automata*. University of Illinois Press (Urbana, 1966).
- R.Nobili, *Phys. Rev. A*, **32**, 3618 (1985); *ibid*, **35**, 1901 (1987); *Proc. INNC-90*, Kluver Acad. Pub., vol.I, 517 (1990); *Neural Networks as Multistable Systems*, Preprint, Padova (1991).
- K.Okajima, *Proc. INNC-90*, Kluver Acad. Pub., vol.II, 504–507 (1990).
- G.Parisi, *Phys. Rev. Lett.*, **50**, 1946–1948 (1983).
- H.Putnam, voce *Logica* in *Enciclopedia*, Einaudi Ed., (Torino, 1979).
- D.E.Rumelhart, G.E.Hinton & R.J.Williams, *Nature*, **323**, 533–536 (1986).
- E.L.Schwartz, *Perception*, **10**, 455–468, (1981).
- C.Shannon, *La teoria matematica delle comunicazioni*. Ed. Eta Compas (1971).
- P.K.Stanton & T.J.Sejnowski, *Nature* **339**, 215–218 (1989).
- M.H.Stone, *Proc. Nat. Acad.* **20**, p. 197 e seg. (1934).
- B.Widrow & M.A.Lehr, *Proc. of the IEEE. Special issue on neural networks, I*, **78**, 1415–1442 (1990).